

DOI: <https://doi.org/10.17721/1812-5409.2025/2>

Наведено результати теоретичних, прикладних і науково-експериментальних досліджень з актуальних питань фізики та математики.

Для науковців, викладачів, студентів та аспірантів.

ГОЛОВНИЙ РЕДАКТОР РЕДАКЦІЙНА КОЛЕГІЯ	Розора Ірина, д-р фіз.-мат. наук, доц. (Україна) Анікушин Андрій, канд. фіз.-мат. наук, доц. (Україна); Анісімов Анатолій, акад. НАН України, д-р фіз.-мат. наук, проф. (Україна); Анісімов Ігор, д-р фіз.-мат. наук, проф. (Україна); Бавула Володимир, проф. (Велика Британія); Бартлова Мілада, PhD (Чеська Республіка); Бегін Луїза, проф. (Італія); Безущак Оксана, д-р фіз.-мат. наук, проф. (Україна); Гаврюшенко Дмитро, д-р фіз.-мат. наук, проф. (Україна); Горячко Андрій, д-р фіз.-мат. наук, доц. (Україна); Джуліано Ріта, проф. (Італія); Дмитрук Ігор, д-р фіз.-мат. наук, проф. (Україна); Жук Ярослав, д-р фіз.-мат. наук, проф. (Україна); Заславський Володимир, д-р техн. наук, проф. (Україна); Затула Дмитро, канд. фіз.-мат. наук, доц. (заст. голов. ред.) (Україна); Капустян Олексій, д-р фіз.-мат. наук, проф. (Україна); Капустян Олена, д-р фіз.-мат. наук, ст. наук. співроб. (Україна); Кашпур Олена, д-р фіз.-мат. наук, доц. (Україна); Коган Ніколас, проф. (США); Кондратенко Сергій, д-р фіз.-мат. наук, проф. (Україна); Кудін Володимир, д-р техн. наук, ст. наук. співроб. (Україна); Леоненко Микола, проф. (Велика Британія); Львов Віктор, д-р фіз.-мат. наук, проф. (Україна); Майборода Ростислав, д-р фіз.-мат. наук, проф. (Україна); Макарець Микола, д-р фіз.-мат. наук, проф. (Україна); Медвідс Артурс, д-р наук, проф. (Латвія); Мішура Юлія, д-р фіз.-мат. наук, проф. (Україна); Оленко Андрій, проф. (Австралія); Орсігер Енцо, проф. (Італія); Пашко Анатолій, д-р фіз.-мат. наук, проф. (Україна); Петравчук Анатолій, д-р фіз.-мат. наук, проф. (Україна); Погані Тібор, проф. (Хорватія); Погорілий Сергій, д-р техн. наук, проф. (Україна); Ронто Міклос, д-р наук, проф. (Угорщина); Савенков Сергій, д-р фіз.-мат. наук, доц. (Україна); Сільвестров Дмитро, д-р наук, проф. (Швеція); Скришевський Валерій, д-р фіз.-мат. наук, проф. (Україна); Сливка-Тилищак Ганна, д-р фіз.-мат. наук, проф. (Україна); Соттінен Томмі, проф. (Фінляндія); Тору Аокі, проф. (Японія); Трофімчук Сергій, проф. (Чилі); Футорний Вячеслав, проф. (Бразилія); Худак Стефан, д-р наук, проф. (Словацька Республіка); Яневич Тетяна, канд. фіз.-мат. наук, ст. наук. співроб. (відп. ред.), (Україна)
Адреса редколегії	Факультет комп'ютерних наук та кібернетики просп. Академіка Глушкова, 4д, м. Київ, 03022 ☎ (38044) 259 01 49 e-mail: physmath.bulletin@knu.ua web: https://bphm.knu.ua
Затверджено	вченою радою факультету комп'ютерних наук та кібернетики 18.11.25 (протокол № 5)
Зареєстровано	Національною радою України з питань телебачення і радіомовлення Рішення № 1089 від 28.03.24 Ідентифікатор друкованого медіа: R30-03805
Атестовано	Міністерством освіти і науки України (категорія А) Наказ № 641 від 28.04.25
Індексування	Scopus, Zentralblatt MATH (zbMATH), JournalTOCs, Crossref, ICI World of Journals, Road, Google Scholar, Наукова періодика України
Засновник та видавець	Київський національний університет імені Тараса Шевченка Видавничо-поліграфічний центр "Київський університет" Свідоцтво про внесення до Державного реєстру ДК № 1103 від 31.10.02
Адреса видавця	ВПЦ "Київський університет" б-р Тараса Шевченка, 14, м. Київ, 01601 ☎ (38044) 239 32 22, 239 31 58, 239 31 28 e-mail: vpcc@knu.ua

BULLETIN

TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV

ISSN 1812-5409 (Print), ISSN 2218-2055 (Online)

PHYSICS AND MATHEMATICS

2(81)/2025

Established in 2005

DOI: <https://doi.org/10.17721/1812-5409.2025/2>

The Bulletin highlights the results of theoretical, scientific, applied and scientific-experimental researches on the current issues of Physics and Mathematics.

For scientists, researchers, lecturers, PhD students and students.

EDITOR-IN-CHIEF	Rozora Iryna, DSc (Phys. & Math.), Assoc. Prof. (Ukraine)
EDITORIAL BOARD	Anikushyn Andrii, PhD (Phys. & Math.), Assoc. Prof. (Ukraine); Anisimov Anatoliy, Active Member of the National Academy of Sciences of Ukraine, DSc (Phys. & Math.), Prof. (Ukraine); Anisimov Igor, DSc (Phys. & Math.), Prof. (Ukraine); Bartlova Milada, PhD (Czech Republic); Bavula Vladimir, Prof. (Great Britain); Beghin Luisa, Prof. (Italy); Bezuschak Oksana, DSc (Phys. & Math.), Prof. (Ukraine); Cogan Nicholas, Prof. (USA); Dmitruk Igor, DSc (Phys. & Math.), Prof. (Ukraine); Futorny Vyacheslav, Prof. (Brazil); Gavryushenko Dmytro, DSc (Phys. & Math.), Prof. (Ukraine); Giuliano Rita, Prof. (Italy); Goriachko Andrii, DSc (Phys. & Math.), Assoc. Prof. (Ukraine); Hudak Stefan, DSc, Prof. (Slovak Republic); Ianevych Tetiana, PhD (Phys. & Math.), Senior Researcher (Managing Editor) (Ukraine); Kashpur Olena, DSc (Phys. & Math.), Assoc. Prof. (Ukraine); Kapustian Olena, DSc (Phys. & Math.), Senior Researcher (Ukraine); Kapustyan Oleksiy, DSc (Phys. & Math.), Prof. (Ukraine); Kondratenko Serhiy, DSc (Phys. & Math.), Prof. (Ukraine); Kudin Volodymyr, DSc, (Engin.), Senior Researcher (Ukraine); Leonenko Nikolay, Prof. (Great Britain); Lvov Victor, DSc (Phys. & Math.), Prof. (Ukraine); Maiboroda Rostyslav, DSc (Phys. & Math.), Prof. (Ukraine); Makarets' Nikolay, DSc (Phys. & Math.), Prof. (Ukraine); Medvids Arturs, DSc (habil.), Prof. (Latvia); Mishura Yuliya, DSc (Phys. & Math.), Prof. (Ukraine); Olenko Andriy, Prof. (Australia); Orsingher Enzo, Prof. (Italy); Pashko Anatolij, DSc (Phys. & Math.), Prof. (Ukraine); Petravchuk Anatolij, DSc (Phys. & Math.), Prof. (Ukraine); Pogány Tibor, Prof. (Croatia); Pogorily Serhiy, DSc (Engin.), Prof. (Ukraine); Rontó Miklós, DSc, Prof. (Hungary); Savenkov Serhij, DSc (Phys. & Math.), Assoc. Prof. (Ukraine); Silvestrov Dmitrii, DSc, Prof. (Sweden); Skryshevsky Valery, DSc (Phys. & Math.), Prof. (Ukraine); Slyvka-Tylyshchak Anna, DSc (Phys. & Math.), Prof. (Ukraine); Sottinen Tommi, Prof. (Finland); Toru Aoki, Prof. (Japan); Trofimchuk Sergey, Prof. (Chile); Zaslavs'kyj Volodymyr, DSc (Engin.), Prof. (Ukraine); Zatula Dmytro, PhD (Phys. & Math.), Assoc. Prof. (Deputy Editor-in-Chief) (Ukraine); Zhuk Yaroslav, DSc (Phys. & Math.), Prof. (Ukraine)
Address	Faculty of Computer Science and Cybernetics 4d, Hlushkov ave., Kyiv, 03022 ☎ (38044) 259 01 49 e-mail: physmath.bulletin@knu.ua web: https://bphm.knu.ua
Approved by	the Academic Council of the Faculty of Computer Science and Cybernetics 18.11.25 (protocol № 5)
Registered by	the National Council of Television and Radio Broadcasting of Ukraine Decision № 1089 of 28.03.24 Identifier of printed media: R30-03805
Certified by	the Ministry of Education and Science of Ukraine (category A) Order № 641 dated 28.04.25
Indexing	Scopus, Zentralblatt MATH (zbMATH), JournalTOCs, Crossref, ICI World of Journals, Road, Google Scholar, Scientific periodicals of Ukraine
Founder and publisher	Taras Shevchenko National University of Kyiv Publishing and Polygraphic Center "Kyiv University" Certificate of entry into the State Register ДК № 1103 dated 31.10.02
Address	PPC "Kyiv University" 14, Taras Shevchenko blvd., Kyiv, 01601 ☎ (38044) 239 32 22, 239 31 58, 239 31 28 e-mail: vpc@knu.ua

ЗМІСТ

ПОНОМАРЬОВ Вадим, МОРЕНЕЦЬ Володимир, РОЗОРА Ірина, ШАРАПОВ Михайло, ЧОРНЕЙ Руслан, СЛАБОСПИЦЬКИЙ Олександр Павло Кнопов. Науковий шлях та становлення школи стохастичної оптимізації	7
---	---

АЛГЕБРА, ГЕОМЕТРІЯ ТА ТЕОРІЯ ЙМОВІРНОСТЕЙ

БОНДАРЕНКО Наталія, БОНДАРЕНКО Євген Матричні зображення ітерованих вінцевого добутоків одновимірних алгебр Лі	12
БУНДЮК Євгеній, МІШУРА Юлія, ПІРОЦЦІ Енріка Формули прогнозу для інтегрованого дробового Броунівського руху	19
ГАЛУШКО Олександр Прогностична модель логістичної регресії для оцінки показань до лікувальної імплантації при пролабуванні шлунку в стравохід або кардіальну ділянку шлунку	31
СЛАГІН Володимир Неперервні перетворення відрізка, що зберігають хвости нега-двійкового зображення чисел з відрізка $[0;1]$	36
МАЙБОРОДА Ростислав, СУГАКОВА Олена Вибір параметра згладжування для ядерної оцінки за спостереженнями з суміші	41
МАСЮТКА Олександр, МОКЛЯЧУК Михайло Задача інтерполяції векторних гармонізованих стійких послідовностей	47
ПРИШЛЯК Олександр, БУТОК Катерина Топологічна структура простих гамільтонових потоків на проєктивній площині та плящі Клейна	60
САДОК Микита Граничні теореми для узагальнених опуклих оболонок	73
ТИХОНЕНКО Дмитро, ЯМНЕНКО Ростислав, КУЧІНКА Каталін Уточнення характеристичної нерівності для норм φ -субгауссової випадкової величини	82

ДИФЕРЕНЦІАЛЬНІ РІВНЯННЯ, МАТЕМАТИЧНА ФІЗИКА ТА МЕХАНІКА

ВАЙСФЕЛЬД Наталя, ЖУРАВЛЬОВА Зінаїда Осесиметрична задача поропружності для шаруватого циліндру	87
ГРІНЧЕНКО Віктор, ЛЕБЕДЄВА Ірина, МАЦИПУРА Володимир, ТРИФОНОВ Тихомир Вплив згладжування поверхні скінченного клиноподібного об'єкта на характеристики розсіювання плоскої хвилі. Частина II. Аналіз чисельних результатів	91
ДРЕУС Андрій, ДЕРБАБА Віталій, ГАРКАВЕНКО Дмитро, АЛЕКСЄЄНКО Сергій Аналіз акустичного випромінювання від модифікованого профіля лопаті вітрової турбіни NACA0012	96
ЖУК Ярослав, ФАЛЛАХ Араш Солейман Технологія лазерного термоформування тонкостінних металевих конструкційних елементів	101
КАПУСТЯН Олексій, ЮСИПІВ Тарас, ПЕРЕСТЮК Юрій, ХАЛЕЦЬКА Зоя Рівномірні аттрактори та властивість асимптотичного підсилення для системи типу PDE-ODE	105
КУЗЬМЕНКО Володимир, ТРОЦЕНКО Ярослав Моделювання поширення звуку у плоскому хвилеводі з округлими вигинами	113
КУЛЯН Віктор, ЮНЬКОВА Олена, КОРОБОВА Марина Ідентифікація параметрів моделей динаміки активів	117
МЕЛЬНИЧЕНКО Микола, ЯРОВОЙ Леонід, ПУЧКО Наталя, ВОДОТОВКА Михайло Методика експериментального моделювання проникнення високошвидкісних ударних елементів у бронематеріалах	122
ОНИШКЕВИЧ Володимир, БАРАБАШ Галина Квазістатична контактна взаємодія скінченного жорсткого циліндра, що обертається, та пружного півпростору при фрикційному теплоутворенні	126
ПАСІЧНИК Галина Фундаментальний розв'язок задачі Коші для дисипативного ультрапараболічного рівняння Колмогорова з незалежними від змінних виродження коефіцієнтами	130
САМОЙЛЕНКО Олег Математичні проблеми співставлення ознак для об'єктів з обмеженими ресурсами керованих на основі зору	138

ЗМІСТ

ХАРИТОНОВ Олексій, КУЦЕНКО Олексій, КУЦЕНКО Анастасія	
Аналіз ефективності комбінування участі двигунів великої та малої тяги при перельотах Земля-Марс.....	142
ЧАЙКОВСЬКИЙ Андрій, ЛЮБИМОВ Олександр, ЛАГОДА Оксана	
Обмеженість розв'язків лінійного різницевого рівняння першого порядку у багатовимірному випадку	146
ШАМРАЙ Максим	
Неасимптотичні межі погіршення дисконтованої кумулятивної винагороди для OBD-проріджених нейронних контролерів	155

КОМП'ЮТЕРНІ НАУКИ ТА ІНФОРМАТИКА

ЗАВАДСЬКИЙ Ігор, КОВАЛЬЧУК Максим	
Оптимізація канонічних кодів Хафмана для послідовного стиснення природномовних текстів.....	159
ЛІПСЬКИЙ Даниїл	
Архітектура та функціональні особливості платформи для автоматизованого тестування веб-застосунків	166
МАКУШЕНКО Ігор, УСАР Ірина, ЛІВІНСЬКА Ганна	
Гістерезисна стратегія для одноканальної системи з чергою та повторними викликами.....	174
МАРТИНЮК Сергій, ЦУРКАН Вячеслав	
Вибір оптимальної стратегії для дискретної динамічної кооперативної гри	182
МАХНО Михайло, ФЕДОРУС Олексій, БОРИСЕНКО Олександр, ВЕРЕМЧУК Максим	
Архітектура системи виявлення ботів в соціальних мережах.....	187
ПРАВОСУД Руслан, МАРЧЕНКО Олександр	
Перенесення стилю діалогу за допомогою параметрично ефективного донавчання.....	193
СЛАБОСПИЦЬКИЙ Олександр	
Метод найменших квадратів відносних похибок для оцінювання параметрів систем за некласичних припущень.....	197
СТОЯН Володимир, ЧЕРНІЙ Дмитро, ВОЛОЩУК Сергій, ЗАТОРСЬКИЙ Роман	
Тривимірні задачі керування динамікою неповно спостережуваних товстих пружних плит. Частина II. Випадок дискретно заданого бажаного стану	202
СТЬОПКІН Андрій, ТУРКА Тетяна	
Алгоритм розпізнавання простих графів колективом агентів	211
ТЮРДЬО Іван, КІЗІЛОВА Наталія	
Застосування методів машинного навчання в сучасній антираковій терапії.....	217

СУЧАСНА ФІЗИКА

ЯКУНОВ Андрій, ДІБРОВ Володимир, САЧКО Артем	
Кореляція між фрактальною розмірністю текстури поверхні та її лазерними спеклами як основа безконтактного методу аналізу керамічних матеріалів	227

CONTENTS

PONOMAROV Vadym, MORENETS Volodymyr, ROZORA Iryna, SHARAPOV Mykhailo, CHORNEI Ruslan, SLABOSPITSKY Alexander Pavlo Knopov. The scientific path and the formation of the Kyiv school of stochastic optimization	7
--	---

ALGEBRA, GEOMETRY, AND PROBABILITY THEORY

BONDARENKO Nataliia, BONDARENKO Ievgen Matrix representations of iterated wreath products of one-dimensional Lie algebras	12
BUNDIUK Yevhenii, MISHURA Yuliya, PIROZZI Enrica Prediction formulas for integrated fractional Brownian motion	19
HALUSHKO Oleksandr Prognostic logistic regression model for assessing indications for therapeutic implantation with gastric prolapse into the esophagus or the cardiac part of the stomach	31
YELAHIN Volodymyr Continuous transformations preserving tails of negabinary representations of numbers from the unit interval	36
MAIBORODA Rostyslav, SUGAKOVA Olena Bandwidth selection for density estimation by mixture with varying concentrations	41
MASYUTKA Oleksandr, MOKLYACHUK Mikhail Interpolation problem for multidimensional harmonizable stable sequences	47
PRISHLYAK Alexandr, BUTOK Kateryna Topological structure of simple Hamiltonian flows on the projective plane and the Klein bottle	60
SADOK Mykyta Limit theorems for generalised convex hulls	73
TYKHONENKO Dmytro, YAMNENKO Rostyslav, KUCHINKA Katalin Improvement of characterization inequality for norms of ϕ -sub-Gaussian random variable	82

DIFFERENTIAL EQUATIONS, MATHEMATICAL PHYSICS AND MECHANICS

VAYSFELD Natalya, ZHURAVLOVA Zinaida Axisymmetric poroelasticity problem for a multilayered cylinder	87
GRINCHENKO Viktor, LEBEDYEVA Iryna, MATSYURA Volodymyr, TRIFONOV Tihomir Effect of smoothing of a finite wedge-shaped object on the scattering characteristics of a plane wave. Part II. Analysis of numerical results	91
DREUS Andrii, DERBABA Vitalii, HARKAVENKO Dmytro, ALEKSIEIENKO Serhii Numerical analysis of acoustic emission from a modified NACA0012 wind turbine blade profile	96
ZHUK Yaroslav, FALLAH Arash Soleiman Technology for laser thermoforming of thin-walled metal structural elements	101
KAPUSTYAN Oleksii, YUSYPIV Taras, PERESTYUK Yuriy, KHALETSKA Zoia Uniform attractors and asymptotic gain property for the PDE-ODE system	105
KUZMENKO Volodymyr, TROTSENKO Yaroslav Modeling of sound propagation in a planar waveguide with rounded bends	113
KULIAN Victor, YUNKOVA Olena, KOROBOVA Maryna Parametric identification of asset dynamic models	117
MELNICHENKO Mykola, YAROVOL Leonid, PUCHKO Natalia, VODOTOVKA Mykhailo Method of experimental modeling of penetration of high-velocity impact elements in armor materials	122
ONYSHKEVYCH Volodymyr, BARABASH Galyna The unsteady contact interaction of a rotating stiff finite cylinder and elastic half-space with frictional heat generation	126
PASICHNYK Halyna Fundamental solution of the Cauchy problem for the dissipative ultraparabolic Kolmogorov equation with coefficients independent on variable degeneration	130
SAMOILENKO Oleh Mathematical problems of feature matching for vision-guided vehicles with limited resources	138
KHARYTONOV Oleksii, KUTSENKO Oleksii, Anastasiia KUTSENKO Analysis of the effectiveness of combining high-and low-thrust engines in Earth-Mars flights	142

CONTENTS

CHAIKOVSKIY Andrii, LIUBIMOV Oleksandr

Boundedness of solutions of the first order linear multidimensional difference equations146

SHAMRAI Maksym

Nonasymptotic bounds on return degradation for OBD-pruned neural controllers.....155

COMPUTER SCIENCES AND INFORMATICS

ZAVADSKYI Igor, KOVALCHUK Maksym

Optimization of the canonical Huffman codes for word-based natural language text compression159

LIPSKYI Danyil

Architecture and functional capabilities of a platform for automated web application testing166

MAKUSHENKO Igor, USAR Iryna, LIVINSKA Hanna

Hysteresis strategy for a single-channel system with a queue and repeated calls174

MARTYNIUK Sergiy, TSURKAN Viacheslav

Choosing the optimal strategy for a discrete dynamic cooperative game182

MAKHNO Mykhailo, FEDORUS Oleksii, BORYSENKO Oleksandr, VEREMCHUK Maksym

Architecture of a social media bot detection system187

PRAVOSUD Ruslan, MARCHENKO Oleksandr

Dialog style transfer with parameter-efficient fine-tuning193

SLABOSPITSKY Alexander

Least squares method of relative errors for parameter estimation of systems under non-classical assumptions197

STOYAN Volodymyr, CHERNIY Dmytro, VOLOSHCHUK Serhii, ZATORSKY Roman

Three-dimensional problems of controlling the dynamics of incompletely observed thick elastic plates.
Part II. The case of a discretely specified desired condition202

STOPKIN Andrii, TURKA Tetiana

Algorithm for recognizing simple graphs by a collective of agents211

TIURDO Ivan, KIZILOVA Natalya

The application of machine learning methods in modern cancer therapy217

MODERN PHYSICS

YAKUNOV Andrii, DIBROV Volodymyr, SACHKO Artem

Correlation between the fractal dimension of surface texture and its laser speckles as the basis
for a non-contact method for analyzing ceramic materials227

Вадим ПОНОМАРЬОВ¹, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0002-7813-3298
e-mail: vponomarov@knu.ua

Володимир МОРЕНЕЦЬ¹, канд. фіз.-мат. наук, асист.
ORCID ID: 0009-0006-0899-4400
e-mail: vmorenets@knu.ua

Ірина РОЗОРА¹, д-р фіз.-мат. наук, доц.
ORCID ID: 0000-0002-8733-7559
e-mail: irozora@knu.ua

Михайло ШАРАПОВ¹, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0002-8047-995X
e-mail: mmsharapov@knu.ua

Руслан ЧОРНЕЙ², канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0003-3866-8893
e-mail: r.chornei@ukma.edu.ua

Олександр СЛАБОСПИЦЬКИЙ¹, канд. фіз.-мат. наук, доц.
ORCID ID: 0009-0007-8753-2075
e-mail: alexsl@knu.ua

¹Київський національний університет імені Тараса Шевченка, Київ, Україна

²Національний університет "Києво-Могилянська академія", Київ, Україна

ПАВЛО КНОПОВ: НАУКОВИЙ ШЛЯХ І СТАНОВЛЕННЯ ШКОЛИ СТОХАСТИЧНОЇ ОПТИМІЗАЦІЇ

Присвячено 85-річчю Павла Соломоновича Кнопова (нар. 21.05.1940, Київ) – д-ра фіз.-мат. наук, чл.-кор. НАН України; фахівця зі стохастичної оптимізації, статистики випадкових процесів і полів, теорії ризику та керування; автора понад 250 праць і 13 монографій; від 1999 р. – завідувача відділу математичних методів дослідження операцій Інституту кібернетики ім. В. М. Глушкова НАН України, від 1987 р. – професора КНУ; лауреата Державної премії України (2009), премій ім. В. М. Глушкова (1997) та ім. В. С. Михалевича (2021), заслуженого діяча науки і техніки України (2021).

Київська школа стохастичної оптимізації, до розвитку якої істотно долучився П. С. Кнопов, є науковим феноменом, що поєднує теорію ймовірностей, математичну статистику та теорію керування для розв'язання задач ухвалення рішень в умовах невизначеності. У статті також висвітлено непростий життєвий і науковий шлях ученого, що пройшов крізь випробування епохи й утвердився як один із провідних математиків свого покоління.

Ключові слова: стохастична оптимізація, керування стохастичними системами, робастна статистика, М-оцінювання, стохастичні поля, великі відхилення, київська школа ймовірностей.

Класифікація відповідно до AMS 2020: 01A70, 90C15, 01A72, 93E20.

Вступ

Історія науки складається не лише з формул і теорем, а й із людських дол. Вона народжується у спогадах, у деталях біографій, у розмовах про те, як формувалося покоління науковців, що зуміло створити школу світового рівня. Київ другої половини ХХ ст. став одним із центрів теорії ймовірностей та стохастичної оптимізації. Шлях одного із представників цієї школи добре ілюструє, як поєднувалися складні особисті обставини, сила інтелекту й наукова спільнота, що підтримувала й виховувала нових дослідників.

1. Дитинство у воєнні роки

Майбутній математик народився у 1940 р. – у рік, коли над Україною пройшла хвиля війни. Його батько пішов на фронт і загинув у перших боях восени 1941 р. Дитинство від самого початку було позначене трагедією. Родина була евакуйована до Свердловської обл., де вони перебували аж до визволення Києва.

Повернення додому не принесло відчуття спокою: у повоєнному місті довелося ділити кімнату в комунальній квартирі з кількома родинами. Грошей бракувало, мати працювала, отримувала скромну зарплатню й невелику пенсію на сина. Саме ця суворя реальність, постійне відчуття скрути й водночас наполегливість матері виховали у хлопця характер – стриманий, але рішучий.

Попри труднощі, він ріс у середовищі, де навчання цінувалося. Спершу виховувався у дитячих садках Києва, а згодом став учнем школи № 21. Саме там уперше зустрів учительку Розу Зверюху, яка прищепила любов до математики. "Вона, мабуть, у певному сенсі визначила мій шлях", – згадував згодом учений. Родинні приклади теж були важливими: його родичка, яка ще до війни вступила на механіко-математичний факультет, після війни повернулася до навчання й підтримувала зв'язки з провідними математиками того часу.



У дитинстві не було надлишку книжок чи репетиторів, але була жива атмосфера пошуку й пізнання. У таких умовах майбутній науковець формував свій інтерес до точних наук, навіть не підозрюючи, що попереду на нього чекає ціла наукова школа і десятиліття досліджень.

2. Студентські роки

1957 рік став переломним: він вступив на механіко-математичний факультет Київського університету. Це був час, коли університет переживав піднесення. В аудиторіях панувала атмосфера інтелектуальної свободи, а серед викладачів були ті, чиї імена згодом увійшли в історію математики.

Викладачами в той час працювали Й. І. Гіхман, А. В. Скороход, В. С. Михалевич, В. С. Корольок, І. М. Коваленко, М. Й. Ядренко, А. Я. Дороговцев – постаті, які створили основу київської школи ймовірностей. Їхні лекції були чимось більшим, ніж звичний виклад матеріалу: це були зустрічі з людьми, які жили математикою, відкривали нові напрями, працювали пліч-о-пліч й із зарубіжними науковими школами.

Студентські роки були часом не лише навчання, а й формування світогляду. Спочатку П. С. Кнопов цікавився алгеброю й аналізом, але під впливом А. В. Скорохода й інших викладачів поступово занурився у світ теорії ймовірностей. Важливою стала участь у студентських наукових гуртках, де обговорювалися нові статті, задачі, де молодь навчалася мислити на рівні науки, а не лише підручника.

3. Шлях у науці

Після закінчення університету науковий шлях П. С. Кнопова був тісно пов'язаний з Інститутом кібернетики АН УРСР, нині Інститутом кібернетики ім. В. М. Глушкова НАН України, – одним із найпотужніших світових наукових центрів у галузі інформатики та прикладної математики. Саме тут розгорнулися дослідження у сфері статистики випадкових процесів і полів, негладкої та стохастичної оптимізації, оптимального керування та розпізнавання образів.

Кандидатську дисертацію Павло Соломонович захистив під керівництвом Анатолія Дороговцева. Робота стосувалася статистики випадкових процесів і полів, і вже тоді було зрозуміло, що цей напрям стане провідним у подальшій діяльності.



П. С. Кнопов (ліворуч) із колегами

У 1970–1980-х рр. київська школа активно працювала над теорією стохастичної оптимізації. Це була нова галузь, де поєднувалися ідеї теорії ймовірностей, математичної статистики та теорії оптимального керування. У Києві проходили масштабні конференції, що збирали провідних математиків з усього світу. Для молодого науковця це було не лише середовище роботи, а й школа життя: він бачив, як формувалися нові теорії, як народжувалася мова науки, зрозуміла і фахівцям, що займаються прикладною наукою, і теоретикам.

Серед тем, якими займався П. С. Кнопов, були:

- Статистика випадкових полів – інструмент для моделювання складних систем у фізиці й техніці.
- Оптимальне керування стохастичними системами – від дискретних моделей до неперервних процесів та полів.
- Методи стохастичної оптимізації – підхід, що згодом став одним із фундаментів сучасної теорії машинного навчання.
- Теорія великих відхилень – галузь, яка пояснює рідкісні події й розробляє математичні моделі та методи для визначення ефективних заходів мінімізації ризиків для об'єктів критичної інфраструктури, довкілля та продовольчої безпеки.
- Моделі теорії запасів, оптимального керування системами, масового обслуговування, забезпечення надійного функціонування складних технічних систем – прикладна математика у сфері економіки, техніки, логістики.
- Випадкові процеси із локальною взаємодією – моделі, що знайшли застосування у фізиці, економіці й теорії розпізнавання образів.

Його праці поєднували фундаментальність і прикладне застосування: вони формували математичний апарат, який можна було використовувати для управління ресурсами, прогнозування надійності систем, аналізу складних процесів.

4. Міжнародна співпраця

Розпад Радянського Союзу відкрив для українських учених нові горизонти. Якщо раніше виїзд за кордон був майже неможливим, то з початку 1990-х рр. почалася активна інтеграція у світову наукову спільноту. Саме тоді Павло Кнопов отримав змогу не лише долучатися до міжнародних досліджень, а й стати повноправним учасником і керівником великих проєктів.

Учений працював у грантових програмах *London Mathematical Society* (1996), *Royal Society* (1997), *DFG* (Німеччина, 1998–2010), співпрацював з університетом Больцано (Італія, 2000), реалізовував спільні проєкти між НАН України

та IASA (2007–2025). Одним із ключових партнерів став університет у Тронгеймі (Норвегія), де було здійснено багаторічну програму досліджень разом з українськими науковцями.

Тематика цих робіт була напрочуд широкою: від розроблення методів оцінювання ризиків для атомних електростанцій і моделювання надійності технічних систем – до прикладної статистики й оптимізації енергетичних процесів. Особливе місце займали проекти, пов'язані з ліквідацією наслідків Чорнобильської катастрофи: вони охоплювали оцінювання ризиків, оптимізацію технічного обслуговування, створення моделей безпеки.

Павло Кнопов виступав на багатьох міжнародних конференціях, серед яких: міжнародні конференції з теорії ймовірностей та математичної статистики (1973, 1985, 1989, 1994, 1998), Всесвітній конгрес математиків у Берліні (1998), Європейський конгрес математиків у Стокгольмі (2004), 16-й Міжнародний конгрес з математичного програмування в Лозанні (1997), Scandinavian – Ukraine Conference in Mathematical Statistics в Умео (1997), конференція з проблем ризику SAR and SRA – Europe Annual Conference у Лондоні (2000), міжнародні конференції зі стохастичного програмування в Колумбійському університеті (США, 1998) та в Києві (1984), International Conference on Stochastic Optimization: Algorithms and Applications у Флоридському університеті (2002), а також конференції зі стохастичного аналізу в Києві (2012, 2015).

Участь у міжнародних дослідженнях принесла не лише фінансову підтримку й сучасне обладнання, а й головне – відчуття приналежності до глобальної наукової спільноти. Це дало змогу українській школі стохастичної оптимізації не замкнутися у собі, а впевнено вийти на рівень, де її результати стали широко відомими у світі.

5. Викладання та школа учнів

Викладацька діяльність Павла Соломоновича посідала важливе місце в його житті. У Київському національному університеті імені Тараса Шевченка він викладав на кафедрі прикладної статистики, читав курси з математичної статистики, фінансової математики, теорії керованих стохастичних процесів.

За десятиліття викладання він виховав цілу школу учнів. Серед них – автори монографій із теорії ризику, оптимізації, статистики випадкових процесів. Частина з них створила власні наукові групи, працює у провідних університетах Європи та США.

Учні згадують, що його стиль викладання відзначався вимогливістю й водночас уважністю до кожного студента. Він умів зацікавити складними темами, показати їхній зв'язок із реальними задачами. Для нього навчання було не формальністю, а шляхом формування мислення.

6. Внесок у розвиток науки

Праці Павла Кнопова та сформованої ним наукової школи зробили вагомий внесок у світову математику й прикладну стохастичну. Київська школа теорії ймовірностей і стохастичної оптимізації, до розвитку якої він істотно долучився, здобула міжнародне визнання завдяки поєднанню високих теоретичних стандартів з увагою до реальних інженерних і соціально-економічних задач. Учні та співробітники школи отримували міжнародні відзнаки, а результати було опубліковано у монографіях і розділах провідних видавництв – *Kluwer, Springer, CRC, ISTE-Wiley*.

Дослідження Павла Соломоновича поєднують фундаментальність і практичну орієнтацію. Вони стали основою для розвитку нових сучасних напрямів математики та інформатики. Ключовою віссю його внеску є проблематика прийняття рішень у умовах невизначеності. Тут закладено базові постановки мінімізації ризику за неповної апіорної інформації, зокрема з використанням багатомодельних класів і баєсівських робастних підходів. Показано, як випадкові критерії можна перетворювати на керовані детерміновані функціонали ризику та працювати з імовірнісними обмеженнями (*chance constraints*).

Другим важливим напрямом стала робастна статистика й М-оцінювання для складних моделей – нелінійних, нестационарних та із залежними спостереженнями. Ці дослідження створили фундамент для методів, які нині застосовують у фінансовій математиці, машинному навчанні й інженерії надійності.

Третій напрям пов'язаний із керуванням стохастичними системами, що мають просторову структуру. Тут запропоновано моделі керованих марковських полів і стохастичних ігор на графах із розподіленими агентами та локальними взаємодіями. Було описано умови існування рівноваг і політик із доведеними властивостями стабільності, що стало новим словом у моделюванні багатокомпонентних систем.

Нарешті, важливою складовою є спадкоємність школи. Підготовлено чимало учнів, створено цикл монографій і навчальних матеріалів, що лягли в основу університетських і післядипломних курсів із робастної статистики, стохастичної оптимізації та керування. Завдяки цьому напрацьовані підходи не лише зберігаються, а й розвиваються у нових галузях – від фінансової інженерії до сучасних методів машинного навчання в умовах невизначеності.

7. Публікації

Павло Соломонович Кнопов є автором понад 260 наукових робіт, серед них 13 монографій (вид. Springer, Kluwer, Nova Science Publishers, CRC PRESS, ISTE-WILEY та ін.).

Вибрані монографії:

1. Kasitskaya, E. J., & Knopov, P. S. (2002). *Empirical Estimates in stochastic optimization and identification*. Kluwer.
2. Chornei, R. K., Daduna, H., & Knopov, P. S. (2006). *Control of Spatially Structured Random Processes and Random Fields with Applications*. Springer.
3. Knopov, P. S., & Pardalos, P. M. (2009). *Simulation and Optimization Methods in Risk and Reliability Theory*. Nova Science Publishers, Inc.
4. Knopov, P. S., & Korkhin, A. S. (2013). *Regression Analysis Under A Priory Parameter Restrictions*. Springer.
5. Knopov, P. S., & Derieva, O. N. (2013). *Estimation and Control problems for Stochastic Partial differential Equations*. Springer.
6. Gaivoronskii, A., Knopov, P., & Zaslavskii, V. (2023). *Modern Optimization Methods for Decision Making Under Risk and Uncertainty*. CRC PRESS, Taylor Francis Group, Boca Raton, London, New York.
7. Gaivoronskii, A., Knopov, P., Norkin, V., & Zaslavski, V. (2025). *Stochastic Modeling and Optimization Methods for Critical Infrastructure*. Protection 1 Stochastic Modeling, ISTE-WILEY.

8. Gaivoronskii, A., Knopov, P., Norkin, V., & Zaslavski, V. (2025). *Stochastic Modeling and Optimization Methods for Critical Infrastructure*. Protection 2 Methods and Tools, ISTE-WILEY.
9. Каниовский, М., Некрылова, З. В., & Кнопов, П. С. (1980). *Предельные теоремы для процессов стохастического программирования*. Наукова думка.
10. Кнопов, П. С. (1981). *Оптимальные оценки для стохастических систем*. Наукова думка.
11. Голодніков, О. Н., Кнопов, П. С., & Пепеляев, В. А. (2008). *Робастне оцінювання ризику*. НАНУ, Інститут кібернетики ім. В. М. Глушкова НАН України.
12. Кнопов, П. С., & Некрилова, З. В. (2008). *Методи моделювання та оцінювання в економетриці*. Київський національний університет імені Тараса Шевченка.
13. Кнопов, П. С., & Цикунов, І. К. (2009). *Методи комплексної оцінки та прогнозування стану соціально-економічної сфери України*. Інститут кібернетики. ім. В. М. Глушкова НАН України.

Вибрані статті:

1. Knopov, P. S., & Kasitskaya, E. J. (1995). Properties of empirical estimates in stochastic optimization and identification problems. *Annals of Operations Research*, 56(1), 225–239
2. Knopov, P. S. (1997). On some classes of M-estimates for non-stationary regression models. *Theory of Stochastic Processes*, 3(19), 222–228.
3. Knopov, P. S. (1999). Markov fields and their applications in economic. *Journal of Mathematical Sciences*, 3923–3931.
4. Golodnikov, A. N., Knopov, P. S., Pardalos, P., & Uryasev, S. (2000). Optimization in the space of Distribution Functions and Applications in the Bayes Analysis. In *Pardalos, P., Uryasev, S. (Eds.). Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, 102–131.
5. Knopov, P. S., Pardalos, P., & Yatsenko, V. A. (2001). Optimal Estimation of signal parameters using linear observation. *Optimization and related topics*. Kluwer Academic Publishers, 103–117.
6. Chornei, R., Daduna, H., & Knopov, P. (2004). Stochastic games for distributed players on graphs. *Mathematical Methods of Operations Research*, 60(2), 279–298
7. Chornei, R., Daduna, H., & Knopov, P. (2005). Controlled Markov fields with finite state space on graphs. *Stochastic Models*, 21(4), 847–874.
8. Kasitskaya, E. I., & Knopov, P. S. (2017). Large deviations for the method of empirical means in stochastic optimization problems with continuous time observations. *Springer Optimization and Its Applications*, 130, 263–276.
9. Knopov, P., & Norkin, V. (2022). Stochastic Optimization Methods for the Stochastic Storage Process Control. *Springer Optimization and Its Applications*, 181, 79–111
10. Knopov, P., & Korkhin, A. (2023). Dynamic models of epidemiology in discrete time taking into account processes with lag. *International Journal of Dynamics and Control*, 11(5), 2193–2214
11. Knopov, P. (2023). Optimization and Identification of Stochastic Systems. *Cybernetics and Systems Analysis*, 375–384.
12. Knopov, P. S., & Pepelyaeva, T. V. (2025). Some Applied Problems of the Theory of Controlled Random Processes. *Cybernetics and Systems Analysis*, 41–52.

Висновки і побажання студентам

У своїх побажаннях молодим дослідникам Павло Соломонович завжди підкреслює: математика залишається фундаментом сучасних технологій і науки. Попри матеріальні труднощі та зниження престижу академічної кар'єри, вона дає можливість вирішувати завдання, що змінюють світ.

"Незважаючи на всі труднощі сучасного життя, не варто забувати про математику та зберігати до неї інтерес. Якщо дослідження справді мають практичну цінність, вони знайдуть визнання і підтримку", – зазначає він.

Цей шлях – від дитинства у воєнному Києві до світового визнання – є свідченням того, що сила науки полягає не лише у формальних здобутках, а й у здатності долати перешкоди, зберігати інтерес до пізнання й передавати його наступним поколінням.



Внесок авторів: Вадим Пономарьов – концептуалізація, методологія, формальний аналіз, написання – оригінальний текст; Володимир Моренець – написання – рецензування та редагування, підготовка рукопису до подання; Ірина Розора – концептуалізація, рецензування; Михайло Шарапов – редагування; Руслан Чорней – підбір ілюстрацій і матеріалів; Олександр Слабоспицький – редагування та підбір матеріалів.

Джерела фінансування. Фінансування дослідження здійснюється власним коштом авторів та частково забезпечено Київським національним університетом імені Тараса Шевченка.

Отримано редакцією журналу / Received: 21.05.25
Прорецензовано / Revised: 23.09.25
Схвалено до друку / Accepted: 10.10.25

Vadym PONOMAROV¹, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-7813-3298
e-mail: vponomarov@knu.ua

Volodymyr MORENETS¹, PhD (Phys. & Math.), Assist.
ORCID ID: 0009-0006-0899-4400
e-mail: vmorenets@knu.ua

Iryna ROZORA¹, DSc (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-8733-7559
e-mail: irozora@knu.ua

Mykhailo SHARAPOV¹, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-8047-995X
e-mail: mmsharapov@knu.ua

Ruslan CHORNEI², PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0003-3866-8893
e-mail: r.chornei@ukma.edu.ua

Alexander SLABOSPITSKY¹, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0009-0007-8753-2075
e-mail: alexsl@knu.ua

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

²National University of Kyiv Mohyla Academy, Kyiv, Ukraine

PAVLO KNOPOV: THE SCIENTIFIC PATH AND THE FORMATION OF THE KYIV SCHOOL OF STOCHASTIC OPTIMIZATION

The article is dedicated to the 85th anniversary of Pavlo Solomonovych Knopov (b. May 21, 1940, Kyiv) – Doctor of Physical and Mathematical Sciences, Corresponding Member of the National Academy of Sciences of Ukraine; a leading specialist in stochastic optimization, statistics of random processes and fields, risk theory and control; author of over 250 scientific works and 13 monographs. Since 1999, he has headed the Department of Mathematical Methods of Operations Research at the V. M. Glushkov Institute of Cybernetics of the NAS of Ukraine, and since 1987 he has been a professor at Taras Shevchenko National University of Kyiv. He is a laureate of the State Prize of Ukraine in Science and Technology (2009), the V. M. Glushkov Prize (1997), the V. S. Mikhakevich (2021), and was awarded the title of Honored Scientist and Engineer of Ukraine (2021).

The Kyiv school of stochastic optimization, to the development of which P. S. Knopov made a significant contribution, represents a scientific phenomenon that combines probability theory, mathematical statistics, and control theory for solving decision-making problems under uncertainty. The article also outlines the difficult life and scientific path of the scholar, who went through the trials of his time and established himself as one of the leading mathematicians of his generation.

Key words: *stochastic optimization, control of stochastic systems, robust statistics, M-estimation, random fields, large deviations, Kyiv school of probability.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

АЛГЕБРА, ГЕОМЕТРІЯ ТА ТЕОРІЯ ЙМОВІРНОСТЕЙ

UDC 512.554.3

DOI: <https://doi.org/10.17721/1812-5409.2025/2.2>

Nataliia BONDARENKO, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0002-6078-9467

e-mail: natvbond@gmail.com

Kyiv National University of Construction and Architecture, Kyiv, Ukraine

Ievgen BONDARENKO, DSc (Phys. & Math.), Prof.

ORCID ID: 0000-0002-9491-211X

e-mail: ievgen.bondarenko@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

MATRIX REPRESENTATIONS OF ITERATED WREATH PRODUCTS OF ONE-DIMENSIONAL LIE ALGEBRAS

The wreath product is a classical construction in group theory, known for its ability to generate complex groups from simpler ones and its deep connections to permutation and automorphism groups. Its analogue in Lie algebra theory, however, remains less developed, despite several notable contributions by A. Shmelkin, F. Sullivan, V. Sushchansky, N. Bondarenko, V. Petrogradsky, and others. In particular, the Lie algebras $L_{p,n}$, associated via the Lazard correspondence with Sylow p -subgroups of symmetric groups $Sym(p^n)$, admit a decomposition as iterated wreath products of one-dimensional Lie algebras over the field $\mathbb{F}_p$, along with a tableau-based representation of elements. In this paper, we consider the Lie algebras $L_n := L_{2,n}$ and their inverse limit L_∞ , which is the infinitely iterated wreath product of one-dimensional Lie algebras over $\mathbb{F}_2$. We construct explicit embeddings of L_n into the Lie algebra $UT_m(\mathbb{F}_2)$ of strictly upper triangular matrices of minimal possible order $m = 2^{n-1} + 1$, and extend this to a recursive embedding of the Lie algebra L_∞ into the infinite-dimensional Lie algebra $UT_\infty(\mathbb{F}_2)$. The construction employs matrix schemes inspired by earlier representations of iterated wreath products of cyclic groups. These results provide a concrete realization of L_∞ and further deepen the connection between combinatorial Lie algebra structures and their matrix representations.

Key words: Lie algebra representation, wreath product, infinitely iterated wreath product, matrix scheme.

AMS 2020 classification: 17B10, 17B70, 17B40.

Introduction

The wreath product is a fundamental algebraic construction that combines two structures into a more complex one. It is particularly important in group theory, where it provides a method to build intricate groups from simpler components, preserving key structural features and playing a central role in the study of permutation groups, automorphism groups, and the construction of examples and counterexamples. In the context of Lie algebras, the wreath product has been less extensively explored, with notable contributions such as the verbal wreath product introduced in (Shmelkin, 1973), wreath products with abelian Lie algebras studied in (Sullivan, 1985; Sushchansky, & Netreba, 2005b; Bondarenko, 2006b), a general framework for wreath products of Lie algebras presented in (Petrogradsky, Razmyslov, & Shishkin, 2007), and a matrix wreath product of algebras introduced in (Alahmadi et al., 2017, 2019). This has further contributed to the development of self-similarity ideas in the theory of Lie algebras, see (Bartholdi, 2015; Petrogradsky, 2016; Futorny, Kochloukova, & Sidki, 2019).

One of the earliest applications of the wreath product is the description of the Sylow p -subgroup $S_{p,n}$ of the symmetric group $Sym(p^n)$ as the n -th iterated wreath product of the cyclic group C_p , see (Kaloujnine, 1945). This approach allows for representing group elements as special tableaux, which have been used to establish many properties of these groups. In particular, the representation of the group $S_{p,n}$ by unitriangular matrices is constructed in (Leonov, Nekrashevych, & Sushchansky, 2005) and extended to infinitely iterated wreath products of C_p in (Leonov, 2004, 2012).

An important connection between discrete groups and Lie algebras is provided by the construction of a graded Lie algebra associated with the lower central series of a group. This construction is especially effective for nilpotent groups and p -groups — a relation known as the Lazard correspondence. The group $S_{p,n}$ is a p -group, and we denote the associated graded Lie algebra by $L_{p,n}$. It is shown in (Sushchansky, & Netreba, 2005a, 2005b) that the algebra $L_{p,n}$ decomposes as the n -th iterated wreath product of one-dimensional Lie algebras over the field $\mathbb{F}_p$. In particular, elements of $L_{p,n}$ admit a special tableau representation, analogous to that of the group $S_{p,n}$. This representation was used to describe characteristic subalgebras, as well as the lower and upper central series of the Lie algebra $L_{p,n}$. Subsequently, iterated wreath products of n -dimensional abelian Lie algebras over the field $\mathbb{F}_p$ were studied using the same tableau-based approach in (Bondarenko, 2006b), and the Lie algebra associated with the group of finitary automorphisms of the p -adic tree was described in (Bondarenko, Gupta, & Sushchansky, 2010). Recently, the algebra $L_{p,n}$ has found applications in algebraic combinatorics related to integer partitions with a special role of the algebra $L_{2,n}$, see (Aragona, Civino, & Gavioli, 2024a, 2024b).

In this paper, we consider matrix representations of the Lie algebras $L_n := L_{2,n}$ and the infinite-dimensional Lie algebra L_∞ , which is the inverse limit of L_n , the infinitely iterated wreath product of one-dimensional Lie algebras over the field $\mathbb{F}_2$. We construct an explicit embedding of L_n into the Lie algebra $UT_m(\mathbb{F}_2)$ of strictly upper triangular matrices for the minimal possible order m . As a consequence, we obtain an explicit embedding of the Lie algebra L_∞ into the Lie algebra $UT_\infty(\mathbb{F}_2)$ of infinite strictly upper triangular matrices. Further, we show that this embedding can be constructed recursively using

matrix schemes, introduced in (Leonov, 2004). It is worth noting that an embedding of the Lie algebra $UT_m(\mathbb{F}_p)$ into the Lie algebra $L_{p,n}$ was previously constructed in (Bondarenko, 2006a).

1. The matrix representation of the n -th iterated wreath product of one-dimensional Lie algebras

Let $L_n = \wr_{i=1}^n A_1$ be the n -th iterated wreath product of the one-dimensional Lie algebras A_1 over the field $\mathbb{F}_2$. The elements of L_n are tableaux of the form:

$$u = [u_1, u_2(x_1), u_3(x_1, x_2), \dots, u_n(x_1, \dots, x_{n-1})],$$

where $u_1 \in \mathbb{F}_2$ and $u_i(x_1, \dots, x_{i-1}) \in \mathbb{F}_2[x_1, \dots, x_{i-1}] / \langle x_1^2, x_2^2, \dots, x_{i-1}^2 \rangle$. The elements of the quotient ring are uniquely represented by polynomials, called reduced polynomials, in which each variable x_i appears with exponent either zero or one. The k -th coordinate of the tableau u is denoted as u_k . The operations in L_n are defined as follows:

- 1) The sum of two elements u and v is given component-wise by $(u + v)_k = u_k + v_k$.
- 2) The multiplication of an element u by a scalar $\alpha \in \mathbb{F}_2$ is defined as $(\alpha u)_k = \alpha u_k$.
- 3) The Lie bracket operation of two elements u and v is defined by the rule

$$(u, v)_k = \sum_{i=1}^{k-1} \left(\frac{\partial v_k}{\partial x_i} u_i + v_i \frac{\partial u_k}{\partial x_i} \right).$$

An important concept in studying the properties of the Lie algebra L_n is the notion of height. The height of a reduced monomial $x_1^{k_1} x_2^{k_2} \dots x_{n-1}^{k_{n-1}}$, $k_i \in \{0, 1\}$, is the number

$$h(x_1^{k_1} x_2^{k_2} \dots x_{n-1}^{k_{n-1}}) = 1 + k_1 + k_2 \cdot 2 + k_3 \cdot 4 + \dots + k_{n-1} \cdot 2^{n-2}.$$

Note that for every height $i \in \{1, 2, \dots, 2^{n-1}\}$ there exists a unique monomial of height i , which we denote by $t(i)$. A few first monomials:

$$\begin{array}{llll} t(1) = 1, & t(2) = x_1, & t(3) = x_2, & t(4) = x_1 x_2, \\ t(5) = x_3, & t(6) = x_1 x_3, & t(7) = x_2 x_3, & t(8) = x_1 x_2 x_3. \end{array}$$

The height of a reduced polynomial is defined as the largest height among the heights of its monomials. Every reduced polynomial $u(x_1, x_2, \dots, x_{n-1})$ can be uniquely represented as a sum of monomials $t(i)$ for different values of $i \in \{1, 2, \dots, 2^{n-1}\}$.

Let us construct the matrix representation of L_n within the Lie algebra UT_m of strictly upper triangular matrices over the field $\mathbb{F}_2$. The Lie algebra L_n is nilpotent of class 2^{n-1} , while UT_m is nilpotent of class $m - 1$. Therefore, it is required that $m \geq 2^{n-1} + 1$, and we will demonstrate that this lower bound is achievable.

For a reduced polynomial $g(x_1, x_2, \dots, x_{n-1})$ and a tableau $u \in L_n$, define the polynomial

$$g^u = \sum_{k=1}^{n-1} \frac{\partial g}{\partial x_k} u_k.$$

It was shown in (Sushchansky, & Netreba, 2005b) that $h(g^u) \leq h(g) - 1$.

We define the mapping $\phi : L_n \rightarrow UT_{2^{n-1}+1}$ as follows. Given a tableau $u \in L_n$, consider the monomials $t(j)$ for $j \in \{1, 2, \dots, 2^{n-1} + 1\}$, apply u , and express the result as a linear combination of monomials:

$$t(j)^u = \sum_{k=1}^{n-1} \frac{\partial t(j)}{\partial x_k} u_k = \sum_{i=1}^{j-1} a_{ij} \cdot t(i), \quad a_{ij} \in \mathbb{F}_2.$$

The coefficients a_{ij} form an upper triangular matrix $A^u = (a_{ij})$ of order $2^{n-1} + 1$. We define $\phi(u) = A^u$.

Theorem 1. The mapping $\phi : L_n \rightarrow UT_{2^{n-1}+1}$, $\phi(u) = A^u$, is a monomorphism.

Proof. The mapping ϕ is linear by construction. We need to verify that ϕ preserves the Lie bracket:

$$\phi((u, v)) = [\phi(u), \phi(v)] \quad \text{for all } u, v \in L_n.$$

It is sufficient to verify the equality on the basis elements of L_n , which are tableaux where all coordinates are zero except for one, which contains the monomial corresponding to all possible heights at the respective coordinate.

Consider the element

$$u = [0, \dots, 0, t(l), 0, \dots, 0] \in L_n,$$

where the monomial $t(l)$ of the variables $x_1, x_2, \dots, x_{p-1}$ is located at position p , here $l \leq 2^{p-1}$. Let us describe the matrix $A = \phi(u)$. We compute:

$$t(j)^u = \frac{\partial t(j)}{\partial x_p} t(l) = t(i), \quad \text{where } i = h\left(\frac{\partial t(j)}{\partial x_p} t(l)\right) < j.$$

It follows that every row and every column of A contains at most one non-zero coefficient. The ij -coefficient of A is non-zero if and only if the variable x_p appears in the monomial $t(j)$, and the monomials $t(j)$ and $t(l)$ do not have common variables. Let $I_{p,l}$ be the set of heights of the monomials $t(j)$ satisfying these conditions. Then

$$A_{ij} = 1 \text{ if and only if } i = j + l - 2^{p-1} - 1 \text{ for every } j \in I_{p,l}.$$

Consider another element $v = [0, \dots, 0, t(m), 0, \dots, 0] \in L_n$, where the monomial $t(m)$ of the variables $x_1, x_2, \dots, x_{q-1}$ is located at position q . Let $B = \phi(v)$ and consider (u, v) and $[A, B] = AB - BA$.

1. If $p = q$, then $(u, v) = 0$. Consider the matrix AB . The ij -coordinate of AB is non-zero if and only if the following monomials are non-zero:

$$t(j)^u = \frac{\partial t(j)}{\partial x_p} t(l) = t(s) \text{ and } t(s)^v = \frac{\partial t(s)}{\partial x_p} t(m) = t(i).$$

Note, however, that the first equality implies that $t(s)$ does not contain x_p , and therefore the second equation is always equal to zero. It follows that $AB = 0$. Similarly, $BA = 0$ and $[A, B] = 0$.

2. Let $p < q$. The tableau (u, v) has the following coordinates:

$$(u, v)_k = 0 \text{ for } k \neq q \text{ and } (u, v)_q = \frac{\partial v_q}{\partial x_p} u_p = \frac{\partial t(m)}{\partial x_p} t(l).$$

Then the ij -coefficient of the matrix $C = \phi(w)$ for $w = (u, v)$ is non-zero if and only if the following equality holds:

$$t(i) = t(j)^w = \sum_{k=1}^{n-1} \frac{\partial t(j)}{\partial x_k} w_k = \frac{\partial t(j)}{\partial x_q} w_q = \frac{\partial t(j)}{\partial x_q} \cdot \frac{\partial t(m)}{\partial x_p} \cdot t(l).$$

Let us compare this with the matrix $[A, B] = AB - BA$. The ij -coordinate of AB is non-zero if and only if the following equalities hold:

$$\begin{aligned} t(j)^u &= \frac{\partial t(j)}{\partial x_p} t(l) = t(s), \\ t(i) &= t(s)^v = \frac{\partial t(s)}{\partial x_q} t(m) = \frac{\partial^2 t(j)}{\partial x_q \partial x_p} \cdot t(l) \cdot t(m) + \frac{\partial t(j)}{\partial x_p} \cdot \frac{\partial t(l)}{\partial x_q} \cdot t(m). \end{aligned}$$

Since $p < q$, the monomial $t(l)$ does not contain the variable x_p ; therefore, the last term in the above sum is equal to zero. Similarly, for the ij -coordinate of BA , we get:

$$\begin{aligned} t(j)^v &= \frac{\partial t(j)}{\partial x_q} t(m) = t(s), \\ t(i) &= t(s)^u = \frac{\partial t(s)}{\partial x_p} t(l) = \frac{\partial^2 t(j)}{\partial x_p \partial x_q} \cdot t(m) \cdot t(l) + \frac{\partial t(j)}{\partial x_q} \cdot \frac{\partial t(m)}{\partial x_p} \cdot t(l). \end{aligned}$$

Since we are working over the field $\mathbb{F}_2$, the ij -coordinate of $[A, B] = AB - BA$ is non-zero (equal to one) if and only if we have

$$t(i) = \frac{\partial t(j)}{\partial x_q} \cdot \frac{\partial t(m)}{\partial x_p} \cdot t(l),$$

which coincides with the condition for C . Therefore, $C = [A, B]$.

It remains to prove that ϕ is injective. Let us take a nonzero element $u \in L_n$, and let k be the smallest position such that $u_k \neq 0$. For the monomial $t(2^{k-1} + 1) = x_k$ we have $t(2^{k-1} + 1)^u = u_k$. Hence, for every monomial $t(i)$ appearing in u_k , the $(i, 2^{k-1} + 1)$ -coefficient of $\phi(u)$ is equal to one. This implies that the kernel of ϕ is trivial, and thus ϕ is a monomorphism. $\square$

2. The matrix representation of the infinitely iterated wreath product of one-dimensional Lie algebras

The sequence of Lie algebras L_n naturally forms an inverse system,

$$\dots \rightarrow L_3 \rightarrow L_2 \rightarrow L_1,$$

where the transition map $\pi_n : L_n \rightarrow L_{n-1}$ is defined by restricting tableaux with n components to their first $n-1$ components. The inverse limit of this system is called the *infinitely iterated wreath product* $L_\infty = \varprojlim L_n = \varprojlim_{i=1}^n A_1$ of the one-dimensional Lie algebras A_1 over the field $\mathbb{F}_2$. The elements of L_∞ are infinite tableaux of the form:

$$u = [u_1, u_2(x_1), u_3(x_1, x_2), \dots, u_n(x_1, \dots, x_{n-1}), \dots],$$

where $u_1 \in \mathbb{F}_2$ and $u_i(x_1, \dots, x_{i-1}) \in \mathbb{F}_2[x_1, \dots, x_{i-1}] / \langle x_1^2, x_2^2, \dots, x_{i-1}^2 \rangle$, $i \geq 2$.

The representations $\phi_n : L_n \rightarrow UT_{2^{n-1}+1}$ constructed in the previous section are compatible with the inverse system of L_n in the following sense. The Lie algebras UT_m form an inverse system, where the transition map $\tau_m : UT_m \rightarrow UT_{m-1}$ is given by restricting a matrix to its $(m-1) \times (m-1)$ submatrix. The inverse limit is the Lie algebra UT_∞ of infinite strictly

upper triangular matrices over the field $\mathbb{F}_2$. The definition of ϕ_n implies that the following diagram is commutative:

$$\begin{array}{ccccccc} \dots & \xrightarrow{\pi_5} & L_4 & \xrightarrow{\pi_4} & L_3 & \xrightarrow{\pi_3} & L_2 & \xrightarrow{\pi_2} & L_1 \\ & & \downarrow \phi_4 & & \downarrow \phi_3 & & \downarrow \phi_2 & & \downarrow \phi_1 \\ \dots & \xrightarrow{\tau_{17}} & UT_9 & \xrightarrow{\tau_9} & UT_5 & \xrightarrow{\tau_5} & UT_3 & \xrightarrow{\tau_3} & UT_1 \end{array}$$

Passing to the limit, we get a monomorphism $\phi : L_\infty \rightarrow UT_\infty$. Explicitly, given a tableau $u \in L_\infty$ and a column index $j \geq 1$, compute $t(j)^u$ and express it as a linear combination of monomials:

$$t(j)^u = \sum_{k=1}^{\infty} \frac{\partial t(j)}{\partial x_k} u_k = \sum_{i=1}^{j-1} a_{ij} \cdot t(i), \quad a_{ij} \in \mathbb{F}_2.$$

Then, we have

$$\phi(u) = A, \quad \text{where } A = (a_{i,j})_{i,j \geq 1} \in UT_\infty.$$

We will show that the representation ϕ could be described recursively (or self-similarly) using the notion of matrix schemes introduced in (Leonov, 2004; Leonov, Nekrashevych, & Sushchansky, 2005). Let M_∞ be the set of infinite matrices over the field $\mathbb{F}_2$. For a matrix $A = (a_{ij})_{i,j \geq 1} \in M_\infty$, we define four submatrices $A_{11}, A_{12}, A_{21}, A_{22}$ as follows:

$$A_{i_0 j_0} = (b_{i,j}), \quad \text{where } b_{i,j} = a_{i_0+2(i-1), j_0+2(j-1)}, \quad i_0, j_0 \in \{1, 2\},$$

$$A_{i_0 j_0} = \begin{pmatrix} a_{i_0, j_0} & a_{i_0, j_0+2} & a_{i_0, j_0+4} & \dots \\ a_{i_0+2, j_0} & a_{i_0+2, j_0+2} & a_{i_0+2, j_0+4} & \dots \\ a_{i_0+4, j_0} & a_{i_0+4, j_0+2} & a_{i_0+4, j_0+4} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Definition 1. The (parity) *scheme* of a matrix $A \in M_\infty$ is the 2×2 matrix, whose coefficients are matrices:

$$sh(A) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Example. Let us construct the scheme of the matrix $A = \phi(u)$ for the table $u = [1, 0, 0, \dots] \in L_\infty$. The ij -coefficient of A is determined by the equation:

$$t(j)^u = \sum_{k=1}^{\infty} \frac{\partial t(j)}{\partial x_k} u_k = \frac{\partial t(j)}{\partial x_1} = t(i).$$

We have $a_{ij} = 1$ if and only if $t(j)$ contains the variable x_1 , which occurs when j is even. In this case, $i = j - 1$ and $a_{j-1, j} = 1$. Consequently, we get the matrix and its scheme:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad \text{and} \quad sh(A) = \begin{pmatrix} 0 & E \\ 0 & 0 \end{pmatrix}.$$

Definition 2. For a tableau $u = [u_1, u_2(x_1), u_3(x_1, x_2), \dots] \in L_\infty$, define two tableaux $u|_0$ and $u|_1$ as follows:

$$u|_c = [u'_1, u'_2(x_1), u'_3(x_1, x_2), \dots] \in L_\infty, \quad c = 0, 1,$$

where $u'_k(x_1, \dots, x_{k-1})$ is obtained from $u_{k+1}(x_1, \dots, x_k)$ by applying the substitution:

$$x_1 \mapsto c, \quad x_2 \mapsto x_1, \quad x_3 \mapsto x_2, \quad \dots, \quad x_k \mapsto x_{k-1}. \quad (1)$$

Theorem 2. The mapping $\phi : L_\infty \rightarrow UT_\infty$ satisfies the recursion:

$$sh(\phi(u)) = \begin{pmatrix} \phi(u|_0) & u_1 E \\ \phi(u|_0) + \phi(u|_1) & \phi(u|_0) \end{pmatrix}, \quad u \in L_\infty.$$

Proof. Let $A = \phi(u) = (a_{i,j})_{i,j \geq 1}$ and $B = \phi(u|_0) = (b_{i,j})_{i,j \geq 1}$. We need to show that $A_{11} = B$, which is equivalent to the condition $a_{1+2i, 1+2j} = b_{1+i, 1+j}$ for all $i, j \geq 0$. The coefficients of A and B comes from the equations:

$$t(1+2j)^u = \sum_{k=1}^{\infty} \frac{\partial t(1+2j)}{\partial x_k} u_k = \sum_{i=1}^{2j} a_{i, 1+2j} \cdot t(i),$$

$$t(1+j)^{u|_0} = \sum_{k=1}^{\infty} \frac{\partial t(1+j)}{\partial x_k} (u|_0)_k = \sum_{i=1}^j b_{i, 1+j} \cdot t(i).$$

Note that a monomial has odd height if and only if it does not contain the variable x_1 . Consequently, monomials of even height appearing in u_k do not contribute to the coefficients $a_{i,1+2j}$ for odd i . The key observation we need is that, the construction of $(u|_0)_k$ from u_{k+1} via the substitution (1) implies a bijective correspondence between the monomials $t(1+2l)$ appearing in u_{k+1} and the monomials $t(1+l)$ appearing in $(u|_0)_k$. The monomial $t(1+2l)$ of u_{k+1} contributes to the coefficient $a_{1+2i,1+2j}$, where

$$1+2i = h\left(\frac{\partial t(1+2j)}{\partial x_{k+1}} t(1+2l)\right) = 1+2j+2l-2^k \Rightarrow i = j+l-2^{k-1}.$$

The monomial $t(1+l)$ of $(u|_0)_k$ contributes to the coefficient $b_{1+i,1+j}$, where

$$1+i = h\left(\frac{\partial t(1+j)}{\partial x_k} t(1+l)\right) = 1+j+l-2^{k-1}.$$

This produces the same value of i as above, implying $a_{1+2i,1+2j} = b_{1+i,1+j}$, and hence $A_{11} = B$. The proof that $A_{22} = B$ follows in a similar manner.

Let us prove that $A_{12} = u_1 E$. For an even j , the monomial $t(j)$ contains the variable x_1 . It follows that every monomial appearing in the last sum of the equation

$$t(2+2j)^u = \sum_{k=1}^{\infty} \frac{\partial t(2+2j)}{\partial x_k} u_k = \frac{\partial t(2+2j)}{\partial x_1} u_1 + \sum_{k>1} \frac{\partial t(2+2j)}{\partial x_k} u_k,$$

has even height and does not contribute to A_{12} . Consequently, if $u_1 = 0$, $A_{12} = 0$, and if $u_1 = 1$, then $A_{12} = E$.

Let $C = \phi(u|_1) = (c_{i,j})_{i,j \geq 1}$. We are left to show that $A_{21} = B + C$, i.e.,

$$a_{2+2i,1+2j} = b_{1+i,1+j} + c_{1+i,1+j} \text{ for all } i, j \geq 0.$$

Only monomials of even heights in u_k contribute to the coefficient $a_{i,1+2j}$ for an even i . The k -th component of $u|_1$ is constructed from u_{k+1} via the substitution (1). By construction, every monomial $t(2+2l)$ appearing in u_{k+1} produces the monomial $t(1+l)$ in $(u|_1)_k$, but also every monomial $t(1+2l)$ in u_{k+1} produces the monomial $t(1+l)$ in $(u|_1)_k$. The monomial $t(2+2l)$ of u_{k+1} contributes to the coefficient $a_{2+2i,1+2j}$, where

$$2+2i = h\left(\frac{\partial t(1+2j)}{\partial x_{k+1}} t(2+2l)\right) = 1+2j+1+2l-2^k \Rightarrow i = j+l-2^{k-1}.$$

The monomial $t(1+l)$ of $(u|_1)_k$ contributes to the coefficient $c_{1+i,1+j}$, where

$$1+i = h\left(\frac{\partial t(1+j)}{\partial x_k} t(1+l)\right) = 1+j+l-2^{k-1}.$$

Note that this produces the same value of i as above.

A possible monomial $t(1+l)$ in $(u|_1)_k$ coming from $t(1+2l)$ in u_{k+1} also contributes to the coefficient $c_{1+i,1+j}$, while $t(1+2l)$ does not contribute to $a_{2+2i,1+2j}$. Note that $t(1+2l)$ in u_{k+1} produces $t(1+l)$ in $(u|_0)_k$. Therefore, this possible additional term in $t(2+2j)^{u|_1}$ has a corresponding term coming from $t(1+l)$ appearing in $(u|_0)_k$ that contributes to the coefficient $b_{1+i,1+j}$. Consequently, $a_{2+2i,1+2j} = b_{1+i,1+j} + c_{1+i,1+j}$, and hence $A_{21} = B + C$. $\square$

Example. Consider the tableau $u = [1, 1, 1, \dots] \in L_{\infty}$. Then

$$u|_0 = u|_1 = [1, 1, 1, \dots] = u.$$

Therefore, we obtain the following scheme recursion for $A = \phi(u)$:

$$sh(A) = \begin{pmatrix} A & E \\ A+A & A \end{pmatrix} = \begin{pmatrix} A & E \\ 0 & A \end{pmatrix}.$$

By expanding this recursion, we obtain the matrix:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Example. Consider the tableau $u = [1, x_1, x_2, \dots] \in L_\infty$. We compute:

$$\begin{aligned} v := u|_0 &= [0, x_1, x_2, x_3, \dots], & u|_1 &= [1, x_1, x_2, x_3, \dots] = u, \\ v|_0 &= [0, x_1, x_2, x_3, \dots] = v, & v|_1 &= [1, x_1, x_2, x_3, \dots] = u. \end{aligned}$$

Therefore, we obtain the following scheme recursion for $A = \phi(u)$ and $B = \phi(v)$:

$$sh(A) = \begin{pmatrix} B & E \\ A+B & B \end{pmatrix}, \quad sh(B) = \begin{pmatrix} B & 0 \\ A+B & B \end{pmatrix}.$$

We can simplify the recursion using:

$$sh(A+B) = sh(A) + sh(B) = \begin{pmatrix} B+B & E \\ A+B+A+B & B+B \end{pmatrix} = \begin{pmatrix} 0 & E \\ 0 & 0 \end{pmatrix}.$$

By expanding this recursion, we obtain the matrices:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Example. Consider the tableau $u = [1, x_1, x_1x_2, x_1x_2x_3, \dots] \in L_\infty$. We compute:

$$u|_0 = [0, 0, 0, \dots], \quad u|_1 = [1, x_1, x_1x_2, x_1x_2x_3, \dots] = u.$$

Therefore, we obtain the following scheme recursion for $A = \phi(u)$:

$$sh(A) = \begin{pmatrix} 0 & E \\ A & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

3. Discussion and conclusions

We have constructed explicit matrix representations of the Lie algebras L_n , defined as the n -th iterated wreath products of one-dimensional Lie algebras over the field $\mathbb{F}_2$, as embeddings into the Lie algebra $UT_m(\mathbb{F}_2)$ of strictly upper triangular matrices with minimal possible order $m = 2^{n-1} + 1$. Furthermore, we have extended this construction to a recursive embedding of their inverse limit L_∞ , the infinitely iterated wreath product, into the infinite-dimensional Lie algebra $UT_\infty(\mathbb{F}_2)$, using matrix schemes inspired by similar constructions for iterated wreath products of groups. The recursive nature of the embeddings may provide insights into automorphism groups, growth functions, and homological invariants of these algebras, while also suggesting potential applications in representation theory, combinatorics, and the theory of self-similar Lie algebras. In future work, it would be natural to explore the representations of iterated wreath products of higher-dimensional abelian Lie algebras over $\mathbb{F}_p$, their cohomological dimensions, and potential applications to p -adic and profinite Lie theory.

Authors' contribution: Nataliia Bondarenko — problem formulation, conducting research, literature review, and writing the original draft; levgen Bondarenko — conducting research, writing, and editing of the manuscript.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Alahmadi, A., Alsulami, H., Jain, S. K., & Zelmanov, E. (2017). On matrix wreath products of algebras. *Electronic Research Announcements in Mathematical Sciences*, 24, 78–86. <https://doi.org/10.3934/era.2017.24.009>
- Alahmadi, A., Alsulami, H., Jain, S. K., & Zelmanov, E. (2019). Matrix wreath products of algebras and embedding theorems. *Transactions of the American Mathematical Society*, 372(4), 2389–2406. <https://doi.org/10.1090/tran/7642>
- Aragona, R., Civino, R., & Gavioli, N. (2024a). A modular idealizer chain and unrefinability of partitions with repeated parts. *Israel Journal of Mathematics*, 260(1), 441–461. <https://doi.org/10.1007/s11856-023-2559-8>
- Aragona, R., Civino, R., & Gavioli, N. (2024b). An ultimately periodic chain in the integral Lie ring of partitions. *Journal of Algebraic Combinatorics*, 59(4), 939–954. <https://doi.org/10.1007/s10801-024-01318-x>
- Bartholdi, L. (2015). Self-similar Lie algebras. *Journal of the European Mathematical Society*, 17(12), 3113–3151. <https://doi.org/10.4171/JEMS/581>
- Bondarenko, N. V. (2006a). Lie algebras associated with Sylow p -subgroups of some classical linear groups. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics and Mathematics*, 2, 21–27 [in Ukrainian]. [Бондаренко, Н. В. (2006). Алгебри Лі,

- асоційовані з силовськими p -підгрупами деяких класичних лінійних груп. *Вісник Київського національного університету імені Тараса Шевченка. Фізико-математичні науки*, 2, 21–27].
- Bondarenko, N. V. (2006b). Lie algebras associated with wreath products of elementary abelian groups. *Matematychni Studii*, 26(1), 3–16. http://matstud.org.ua/texts/2006/26_1/26_1_003_016.pdf
- Bondarenko, N. V., Gupta, C. K., & Sushchansky, V. I. (2010). Lie algebra associated with the group of finitary automorphisms of p -adic tree. *Journal of Algebra*, 324(9), 2198–2218. <https://doi.org/10.1016/j.jalgebra.2010.07.042>
- Futorny, V., Kochloukova, D. H., & Sidki, S. N. (2019). On self-similar Lie algebras and virtual endomorphisms. *Mathematische Zeitschrift*, 292(3–4), 1123–1156. <https://doi.org/10.1007/s00209-018-2146-6>
- Kaloujnine, L. (1945). On the sylow p -subgroups of the symmetric group of degree p^m . *Weekly Proceedings of the Sessions of the Academy of Sciences*, 221, 222–224 [in French]. [Kaloujnine, L. (1945). Sur les p -groupes de Sylow du groupe symétrique du degré p^m . *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 221, 222–224].
- Leonov, Y. G. (2004). Representation of residually finite 2-groups by infinite-dimensional unitriangular matrices over a field of two elements. *Matematychni Studii*, 22(2), 134–140 [in Russian]. [Леонов, Ю.Г. (2004). Представление финитно-аппроксимируемых 2-групп бесконечномерными унитарными матрицами над полем из двух элементов. *Математичні студії*, 22(2), 134–140].
- Leonov, Y. G. (2012). On the representation of groups approximated by finite p -groups. *Ukrainian Mathematical Journal*, 63(11), 1706–1718 [in Russian]. [Леонов, Ю.Г. (2012). О представлении групп, аппроксимируемых конечными p -группами. *Український математичний журнал*, 22(11), 1706–1718].
- Leonov, Y. G., Nekrashevych, V. V., & Sushchansky, V. I. (2005). Representations of wreath products by unitriangular matrices. *Reports of the National Academy of Sciences of Ukraine*, 4, 29–33 [in Ukrainian]. [Леонов, Ю., Некрашевич, В., & Суцанський, В. (2005). Зображення вінцевих добутоків унітрикутними матрицями. *Доповіді Національної академії наук України*, 4, 29–33].
- Petrogradsky, V. M. (2016). Fractal nil graded Lie superalgebras. *Journal of Algebra*, 466, 229–283. <https://doi.org/10.1016/j.jalgebra.2016.06.028>
- Petrogradsky, V. M., Razmyslov, Y. P., & Shishkin, E. O. (2007). Wreath products and Kaluzhnin-Krasner embedding for Lie algebras. *Proceedings of the American Mathematical Society*, 135(3), 625–636. <https://doi.org/10.1090/S0002-9939-06-08502-9>
- Shmelkin, A. L. (1973). Wreath products of Lie algebras, and their application in group theory. *Proceedings of the Moscow Mathematical Society*, 29, 247–260 [in Russian]. [Шмелькин, А. Л. (1973). Сплетения алгебр Ли и их применение в теории групп. *Труды Московского математического общества*, 29, 247–260].
- Sullivan, F. E. (1985). Wreath products of Lie algebras. *Journal of Pure and Applied Algebra*, 35(1), 95–104. [https://doi.org/10.1016/0022-4049\(85\)90032-5](https://doi.org/10.1016/0022-4049(85)90032-5)
- Sushchansky, V., & Netreba, N. (2005a). Lie algebras associated with Sylow p -subgroups of finite symmetric groups. *Matematychni Studii*, 24(2), 127–138 [in Russian]. [Суцанський, В., & Нетреба, Н. (2005). Алгебры Ли, ассоциированные с силовскими p -подгруппами конечных симметрических групп. *Математичні студії*, 24(2), 127–138].
- Sushchansky, V., & Netreba, N. (2005b). Wreath product of Lie algebras and Lie algebras associated with Sylow p -subgroups of finite symmetric groups. *Algebra and Discrete Mathematics*, 1, 122–132.

Отримано редакцією журналу / Received: 09.04.25

Прорецензовано / Revised: 24.06.25

Схвалено до друку / Accepted: 10.10.25

Наталія БОНДАРЕНКО, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0002-6078-9467

e-mail: natvbond@gmail.com

Київський національний університет будівництва і архітектури, Київ, Україна

Євген БОНДАРЕНКО, д-р фіз.-мат. наук, проф.

ORCID ID: 0000-0002-9491-211X

e-mail: levgen.bondarenko@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

МАТРИЧНІ ЗОБРАЖЕННЯ ІТЕРОВаних ВІНЦЕВИХ ДОБУТКІВ ОДНОВИМІРНИХ АЛГЕБР ЛІ

Вінцевий добуток є класичною конструкцією в теорії груп, відомою своєю здатністю утворювати складні групи з простіших, а також глибокими зв'язками з групами підстановок та автоморфізмів. Його аналог у теорії алгебр Лі, однак, залишається менш дослідженим, незважаючи на низку вагомих внесків, зроблених А. Шмелькіним, Ф. Салліваном, В. Суцанським, Н. Бондаренко, В. Петроградським та ін. Зокрема, алгебри Лі $L_{p,n}$, що пов'язані із силовськими p -підгрупами симетричних груп $Sym(p^n)$ через відповідність Лазара, допускають декомпозицію як ітеровані вінцеві добуток одновимірних алгебр Лі над полем $\mathbb{F}_p$, а також мають табличне зображення елементів. У цій роботі ми розглядаємо алгебри Лі $L_n := L_{2,n}$ та їхню зворотню границю L_∞ , що є нескінченно ітерованим вінцевим добутком одновимірних алгебр Лі над полем $\mathbb{F}_2$. Ми будемо явні занурення алгебри L_n в алгебру Лі строго верхньотрикутних матриць $UT_m(\mathbb{F}_2)$ мінімально можливого порядку $m = 2^{n-1} + 1$ та поширюємо це на рекурсивне занурення алгебри Лі L_∞ в нескінченновимірну алгебру Лі $UT_\infty(\mathbb{F}_2)$. Конструкція використовує матричні шаблони, побудовані на основі попередніх зображень ітерованих вінцевих добутоків циклічних груп. Ці результати забезпечують конкретну реалізацію L_∞ і ще більше поглиблюють зв'язок між комбінаторною структурою алгебр Лі та їх матричними зображеннями.

Ключові слова: зображення алгебри Лі, вінцевий добуток, нескінченно ітерований вінцевий добуток, шаблон матриці.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 519.21

DOI: <https://doi.org/10.17721/1812-5409.2025/2.3>

Yevhenii BUNDIUK, Student
ORCID ID: 0009-0006-2130-6705

e-mail: minakami1984@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Yuliya MISHURA, DSc (Phys. & Math.), Prof.

ORCID ID: 0000-0002-6877-1800

e-mail: yuliyamishura@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Enrica PIROZZI, Prof.

ORCID ID: 0000-0001-8692-2759

e-mail: enrica.pirozzi@unicampania.it
University of Campania Luigi Vanvitelli, Caserta, Italy

PREDICTION FORMULAS FOR INTEGRATED FRACTIONAL BROWNIAN MOTION

This paper investigates the projection (prediction) of increments of the integrated fractional Brownian motion (ifBm). We introduce ifBm, calculate its covariance function, and establish the stationarity of its increments. Our primary goal is to determine the coefficients for the linear prediction of a future ifBm increment based on a series of past increments. We show that the prediction coefficients exhibit complex behavior, with their signs changing depending on the Hurst index (H). For overlapping increments, this sign change occurs at a non-trivial values of H . In contrast, for non-overlapping increments, the sign change happens at $H=0.5$, consistent with the properties of fractional Brownian motion itself. This work provides explicit formulas where possible and extensive numerical tables, offering insights into the properties and predictability of ifBm, laying the groundwork for further research.

Key words: fractional Brownian motion, integrated fractional Brownian motion, covariance function, projection problem, projection coefficients.

AMS 2020 classification: 60G15, 60G22.

Introduction

Fractional Brownian motion (fBm) is one of the most popular processes in recent decades. The reason is that it is a stochastic process that has memory, and this property is inherent in real processes in physics, biology, electronics, economics, and finance. Investigation of stochastic representations and the main properties of fBm started with the works (Mandelbrot, & Van Ness, 1968; Barton, & Vincent Poor, 1988; Cioczek-Georges, & Mandelbrot, 1995; Dai, & Heyde, 1996; Decreusefond, & Üstünel, 1998; Carmona, & Coutin, 1998; Coutin, & Decreusefond, 1999; Decreusefond, & Üstünel, 1999; Norros, Valkeila, & Virtamo, 1999; Novikov, & Valkeila, 1999). Then, stochastic analysis of fBm and related processes, together with the main applications, was developed in the papers (Alòs, Mazet, & Nualart, 2001; Coutin, Nualart, & Tudor, 2001; Duncan, 2001; Duncan, Pasik-Duncan, & Maslowski, 2002; Biagini et al., 2002; Alòs, & Nualart, 2003; Bender, 2003b, 2003a; Benth, 2003; Boufoussi, & Ouknine, 2003; Nualart, 2003; Nualart, & Ouknine, 2003; Anh, & Inoue, 2004; Carmona, Coutin, & Montseny, 2003; Cheridito, 2001; Cheridito, Kawaguchi, & Maejima, 2003; Bender, & Elliott, 2004; Cheridito, & Nualart, 2005; Corcuera, Nualart, & Woerner, 2006; Bender, Sottinen, & Valkeila, 2008). Let us mention also expansions and transformations of fBm contained in (Dzhaparidze, & van Zanten, 2004, 2005; Jost, 2006). Of course, this list is not exhaustive. The basic properties of fractional Brownian motion have been summarized in many books and papers, and without claiming to be complete, we will mention just the books (Banna et al., 2019; Biagini et al., 2008; Beran, 1994; Nourdin, 2012; Pipiras, & Taqqu, 2017; Samorodnitsky, 2016). Although a fairly convenient and flexible model, fractional Brownian motion still has the disadvantage that fractional exponents in the covariance function make the covariance matrix very inconvenient for analytical research, calculations of the determinant, inversion and other operations. Often, problems related to calculating the determinant do not have an analytical solution at all, but only a numerical one. This point, the accompanying conditions and possible actions were discussed in detail in the papers (Malyarenko et al., 2023; Mishura, Ralchenko, & Shklyar, 2020). In particular, in paper (Mishura, Ralchenko, & Shklyar, 2020), considering applications of fractional Brownian motion to financial markets, we hypothesized that for values of the Hurst index $H \in (\frac{1}{2}, 1)$, all coefficients of projection of the increment of fBm on the subsequent increments are strictly positive. Because of the overly complex form of the determinants associated with the covariance matrix of the increments of fBm that arose when solving this problem, we were able to analytically verify our hypothesis only for a limited number of increments, and we also considered numerical examples. Taking this into account, in paper (Bodnarchuk, & Mishura, 2024) we considered, instead of fBm, an integrated Wiener process, which preserves memory but has not a complicated covariance matrix. In this case we could completely resolve the problem of projection, but the result was unexpected: in this case the signs of the coefficients alternate. Integrated fractional Brownian motion is a natural generalization of an integrated Wiener process. It is natural to average very big datasets in the periods of time. It has also proven its value, in particular, in modeling first-passage time problems in various fields such as computational neuroscience, finance, and queueing theory. Specifically, in case of computational neuroscience, it is used to model the time variability and firing rate adaptation of neurons. (see, e.g., the discussion in (Abundo, & Pirozzi, 2018, 2019)). Now, given the value of studying such integrated processes, we decided in the present paper to consider the problem of projection (prediction) of the increments of the integrated fractional Brownian

motion. Why are the problems of projection (roughly speaking, to find $\mathbb{E}[\Delta I_0 \mid \Delta I_1, \dots, \Delta I_n]$, see Section 3) and prediction (roughly speaking, to find $\mathbb{E}[\Delta I_n \mid \Delta I_0, \dots, \Delta I_{n-1}]$) equivalent? Just because we consider stationary sequences, for which the corresponding coefficients are simply arranged in mutually inverse order. The paper is organized as follows: in Section 1 we give preliminaries on fBm; in Section 2 we introduce integrated fBm, calculate its covariance function and establish its properties; in Section 3 we calculate the coefficients of projection $\mathbb{E}[\Delta I_0 \mid \Delta I_1, \dots, \Delta I_n]$, $n \geq 1$ with overlapping intervals analytically for $n = 1$, partially analytically for $n = 2$ and provide numerical results for $2 \leq n \leq 10$; in Section 4 we do the same for the non-overlapping intervals.

1. Preliminaries

A *fractional Brownian motion* (fBm) with Hurst index $H \in (0, 1)$ is a centered Gaussian process $B^H = \{B_t^H\}_{t \geq 0}$ satisfying the following properties:

- $B_0^H = 0$ almost surely;
- For every $t, s \geq 0$, the covariance function is given by

$$\mathbb{E}[B_t^H B_s^H] = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H});$$

- B^H has stationary increments;
- the trajectories of B^H are almost surely locally Hölder continuous of order α for every $\alpha \in (0, H)$, i.e., for every compact interval $[0, T]$, there exists a random variable $C = C(T, \omega) > 0$ such that

$$|B_t^H - B_s^H| \leq C|t - s|^\alpha, \quad \text{for all } s, t \in [0, T] \quad \text{a.s.}$$

The last property, in particular, implies that for any $0 \leq a < b < \infty$ the integral $\int_a^b B_s^H ds$ is well-defined in the classical Riemann sense and is a zero-mean Gaussian random variable.

2. Integrated Fractional Brownian Motion

In this section, we consider the increments of the integral of a fractional Brownian motion, more precisely, we consider the integrals $I_k = \int_k^{k+1} B_s^H ds$, where B^H denotes fBm with Hurst index $H \in (0, 1)$, and $k \geq 0$ is an integer.

Lemma 1. *The covariance function of I_k is given by the formula*

$$\mathbb{E}[I_k I_l] = \frac{1}{2} (I_{1,k,l} - I_{2,k,l}),$$

where

$$I_{1,k,l} = \frac{1}{2H+1} [(l+1)^{2H+1} - l^{2H+1} + (k+1)^{2H+1} - k^{2H+1}],$$

$$I_{2,k,l} = \frac{||k-l|+1|^{2H+2} - 2|k-l|^{2H+2} + ||k-l|-1|^{2H+2}}{(2H+1)(2H+2)}.$$

Proof. Indeed,

$$\begin{aligned} \mathbb{E}[I_k I_l] &= \mathbb{E} \left[\int_k^{k+1} B_s^H ds \int_l^{l+1} B_t^H dt \right] = \int_k^{k+1} \int_l^{l+1} \mathbb{E}[B_s^H B_t^H] dt ds = \frac{1}{2} \int_k^{k+1} \int_l^{l+1} (s^{2H} + t^{2H} - |s - t|^{2H}) dt ds = \\ &= \frac{1}{2} \left[\int_k^{k+1} \int_l^{l+1} (s^{2H} + t^{2H}) dt ds - \int_k^{k+1} \int_l^{l+1} |s - t|^{2H} dt ds \right] =: \frac{1}{2} (I_{1,k,l} - I_{2,k,l}). \end{aligned}$$

Then, obviously,

$$I_{1,k,l} = \frac{1}{2H+1} [(l+1)^{2H+1} - l^{2H+1} + (k+1)^{2H+1} - k^{2H+1}].$$

Consider two cases. If $k = l$, then the integral $I_{2,k,l}$ simplifies to

$$\begin{aligned} I_{2,k,k} &= \int_k^{k+1} \int_k^{k+1} |s - t|^{2H} dt ds = \int_0^1 \int_0^1 |s - t|^{2H} dt ds = \\ &= 2 \int_0^1 \int_0^s (s - t)^{2H} dt ds = 2 \int_0^1 \frac{s^{2H+1}}{2H+1} ds = \frac{2}{(2H+1)(2H+2)}. \end{aligned}$$

Further, for $k > l$, the integral equals

$$\begin{aligned} I_{2,k,l} &= \int_k^{k+1} \int_l^{l+1} (s - t)^{2H} dt ds = \int_k^{k+1} \frac{(s - l)^{2H+1} - (s - l - 1)^{2H+1}}{2H+1} ds = \\ &= \frac{1}{2H+1} \int_{k-l}^{k-l+1} [v^{2H+1} - (v-1)^{2H+1}] dv = \frac{(k-l+1)^{2H+2} - 2(k-l)^{2H+2} + (k-l-1)^{2H+2}}{(2H+1)(2H+2)}. \end{aligned}$$

Since the initial integral is symmetric in s and t , the previous case can be naturally generalized for $k < l$. Finally, we get

$$I_{2,k,l} = \frac{||k-l|+1|^{2H+2} - 2|k-l|^{2H+2} + ||k-l|-1|^{2H+2}}{(2H+1)(2H+2)},$$

and the proof is completed. $\square$

Corollary 1. *The variance of I_k is given by*

$$\mathbb{E}[I_k^2] = \frac{(H+1)[(k+1)^{2H+1} - k^{2H+1}] - 1}{(2H+1)(H+1)}.$$

The following statements present additional properties of the discrete-time process $\{I_k, k \geq 0\}$.

Corollary 2. *The discrete-time process $\{I_k, k \geq 0\}$ is not stationary, since the covariance does not entirely depend on the difference $|k-l|$.*

Lemma 2. *The discrete-time process $\{I_k, k \geq 0\}$ has stationary increments.*

Proof. We can prove even a bit more general fact. Indeed, introduce $\Delta I_k = I_{k+1} - I_k$, $k \geq 0$ and consider a shift of the integration interval, not necessarily by an integer, but by any $h \geq 0$. Then we need to check whether $\mathbb{E}[\Delta I_k \Delta I_{k+h}]$ depends solely on h .

$$\begin{aligned} \mathbb{E}[\Delta I_k \Delta I_{k+h}] &= \int_k^{k+1} \int_{k+h}^{k+h+1} \mathbb{E}[(B_{s+1}^H - B_s^H)(B_{t+1}^H - B_t^H)] dt ds = \\ &= \int_k^{k+1} \int_{k+h}^{k+h+1} \frac{1}{2} (|s-t+1|^{2H} + |s-t-1|^{2H} - 2|s-t|^{2H}) dt ds = \\ &= \frac{1}{2} \int_0^1 \int_0^1 (|u-v-h+1|^{2H} + |u-v-h-1|^{2H} - 2|u-v-h|^{2H}) dv du, \end{aligned} \quad (1)$$

which is a function just of h . $\square$

In what follows we use the notation for the covariance

$$\rho(d, H) := \mathbb{E}[\Delta I_0 \Delta I_d], d \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}.$$

Note that $\rho(d, H) = \mathbb{E}[\Delta I_k \Delta I_{k+d}]$, $k, k+d \geq 0$.

Lemma 3. *The covariance function $\rho(d, H)$ is given by the formula*

$$\rho(d, H) = \frac{|d+2|^{2H+2} - 4|d+1|^{2H+2} + 6|d|^{2H+2} - 4|d-1|^{2H+2} + |d-2|^{2H+2}}{2(2H+2)(2H+1)}.$$

Proof. Consider the covariance $\mathbb{E}[\Delta I_k \Delta I_l]$ for $k, l \geq 0$ and rewrite it in the following way:

$$\begin{aligned} \mathbb{E}[\Delta I_k \Delta I_l] &= \mathbb{E}[(I_{k+1} - I_k)(I_{l+1} - I_l)] = \mathbb{E}[I_{k+1}I_{l+1}] - \mathbb{E}[I_{k+1}I_l] - \mathbb{E}[I_kI_{l+1}] + \mathbb{E}[I_kI_l] = \\ &= -I_{2,k+1,l+1} + I_{2,k+1,l} + I_{2,k,l+1} - I_{2,k,l} = \\ &= \frac{|d+2|^{2H+2} - 4|d+1|^{2H+2} + 6|d|^{2H+2} - 4|d-1|^{2H+2} + |d-2|^{2H+2}}{2(2H+2)(2H+1)}, \quad d = |k-l|, \end{aligned}$$

from which the result of lemma immediately follows. $\square$

Remark 1. Note that in the case $l = k \pm 1$ ($d = 1$) the increments $\Delta I_k = I_{k+1} - I_k$ and $\Delta I_l = I_{l+1} - I_l$ are considered under overlapping intervals, while in the case $d \geq 2$ the underlying intervals do not overlap. This directly affects the properties of the respective covariance.

It is well-known that the increments of fractional Brownian motion are positively (negatively) correlated when $H > 1/2$ ($H < 1/2$) if the respective intervals are non-overlapping. However, as it was mentioned in Remark 1, the intervals overlap if $d = 1$ and do not overlap if $d \geq 2$. This affects the sign of the covariance. In the following lemma, we investigate the sign of $\rho(d, H)$.

Lemma 4. *The function $\rho : \mathbb{N}_0 \times (0, 1) \rightarrow \mathbb{R}$ has the following properties:*

- (i) *for $d \geq 2$, $\rho(d, H)$ is negative (positive) for $H < \frac{1}{2}$ ($H > \frac{1}{2}$), and zero otherwise.*
- (ii) *$\rho(1, H)$ is negative (positive) for $H < \alpha_H$ ($H > \alpha_H$), and zero otherwise, where $\alpha_H \approx 0.2626$.*

Proof. Similarly to (1),

$$\begin{aligned} \mathbb{E}[\Delta I_k \Delta I_l] &= \mathbb{E}[(I_{k+1} - I_k)(I_{l+1} - I_l)] = \\ &= \mathbb{E}\left[\left(\int_{k+1}^{k+2} B_s^H ds - \int_k^{k+1} B_s^H ds\right)\left(\int_{l+1}^{l+2} B_t^H dt - \int_l^{l+1} B_t^H dt\right)\right] = \\ &= \mathbb{E}\left[\int_0^1 (B_{u+k+1}^H - B_{u+k}^H) du \int_0^1 (B_{v+l+1}^H - B_{v+l}^H) dv\right] = \int_0^1 \int_0^1 \mathbb{E}[(B_{u+k+1}^H - B_{u+k}^H)(B_{v+l+1}^H - B_{v+l}^H)] du dv. \end{aligned}$$

Inside of the expectation we have a product of two increments. If $d = |k - l| \geq 2$, the intervals are non-overlapping, and we get the first statement.

We are left to consider the case $d = 1$, for which the intervals overlap. Using the expression for ρ , obtained in Lemma 4, we get:

$$\rho(1, H) = \frac{3^{2H+2} - 4 \cdot 2^{2H+2} + 7}{2(2H+2)(2H+1)}.$$

Since the denominator is strictly positive, the sign of $\rho(1, H)$ is determined by the sign of the numerator. Let us define the numerator as a function of H

$$f(H) = 3^{2H+2} - 4 \cdot 2^{2H+2} + 7.$$

We need to analyze the sign of $f(H)$ for $H \in (0, 1)$. First, observe the limits at boundaries

$$\lim_{H \rightarrow 0} f(H) = 0, \quad \lim_{H \rightarrow 1} f(H) = 24.$$

We compute the derivative to study monotonicity

$$f'(H) = \frac{d}{dH} (3^{2H+2} - 4 \cdot 2^{2H+2} + 7) = 18 \log 3 \cdot 9^H - 32 \log 2 \cdot 4^H.$$

Setting $f'(H) = 0$ gives the critical point β_H

$$18 \log 3 \cdot 9^{\beta_H} = 32 \log 2 \cdot 4^{\beta_H} \Rightarrow \beta_H = \frac{\log\left(\frac{16 \log 2}{9 \log 3}\right)}{\log \frac{9}{4}} \approx 0.14157.$$

Since $g(H) = \left(\frac{9}{4}\right)^H$ is strictly increasing, if $H < \beta_H$, then $f'(H) < 0$ and if $H > \beta_H$, then $f'(H) > 0$. Therefore, $f(H)$ is strictly decreasing on $(0, \beta_H)$ and strictly increasing on $(\beta_H, 1)$, attaining a local minimum at $H = \beta_H$. Since $f(H)$ starts at 0, becomes negative, and then increases up to a positive value, it changes the sign exactly once at some point α_H . Solving the equation $f(\alpha_H) = 3^{2\alpha_H+2} - 4 \cdot 2^{2\alpha_H+2} + 7 = 0$ numerically yields $\alpha_H \approx 0.2626$. Therefore, for $0 < H < \alpha_H$, the function $f(H)$ starts at 0, decreases to a minimum at β_H , and increases back to 0 at α_H . Thus, $f(H) < 0$ on $(0, \alpha_H)$. For $H > \alpha_H$, the function $f(H)$ continues to increase from $f(\alpha_H) = 0$. Thus, $f(H) > 0$ on $(\alpha_H, 1)$. For $H = \alpha_H$, $f(H) = 0$.

Since the sign of $\rho(1, H)$ is the same as the sign of $f(H)$, we conclude that $\rho(1, H)$ is negative for $H \in (0, \alpha_H)$, positive for $H \in (\alpha_H, 1)$, and zero at $H = \alpha_H$, where $\alpha_H \approx 0.2626$. $\square$

3. Coefficients of projection

In this section we consider the projection of ΔI_0 onto $(\Delta I_1, \dots, \Delta I_n)$, that is, the conditional expectation $\mathbb{E}[\Delta I_0 | \Delta I_1, \dots, \Delta I_n]$. Note that the joint distribution of $(\Delta I_0, \dots, \Delta I_n)$ is centered and Gaussian. Applying the theorem of normal correlation, we obtain the following representation

$$\mathbb{E}[\Delta I_0 | \Delta I_1, \dots, \Delta I_n] = \sum_{k=1}^n b_k^{(n)} \Delta I_k, \quad n \geq 1, \quad (2)$$

where the coefficients $b_k^{(n)} \in \mathbb{R}$. Since the increments are stationary (by Lemma 2), clearly $\mathbb{E}[\Delta I_k | \Delta I_{k+1}, \dots, \Delta I_{k+n}]$ contains the same coefficients. Now, let us consider the coefficients $b_k^{(n)}$ in the projection (2). The approach to obtain the coefficients $b_k^{(n)}$ is to solve the system of linear equations, which we get by multiplying both sides of (2) by ΔI_l , $l \geq 1$ and taking the expectation. We get the system

$$\rho_l = \sum_{k=1}^n b_k^{(n)} \rho_{|k-l|}, \quad l = 1, \dots, n,$$

where $\rho_d := \rho(d, H) = \mathbb{E}[\Delta I_0 \Delta I_d]$ and is given explicitly by Lemma 3.

Let us define the matrix $A^{(n)} \in \mathbb{R}^{n \times n}$ with entries $A_{ij} = \rho_{|i-j|}$ and the vector $\mathbf{r}^{(n)} = (\rho_1, \dots, \rho_n)^T$. Then

$$A^{(n)} \mathbf{b}^{(n)} = \mathbf{r}^{(n)}, \quad \text{and} \quad b_k^{(n)} = \frac{\det A_k^{(n)}}{\det A^{(n)}}, \quad (3)$$

where $A_k^{(n)}$ is the matrix obtained from $A^{(n)}$ by replacing the k th column with the vector $\mathbf{r}^{(n)}$. Note that $\det A^{(n)}$ is strictly positive, it cannot be proved similarly to Theorem 1.1 from (Banna et al., 2019). Therefore, the sign of each coefficient is determined by the $\det A_k^{(n)}$. Now we will proceed as follows: we will calculate analytically those coefficients for which this is possible, and we will carry out the rest of the analysis by numeric. Let us write down the expressions for ρ_0, ρ_1, ρ_2 , where $\rho_d := \rho(d, H)$, using the formula obtained in Lemma 3:

$$\rho_0 = \frac{8 \cdot 4^H - 8}{4(H+1)(2H+1)}, \quad \rho_1 = \frac{9 \cdot 9^H - 16 \cdot 4^H + 7}{4(H+1)(2H+1)}, \quad \rho_2 = \frac{16 \cdot 16^H - 36 \cdot 9^H + 24 \cdot 4^H - 4}{4(H+1)(2H+1)}.$$

Case $n = 1$: In this case, the matrix $A^{(n)}$ and the vector $\mathbf{r}^{(n)}$ contain only one element each, and they are equal

$$A^{(1)} = \rho_0 = \frac{8 \cdot 4^H - 8}{4(H+1)(2H+1)}, \quad r^{(1)} = \rho_1 = \frac{9 \cdot 9^H - 16 \cdot 4^H + 7}{4(H+1)(2H+1)}.$$

Solving the equation $A^{(1)}b_1^{(1)} = r^{(1)}$ for $b_1^{(1)}$, we get that

$$b_1^{(1)} = \frac{\rho_1}{\rho_0} = \frac{9 \cdot 9^H - 16 \cdot 4^H + 7}{8 \cdot 4^H - 8}.$$

Since ρ_0 is a variance, it is strictly positive for all $H \in (0, 1)$. Therefore, the sign of $b_1^{(1)}$ is determined by the ρ_1 in the numerator. The next result is a direct consequence of Lemma 4.

Proposition 1. *The coefficient $b_1^{(1)}$ is negative (positive) for $H < H_{1,1}$ ($H > H_{1,1}$), where $H_{1,1} := \alpha_H \approx 0.2626$.*

Case $n = 2$: In this case

$$A^{(2)} = \begin{pmatrix} \rho_0 & \rho_1 \\ \rho_1 & \rho_0 \end{pmatrix}, \quad \mathbf{r}^{(2)} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}.$$

Therefore

$$b_1^{(2)} = \frac{\rho_1(\rho_0 - \rho_2)}{\rho_0^2 - \rho_1^2}, \quad b_2^{(2)} = \frac{\rho_0\rho_2 - \rho_1^2}{\rho_0^2 - \rho_1^2}.$$

As mentioned previously, the sign of each projection coefficient is determined by its numerator, since the denominators are the determinants of a covariance matrix and therefore always positive.

Proposition 2. *The coefficient $b_1^{(2)}$ is negative (positive) for $H < H_{2,1}$ ($H > H_{2,1}$), where $H_{2,1} := \alpha_H \approx 0.2626$, and the coefficient $b_2^{(2)}$ is negative (positive) for $H < H_{2,2}$ ($H > H_{2,2}$), where $H_{2,2} \approx 0.8171$.*

Proof. Note that $\rho_0 > \rho_2$ for all $H \in (0, 1)$, because of the general variance-covariance inequality. So, the sign of the coefficient $b_1^{(2)}$ is determined by ρ_1 , which we addressed earlier. Therefore, according to Lemma 4, $b_1^{(2)}$ is negative for $H < \alpha_H$ and positive afterward, where $\alpha_H \approx 0.2626$.

For $b_2^{(2)}$, its sign is determined by the function $\rho_0\rho_2 - \rho_1^2 := g_2(H)$, which can be written in closed form as

$$g_2(H) = \frac{-81 \cdot 81^H + 128 \cdot 64^H - 192 \cdot 16^H + 162 \cdot 9^H - 17}{16(H+1)^2(2H+1)^2} = \frac{g_2^*(H)}{16(H+1)^2(2H+1)^2}.$$

Due to the complex nature of $g_2^*(H)$, a complete analytical study of its sign on the interval $(0, 1)$ or even establishing its behavior on certain subintervals using elementary analytical tools seems to be an unsolvable problem.

Consequently, we rely on numerical methods to analyze the behavior of function $g_2^*(H)$. These numerical investigations, which are further supported by a plot of $b_2^{(2)}$ presented at (Fig. 1), indicate that $g_2^*(H)$ has a unique root within the interval $(0, 1)$ at point $H_{2,2} \approx 0.8171$. The numerical analysis shows that $g_2^*(H) < 0$ for $H < H_{2,2}$ and $g_2^*(H) > 0$ for $H > H_{2,2}$. $\square$

Remark 2. While a comprehensive analytical proof of the behavior of $g_2^*(H)$ on the interval $(0, 1)$ is not provided due to its complexity, we can investigate its characteristics around specific points of interest using Taylor expansions. This approach does not serve as a general proof of the function's behavior across the entire interval but offers insights into its local properties. In our primary consideration, $H \in (0, 1)$, but since $g_2^*(H)$ is continuous, we can extend the domain to $H \in [0, 1]$ to include boundary points. To explore the behavior near $H = 0$, we can employ Taylor expansion. It can be verified that $g_2^*(0) = 0$. The first and second derivatives of $g_2^*(H)$ are

$$g_2^{*'}(H) = -81 \cdot 81^H \log 81 + 128 \cdot 64^H \log 64 - 192 \cdot 16^H \log 16 + 162 \cdot 9^H \log 9.$$

$$g_2^{*''}(H) = -81 \cdot 81^H (\log 81)^2 + 128 \cdot 64^H (\log 64)^2 - 192 \cdot 16^H (\log 16)^2 + 162 \cdot 9^H (\log 9)^2.$$

Evaluating at $H = 0$, we find that $g_2^{*'}(0) = 0$ and $g_2^{*''}(0) < 0$. The Taylor expansion for $g_2^*(H)$ around $H_0 = 0$ is thus

$$g_2^*(H) = g_2^*(0) + g_2^{*'}(0)H + \frac{g_2^{*''}(0)}{2}H^2 + o(H^2).$$

Since $g_2^{*''}(0) < 0$, this suggests that $g_2^*(H)$ is negative for small positive values of H . To understand the function's behavior around the numerically identified root $H_{2,2} \approx 0.8171$, where $g_2^*(H_{2,2}) \approx 0$, we can examine its first derivative. It can be

checked that $g_2^*(H_{2,2}) > 0$. The Taylor expansion around $H_0 = H_{2,2}$ is

$$g_2^*(H) = g_2^*(H_{2,2}) + g_2^{*'}(H_{2,2})(H - H_{2,2}) + o(H - H_{2,2}).$$

Since $g_2^*(H_{2,2}) = 0$ and $g_2^{*'}(H_{2,2}) > 0$, this expansion indicates that the function is increasing as it passes through this root, transitioning from negative to positive values. Furthermore, direct evaluation shows that $g_2^*(1) = 0$. Of course, while the fact that we dropped the denominator from the analysis affected the position of the critical points, it did not affect the roots and thus the behavior of the sign of $b_2^{(2)}$ we were initially interested in.

Numerical results

In this section, we present the numerically computed coefficients $b_k^{(n)}$ for $n = 1, \dots, 10$. The results below are listed for the Hurst indices $H = 0.1, 0.4, 0.7, 0.9$ (Tab. 1-4).

Table 1

Coefficients $b_k^{(n)}$ for $H=0.1$										
n/k	1	2	3	4	5	6	7	8	9	10
1	-0.1409	-	-	-	-	-	-	-	-	-
2	-0.1637	-0.1621	-	-	-	-	-	-	-	-
3	-0.1808	-0.1793	-0.1050	-	-	-	-	-	-	-
4	-0.1897	-0.1946	-0.1204	-0.0854	-	-	-	-	-	-
5	-0.1957	-0.2031	-0.1341	-0.0986	-0.0700	-	-	-	-	-
6	-0.1999	-0.2089	-0.1421	-0.1108	-0.0817	-0.0596	-	-	-	-
7	-0.2030	-0.2132	-0.1478	-0.1181	-0.0925	-0.0700	-0.0519	-	-	-
8	-0.2053	-0.2164	-0.1521	-0.1236	-0.0993	-0.0798	-0.0612	-0.0459	-	-
9	-0.2072	-0.2189	-0.1553	-0.1276	-0.1044	-0.0861	-0.0701	-0.0544	-0.0412	-
10	-0.2088	-0.2210	-0.1580	-0.1309	-0.1083	-0.0909	-0.0759	-0.0626	-0.0490	-0.0374

Table 2

Coefficients $b_k^{(n)}$ for $H=0.4$										
n/k	1	2	3	4	5	6	7	8	9	10
1	0.1377	-	-	-	-	-	-	-	-	-
2	0.1500	-0.0893	-	-	-	-	-	-	-	-
3	0.1485	-0.0868	-0.0163	-	-	-	-	-	-	-
4	0.1481	-0.0890	-0.0125	-0.0252	-	-	-	-	-	-
5	0.1477	-0.0892	-0.0141	-0.0226	-0.0174	-	-	-	-	-
6	0.1474	-0.0896	-0.0143	-0.0240	-0.0151	-0.0156	-	-	-	-
7	0.1472	-0.0898	-0.0146	-0.0242	-0.0163	-0.0136	-0.0133	-	-	-
8	0.1471	-0.0899	-0.0148	-0.0245	-0.0165	-0.0147	-0.0116	-0.0118	-	-
9	0.1469	-0.0901	-0.0150	-0.0247	-0.0168	-0.0148	-0.0125	-0.0102	-0.0105	-
10	0.1468	-0.0902	-0.0151	-0.0248	-0.0169	-0.0151	-0.0126	-0.0111	-0.0091	-0.0095

Table 3

Coefficients $b_k^{(n)}$ for $H=0.7$										
n/k	1	2	3	4	5	6	7	8	9	10
1	0.5091	-	-	-	-	-	-	-	-	-
2	0.5212	-0.0238	-	-	-	-	-	-	-	-
3	0.5235	-0.0730	0.0945	-	-	-	-	-	-	-
4	0.5200	-0.0704	0.0755	0.0363	-	-	-	-	-	-
5	0.5186	-0.0735	0.0784	0.0149	0.0410	-	-	-	-	-
6	0.5173	-0.0740	0.0760	0.0172	0.0250	0.0309	-	-	-	-
7	0.5164	-0.0747	0.0755	0.0151	0.0271	0.0165	0.0278	-	-	-
8	0.5158	-0.0750	0.0748	0.0147	0.0253	0.0183	0.0153	0.0241	-	-
9	0.5152	-0.0754	0.0745	0.0142	0.0250	0.0167	0.0170	0.0129	0.0216	-
10	0.5148	-0.0756	0.0741	0.0139	0.0245	0.0164	0.0155	0.0144	0.0116	0.0194

Table 4

Coefficients $b_k^{(n)}$ for $H=0.9$										
n/k	1	2	3	4	5	6	7	8	9	10
1	0.8212	-	-	-	-	-	-	-	-	-
2	0.8078	0.0163	-	-	-	-	-	-	-	-
3	0.8048	-0.1323	0.1840	-	-	-	-	-	-	-
4	0.7906	-0.1221	0.1219	0.0772	-	-	-	-	-	-
5	0.7841	-0.1324	0.1322	0.0103	0.0845	-	-	-	-	-
6	0.7788	-0.1331	0.1239	0.0187	0.0352	0.0629	-	-	-	-
7	0.7752	-0.1351	0.1228	0.0116	0.0427	0.0187	0.0568	-	-	-
8	0.7724	-0.1360	0.1207	0.0111	0.0367	0.0253	0.0189	0.0489	-	-
9	0.7703	-0.1368	0.1196	0.0094	0.0363	0.0200	0.0249	0.0150	0.0439	-
10	0.7685	-0.1374	0.1187	0.0087	0.0348	0.0196	0.0202	0.0204	0.0135	0.0394

The situation with the signs of the coefficients, to some extent, reflects the situation with overlapping intervals. However, it is also a mixture with the situation of projection of the increments of fBm: for small values of H all coefficients are negative, as in the case of increments of fBm, but for $H > \frac{1}{2}$ the coefficients become strictly positive, except for the second one, and for the increments of fBm they are all positive.

To better illustrate the behavior of the coefficients, we present the plots of the coefficients as functions of H for $n = 2, 3$ (Fig. 1).

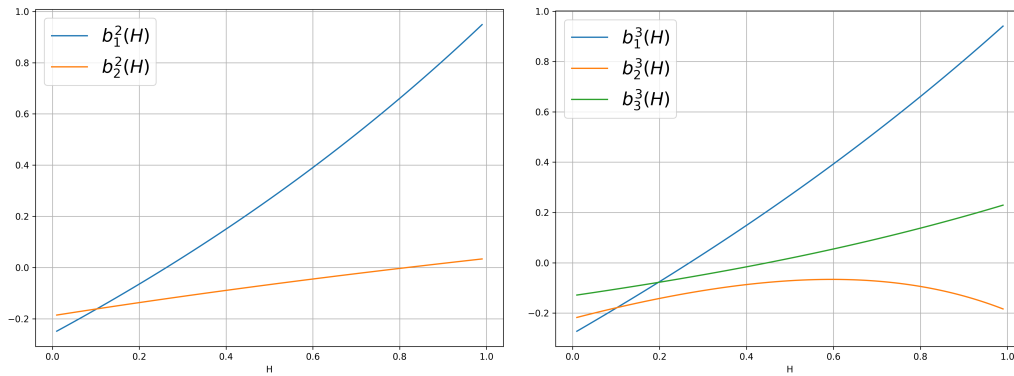


Fig. 1. LEFT – Case $n = 2$: $b_1^{(2)}, b_2^{(2)}$ as functions of H . RIGHT – Case $n = 3$: $b_1^{(3)}, b_2^{(3)}, b_3^{(3)}$ as functions of H

4. Projection of “non-overlapping” increments

Our previous analysis focused on the general form of the projection, which involved increments on the overlapping intervals, leading to complexities in coefficient behavior. Let us now consider the projection $\mathbb{E}[\Delta I_0 \mid \Delta I_2, \Delta I_4, \dots, \Delta I_{2p}]$, where $p \geq 1$. The conditional expectation then can be written as follows

$$\mathbb{E}[\Delta I_0 \mid \Delta I_2, \Delta I_4, \dots, \Delta I_{2p}] = \sum_{k=1}^p \tilde{b}_k^{(p)} \Delta I_{2k}, \quad p \geq 1. \quad (4)$$

To obtain the coefficients $\tilde{b}_k^{(p)}$, we multiply both sides of the projection equation (4) by ΔI_{2l} for $l = 1, \dots, p$, and take the expectation. This yields the following system of linear equations

$$\mathbb{E}[\Delta I_0 \Delta I_{2l}] = \sum_{k=1}^p \tilde{b}_k^{(p)} \mathbb{E}[\Delta I_{2k} \Delta I_{2l}].$$

Using the definition $\rho_d = \mathbb{E}[\Delta I_0 \Delta I_d]$ and the stationarity property $\mathbb{E}[\Delta I_i \Delta I_j] = \rho_{|i-j|}$, we get

$$\rho_{2l} = \sum_{k=1}^p \tilde{b}_k^{(p)} \rho_{|2k-2l|}, \quad l = 1, \dots, p.$$

Let us define the matrix $\tilde{A}^{(p)} \in \mathbb{R}^{p \times p}$ with entries

$$\tilde{A}_{ij} = \rho_{|2i-2j|} = \rho_{2|i-j|}, \quad i, j \in \{1, \dots, p\}.$$

Then the matrix $\tilde{A}^{(p)}$ has the following form

$$\tilde{A}^{(p)} = \begin{pmatrix} \rho_0 & \rho_2 & \rho_4 & \cdots & \rho_{2(p-1)} \\ \rho_2 & \rho_0 & \rho_2 & \cdots & \rho_{2(p-2)} \\ \rho_4 & \rho_2 & \rho_0 & \cdots & \rho_{2(p-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{2(p-1)} & \rho_{2(p-2)} & \rho_{2(p-3)} & \cdots & \rho_0 \end{pmatrix}.$$

Let the vector of coefficients be $\tilde{\mathbf{b}}^{(p)} = (\tilde{b}_1^{(p)}, \dots, \tilde{b}_p^{(p)})^T$, and vector $\tilde{\mathbf{r}}^{(p)}$ be defined as $\tilde{\mathbf{r}}^{(p)} = (\rho_2, \rho_4, \dots, \rho_{2p})^T$. Then

$$\tilde{A}^{(p)} \tilde{\mathbf{b}}^{(p)} = \tilde{\mathbf{r}}^{(p)}, \text{ and } \tilde{b}_k^{(p)} = \frac{\det \tilde{A}_k^{(p)}}{\det \tilde{A}^{(p)}}, \quad (5)$$

where $\tilde{A}_k^{(p)}$ is the matrix obtained from $\tilde{A}^{(p)}$ by replacing the k th column with the vector $\tilde{\mathbf{r}}^{(p)}$. The covariance function $\rho_d = \rho(d, H) = \mathbb{E}[\Delta I_0 \Delta I_d]$ is given by Lemma 3, as before.

We will conduct our further analysis of the coefficients in a way similar to that we did in the previous section. Specifically, we will consider the cases $p = 1$ and $p = 2$, for which we write down useful expressions, namely ρ_0, ρ_2, ρ_4 , in the following closed form

$$\rho_0 = \frac{8 \cdot 4^H - 8}{4(H+1)(2H+1)}, \quad \rho_2 = \frac{16 \cdot 16^H - 36 \cdot 9^H + 24 \cdot 4^H - 4}{4(H+1)(2H+1)},$$

$$\rho_4 = \frac{36 \cdot 36^H - 100 \cdot 25^H + 96 \cdot 16^H - 36 \cdot 9^H + 4 \cdot 4^H}{4(H+1)(2H+1)}.$$

Case $p = 1$: In this case, the projection coefficient $\tilde{b}_1^{(1)}$ is given by $\rho_0 \tilde{b}_1^{(1)} = \rho_2$. Thus, we trivially get

$$\tilde{b}_1^{(1)} = \frac{\rho_2}{\rho_0} = \frac{16 \cdot 16^H - 36 \cdot 9^H + 24 \cdot 4^H - 4}{8 \cdot 4^H - 8}.$$

As before, ρ_0 is a variance and thus always positive. The behavior of $\tilde{b}_1^{(1)}$ is then defined by the numerator ρ_2 .

Proposition 3. *The coefficient $\tilde{b}_1^{(1)}$ is negative (positive) for $H < \frac{1}{2}$ ($H > \frac{1}{2}$).*

Proof. By Lemma 4, we know that ρ_2 , and thus $\tilde{b}_1^{(1)}$, is negative (positive) for $H < \frac{1}{2}$ ($H > \frac{1}{2}$). □

Case $p = 2$: Here, the coefficients $\tilde{b}_1^{(2)}$ and $\tilde{b}_2^{(2)}$ are given by the system of equations $\tilde{A}^{(2)} \tilde{\mathbf{b}}^{(2)} = \tilde{\mathbf{r}}^{(2)}$, where

$$\tilde{A}^{(2)} = \begin{pmatrix} \rho_0 & \rho_2 \\ \rho_2 & \rho_0 \end{pmatrix}, \quad \tilde{\mathbf{r}}^{(2)} = \begin{pmatrix} \rho_2 \\ \rho_4 \end{pmatrix}.$$

The coefficients $\tilde{\mathbf{b}}^{(2)}$ equal

$$\tilde{b}_1^{(2)} = \frac{\rho_2(\rho_0 - \rho_4)}{\rho_0^2 - \rho_2^2}, \quad \tilde{b}_2^{(2)} = \frac{\rho_0\rho_4 - \rho_2^2}{\rho_0^2 - \rho_2^2}.$$

As repeatedly emphasized, the denominators are the determinants of covariance matrices, hence strictly positive. Therefore, the behavior of the coefficients is defined by the numerators $\rho_2(\rho_0 - \rho_4) =: f_1(H)$, $\rho_0\rho_4 - \rho_2^2 =: f_2(H)$ with the following closed form

$$f_1(H) = \frac{f_1^*(H)}{(H+1)^2(2H+1)^2}, \quad f_2(H) = \frac{f_2^*(H)}{(H+1)^2(2H+1)^2},$$

where f_1^*, f_2^* are given by

$$\begin{aligned} f_1^*(H) &= -36 \cdot 576^H + 100 \cdot 400^H + 81 \cdot 324^H - 96 \cdot 256^H - 225 \cdot 225^H + 198 \cdot 144^H + 150 \cdot 100^H - \\ &\quad - 81 \cdot 81^H - 140 \cdot 64^H + 54 \cdot 36^H - 25 \cdot 25^H + 22 \cdot 16^H + 9 \cdot 9^H - 13 \cdot 4^H + 2, \\ f_2^*(H) &= -16 \cdot 256^H + 90 \cdot 144^H - 50 \cdot 100^H - 81 \cdot 81^H + 72 \cdot 36^H + 50 \cdot 25^H - 74 \cdot 16^H + 10 \cdot 4^H - 1. \end{aligned}$$

Proposition 4. *The coefficients $\tilde{b}_1^{(2)}, \tilde{b}_2^{(2)}$ are negative (positive) for $H < \frac{1}{2}$ ($H > \frac{1}{2}$).*

Proof. (i) In case of $\tilde{b}_1^{(2)}$, we use the same reasoning as in the previous section, since the expression in the numerator simply factorizes to $\rho_2(\rho_0 - \rho_4)$. We know that $\rho_0 > \rho_4$ by the general variance-covariance inequality and thus the behavior of the numerator is defined by ρ_2 , which was discussed previously. Hence, $\tilde{b}_1^{(2)}$ is negative for $H < \frac{1}{2}$ and positive for $H > \frac{1}{2}$.

(ii) The determination of the sign of $\tilde{b}_2^{(2)}$ for $H < \frac{1}{2}$ follows from Lemma 4, which establishes that $\rho_0\rho_4 - \rho_2^2 < 0$ within this domain, since ρ_0 is strictly positive and $\rho_2 < 0$, $\rho_4 < 0$ for $H < \frac{1}{2}$. However, the case $H > \frac{1}{2}$ requires a different approach, which is not accessible at the moment. Alternatively, we present a plot of $\tilde{b}_2^{(2)}$ at (Fig. 2) to convey its behavior. □

Remark 3. Similarly to Remark 2, we can investigate the behavior of f_2^* around interesting points using Taylor expansions. First, we check the behavior near $H = 0$. We verify that $f_2^*(0) = 0$, $f_2^{*'}(0) = 0$, $f_2^{*''}(0) < 0$ and then apply the Taylor expansion around $H_0 = 0$, which immediately gives us that $f_2^*(H)$ starts negative for small positive H . Next, we analyze the function near $H = \frac{1}{2}$. We make sure that $f_2^*(\frac{1}{2}) = 0$ and $f_2^{*'}(\frac{1}{2}) > 0$. Since $f_2^{*'}(\frac{1}{2}) > 0$, the Taylor expansion around $H_0 = \frac{1}{2}$ shows that the function starts positive for $H > \frac{1}{2}$. Finally, we analyze the function near $H = 1$. We ensure that $f_2^*(1) = 0$ and $f_2^{*'}(1) < 0$. Since $f_2^{*'}(1) < 0$, the function is decreasing at $H = 1$. This shows that for $H < 1$ and $H \approx 1$, $f_2^*(H) > f_2^*(1) = 0$, so the function was positive and decreases towards zero at $H = 1$.

Numerical results

Here we present the numerically computed coefficients $\tilde{b}_k^{(p)}$ for $p = 1, \dots, 10$. The results below are listed for the Hurst indices $H = 0.1, 0.4, 0.7, 0.9$ (Tab. 5-8).

Table 5

Coefficients $\tilde{b}_k^{(p)}$ for H=0.1										
p/k	1	2	3	4	5	6	7	8	9	10
1	-0.1391	-	-	-	-	-	-	-	-	-
2	-0.1462	-0.0513	-	-	-	-	-	-	-	-
3	-0.1476	-0.0552	-0.0267	-	-	-	-	-	-	-
4	-0.1480	-0.0561	-0.0291	-0.0165	-	-	-	-	-	-
5	-0.1482	-0.0564	-0.0298	-0.0181	-0.0112	-	-	-	-	-
6	-0.1483	-0.0566	-0.0300	-0.0186	-0.0124	-0.0081	-	-	-	-
7	-0.1483	-0.0567	-0.0301	-0.0188	-0.0127	-0.0091	-0.0062	-	-	-
8	-0.1484	-0.0567	-0.0302	-0.0188	-0.0129	-0.0093	-0.0069	-0.0049	-	-
9	-0.1484	-0.0567	-0.0302	-0.0189	-0.0130	-0.0094	-0.0071	-0.0055	-0.0040	-
10	-0.1484	-0.0568	-0.0302	-0.0189	-0.0130	-0.0095	-0.0072	-0.0057	-0.0044	-0.0033

Table 6

Coefficients $\tilde{b}_k^{(p)}$ for H=0.4										
p/k	1	2	3	4	5	6	7	8	9	10
1	-0.0686	-	-	-	-	-	-	-	-	-
2	-0.0708	-0.0314	-	-	-	-	-	-	-	-
3	-0.0714	-0.0328	-0.0202	-	-	-	-	-	-	-
4	-0.0717	-0.0333	-0.0212	-0.0148	-	-	-	-	-	-
5	-0.0719	-0.0335	-0.0216	-0.0156	-0.0116	-	-	-	-	-
6	-0.0720	-0.0337	-0.0218	-0.0159	-0.0123	-0.0095	-	-	-	-
7	-0.0720	-0.0338	-0.0220	-0.0161	-0.0126	-0.0101	-0.0081	-	-	-
8	-0.0721	-0.0339	-0.0221	-0.0162	-0.0127	-0.0104	-0.0086	-0.0070	-	-
9	-0.0721	-0.0339	-0.0221	-0.0163	-0.0128	-0.0105	-0.0088	-0.0074	-0.0061	-
10	-0.0722	-0.0340	-0.0222	-0.0164	-0.0129	-0.0106	-0.0089	-0.0076	-0.0065	-0.0055

Table 7

Coefficients $\tilde{b}_k^{(p)}$ for H=0.7										
p/k	1	2	3	4	5	6	7	8	9	10
1	0.2416	-	-	-	-	-	-	-	-	-
2	0.2172	0.1008	-	-	-	-	-	-	-	-
3	0.2105	0.0864	0.0663	-	-	-	-	-	-	-
4	0.2073	0.0821	0.0559	0.0495	-	-	-	-	-	-
5	0.2053	0.0799	0.0527	0.0413	0.0395	-	-	-	-	-
6	0.2040	0.0786	0.0509	0.0387	0.0327	0.0329	-	-	-	-
7	0.2031	0.0776	0.0499	0.0372	0.0305	0.0271	0.0282	-	-	-
8	0.2024	0.0770	0.0491	0.0363	0.0293	0.0252	0.0232	0.0246	-	-
9	0.2018	0.0765	0.0486	0.0357	0.0285	0.0241	0.0215	0.0202	0.0219	-
10	0.2014	0.0761	0.0481	0.0352	0.0279	0.0234	0.0205	0.0187	0.0179	0.0197

Table 8

Coefficients $\tilde{b}_k^{(p)}$ for H=0.9										
p/k	1	2	3	4	5	6	7	8	9	10
1	0.6797	-	-	-	-	-	-	-	-	-
2	0.5226	0.2311	-	-	-	-	-	-	-	-
3	0.4880	0.1527	0.1500	-	-	-	-	-	-	-
4	0.4715	0.1360	0.0965	0.1095	-	-	-	-	-	-
5	0.4621	0.1276	0.0848	0.0689	0.0862	-	-	-	-	-
6	0.4560	0.1228	0.0788	0.0598	0.0534	0.0711	-	-	-	-
7	0.4517	0.1195	0.0751	0.0550	0.0460	0.0435	0.0604	-	-	-
8	0.4485	0.1172	0.0727	0.0521	0.0420	0.0372	0.0367	0.0526	-	-
9	0.4460	0.1155	0.0710	0.0502	0.0396	0.0339	0.0313	0.0317	0.0465	-
10	0.4441	0.1142	0.0697	0.0488	0.0379	0.0318	0.0283	0.0269	0.0279	0.0417

Here we can see that the behavior of the projection coefficients in the non-overlapping scenario is consistent with the properties of fBm: all coefficients are negative for $H < \frac{1}{2}$, and all coefficients are positive for $H > \frac{1}{2}$. As in the previous case, we present the plots of the coefficients as functions of H for $p = 2, 3$ (Fig. 2).

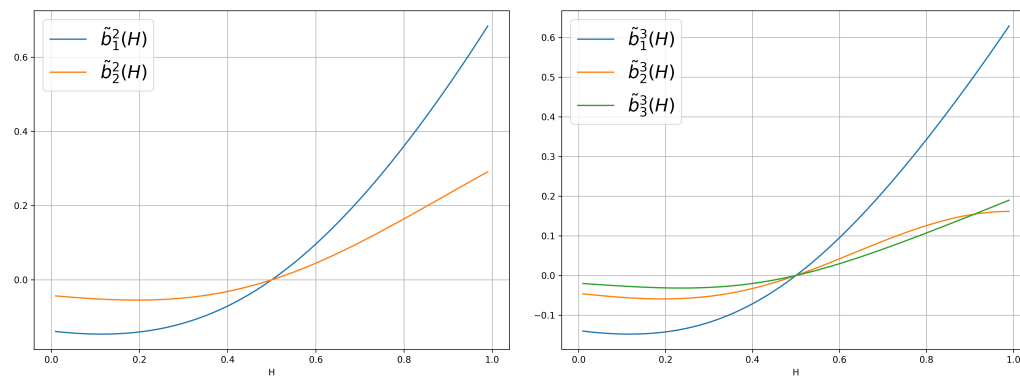


Fig. 2. LEFT – Case $p = 2$: $\tilde{b}_2^{(1)}, \tilde{b}_2^{(2)}$ as functions of H . RIGHT – Case $p = 3$: $\tilde{b}_3^{(1)}, \tilde{b}_3^{(2)}, \tilde{b}_3^{(3)}$ as functions of H

5. Calculation of the Prediction Error

In this section, we derive the formula for the Mean Squared Error (MSE) of the projection discussed in Section 3. The error of the prediction is defined as $e_n = \Delta I_0 - \mathbb{E}[\Delta I_0 | \Delta I_1, \dots, \Delta I_n]$. Since the process is centered, the MSE is simply the variance of this error, $\text{MSE}_n = \mathbb{E}[e_n^2]$, and it equals

$$\text{MSE}_n = \rho_0 - \sum_{k=1}^n \rho_k b_k^{(n)}$$

for overlapping, and

$$\text{MSE}_p = \rho_0 - \sum_{k=1}^p \rho_{2k} \tilde{b}_k^{(p)}$$

for non-overlapping intervals, where the values of coefficients are given by (3) and (5). In Tab. 9, we can see that the MSE decreases as the number of predictors grows, both in the overlapping and non-overlapping scenarios. It is interesting to see that MSE decreases as H approaches its boundary points $H = 0$ and $H = 1$ and increases when H is close to $1/2$. The behaviour around $H = 1$, where fBm and consequently ifBm degenerate to $\xi \cdot \phi(t)$, where $\phi(t) = t$ for fBm, and $\phi(t) = t^2/2$ for ifBm, is very clear. The same is about $H = 1/2$ because the terms of prediction are independent for non-overlapping intervals. When H approaches 0, fBm tends in distribution to white noise, and therefore the norm of projection increases, and consequently MSE decreases.

Table 9

Mean Squared Error (MSE) for the Overlapping Case

n	H = 0.01	H = 0.4	H = 0.7	H = 0.99
1	0.0259	0.5770	0.5952	0.0366
2	0.0250	0.5724	0.5949	0.0366
3	0.0246	0.5723	0.5895	0.0347
4	0.0243	0.5719	0.5888	0.0344
5	0.0242	0.5717	0.5878	0.0340
6	0.0240	0.5716	0.5872	0.0338
7	0.0239	0.5715	0.5868	0.0336
8	0.0239	0.5714	0.5864	0.0335
9	0.0238	0.5714	0.5861	0.0334
10	0.0238	0.5713	0.5859	0.0333

Mean Squared Error (MSE) for the Non-Overlapping Case

p	H = 0.01	H = 0.4	H = 0.7	H = 0.99
1	0.0266	0.5854	0.7566	0.0696
2	0.0266	0.5848	0.7489	0.0637
3	0.0266	0.5846	0.7456	0.0614
4	0.0266	0.5845	0.7437	0.0602
5	0.0266	0.5844	0.7426	0.0595
6	0.0266	0.5843	0.7418	0.0591
7	0.0266	0.5843	0.7412	0.0587
8	0.0266	0.5843	0.7407	0.0585
9	0.0266	0.5842	0.7404	0.0583
10	0.0266	0.5842	0.7401	0.0581

Discussion and conclusions

In this work, we have successfully investigated the projection of increments of integrated fractional Brownian motion (ifBm). We began by formally introducing ifBm, deriving its covariance function, and establishing the key property that its increments are stationary. Our primary contribution lies in the determination of the coefficients for the linear prediction of an ifBm increment, ΔI_0 , based on a series of subsequent increments, $(\Delta I_1, \dots, \Delta I_n)$. We derived exact analytical formulas for the projection onto a single subsequent increment in both overlapping and non-overlapping scenarios. For projections onto two increments, we provided partially analytical results. The corresponding code for reproducing these results is provided in our repository¹.

To handle more complex cases, we computed numerical solutions for projections onto as many as ten increments for both interval configurations. A significant finding of this study is the complex behavior of these projection coefficients, particularly the sign changes observed as a function of the Hurst index H . This reveals a non-trivial dependence structure, especially when the intervals overlap. However, the covariance structure of the fractional Brownian motion poses a significant problem in terms of formal analysis, and thus needs new approaches.

Authors' contribution: Yevhenii Bundiuk — derivation of the main results, writing – original draft; Yuliya Mishura — conceptualization, methodology, writing – revision and editing; Enrica Pirozzi — writing – revision and editing.

Sources of funding. The second author is supported by the Swedish Foundation for Strategic Research, grant UKR24-0004. The third author was partially funded by the project MUR NextGenerationEU PRIN-PNRR P2022XSF5H "Stochastic models in biomathematics and applications", CUP E53D23018050001. The third author is also a member of INdAM-GNCS.

References

- Abundo, M., & Pirozzi, E. (2018). Integrated stationary Ornstein–Uhlenbeck process, and double integral processes. *Physica A: Statistical Mechanics and Its Applications*, 494, 265–275. <https://doi.org/10.1016/j.physa.2017.12.043>
- Abundo, M., & Pirozzi, E. (2019). On the integral of the fractional Brownian motion and some pseudo-fractional Gaussian processes. *Mathematics*, 7(10), 991. <https://doi.org/10.3390/math7100991>
- Alòs, E., Mazet, O., & Nualart, D. (2001). Stochastic calculus with respect to Gaussian processes. *The Annals of Probability*, 29(2), 766–801. <https://doi.org/10.1214/aop/1008956692>
- Alòs, E., & Nualart, D. (2003). Stochastic integration with respect to the fractional Brownian motion. *Stochastics and Stochastics Reports*, 75(3), 129–152. <https://doi.org/10.1080/1045112031000078917>
- Anh, V. V., & Inoue, A. (2004). Prediction of fractional Brownian motion with Hurst index less than 1/2. *Bulletin of the Australian Mathematical Society*, 70(2), 321–328. <https://doi.org/10.1017/S0004972700034535>
- Banna, O., Mishura, Y., Ralchenko, K., & Shklyar, S. (2019). *Fractional Brownian motion: Approximations and Projections*. ISTE Ltd and John Wiley & Sons Inc. <https://doi.org/10.1002/9781119476771>
- Barton, R. J., & Vincent Poor, H. (1988). Signal detection in fractional Gaussian noise. *IEEE Transactions on Information Theory*, 34(5), 943–959. <https://doi.org/10.1109/18.21218>
- Bender, C. (2003a). An Itô formula for generalized functionals of a fractional Brownian motion with arbitrary Hurst parameter. *Stochastic Processes and their Applications*, 104(1), 81–106. [https://doi.org/10.1016/S0304-4149\(02\)00212-0](https://doi.org/10.1016/S0304-4149(02)00212-0)
- Bender, C. (2003b). An S-transform approach to integration with respect to a fractional Brownian motion. *Bernoulli*, 9(6), 955–983. <https://doi.org/10.3150/bj/1072215197>

¹https://github.com/RadicallyUprooted/prediction_formulas_for_integrated_fbm

- Bender, C., & Elliott, R. J. (2004). Arbitrage in a discrete version of the Wick-fractional Black-Scholes market. *Mathematics of Operations Research*, 29(4), 935–945. <https://doi.org/10.1287/moor.1040.0096>
- Bender, C., Sottinen, T., & Valkeila, E. (2008). No-arbitrage Pricing Beyond Semimartingales. *Finance and Stochastics*, 12(4), 441–468. <https://doi.org/10.1007/s00780-008-0074-8>
- Benth, F. E. (2003). On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. *Applied Mathematical Finance*, 10(4), 303–324. <https://doi.org/10.1080/1350486032000174628>
- Beran, J. (1994). *Statistics for Long-memory Processes*. Chapman & Hall. <https://doi.org/10.1201/9780203738481>
- Biagini, F., Hu, Y., Øksendal, B., & Sulem, A. (2002). A stochastic maximum principle for processes driven by a fractional Brownian motion. *Stochastic Processes and their Applications*, 100(1-2), 233–253. [https://doi.org/10.1016/S0304-4149\(02\)00105-9](https://doi.org/10.1016/S0304-4149(02)00105-9)
- Biagini, F., Hu, Y., Øksendal, B., & Zhang, T. (2008). *Stochastic Calculus for Fractional Brownian Motion and Applications*. Springer. <https://doi.org/10.1007/978-1-84628-797-8>
- Bodnarchuk, I., & Mishura, Y. (2024). Combinatorial approach to the calculation of projection coefficients for the simplest Gaussian-Volterra process. *Modern Stochastics: Theory and Applications*, 11(4), 403–419. <https://doi.org/10.15559/24-VMSTA252>
- Boufoussi, B., & Ouknine, Y. (2003). On a SDE driven by a fractional Brownian motion and with monotone drift. *Electronic Communications in Probability*, 8, 122–134. <https://doi.org/10.1214/ECP.v8-1084>
- Carmona, P., & Coutin, L. (1998). Fractional Brownian motion and the Markov property. *Electronic Communications in Probability*, 3, 95–107. <https://doi.org/10.1214/ECP.v3-998>
- Carmona, P., Coutin, L., & Montseny, G. (2003). Stochastic integration with respect to fractional Brownian motion. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 39(1), 27–68. [https://doi.org/10.1016/S0246-0203\(02\)01111-1](https://doi.org/10.1016/S0246-0203(02)01111-1)
- Cheridito, P. (2001). Mixed fractional Brownian motion. *Bernoulli*, 7(6), 913–934. <https://doi.org/10.2307/3318626>
- Cheridito, P., Kawaguchi, H., & Maejima, M. (2003). Fractional Ornstein–Uhlenbeck processes. *Electronic Journal of Probability*, 8, 1–14. <https://doi.org/10.1214/EJP.v8-125>
- Cheridito, P., & Nualart, D. (2005). Stochastic integral of divergence type with respect to the fractional Brownian motion with Hurst parameter $H \in (0, \frac{1}{2})$. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 41(6), 1049–1081. <https://doi.org/10.1016/j.anihpb.2004.09.004>
- Cioczek-Georges, R., & Mandelbrot, B. B. (1995). A class of micropulses and antipersistent fractional Brownian motion. *Stochastic Processes and their Applications*, 60(1), 1–18. [https://doi.org/10.1016/0304-4149\(95\)00046-1](https://doi.org/10.1016/0304-4149(95)00046-1)
- Corcuera, J. M., Nualart, D., & Woerner, J. H. C. (2006). Power variation of some integral long-memory processes. *Bernoulli*, 12(4), 713–735. <https://doi.org/10.3150/bj/1155735933>
- Coutin, L., & Decreusefond, L. (1999). Abstract nonlinear filtering theory in the presence of fractional Brownian motion. *The Annals of Applied Probability*, 9(4), 1058–1090. <https://doi.org/10.1214/aoap/1029962865>
- Coutin, L., Nualart, D., & Tudor, C. A. (2001). Tanaka formula for the fractional Brownian motion. *Stochastic Processes and their Applications*, 94(2), 301–315. [https://doi.org/10.1016/S0304-4149\(01\)00085-0](https://doi.org/10.1016/S0304-4149(01)00085-0)
- Dai, W., & Heyde, C. C. (1996). Itô's formula with respect to fractional Brownian motion and its application. *Journal of Applied Mathematics and Stochastic Analysis*, 9(4), 439–448. <https://doi.org/10.1155/S104895339600038X>
- Decreusefond, L., & Üstünel, A. S. (1998). Fractional Brownian motion: Theory and applications. *ESAIM: Proceedings*, 5, 75–86. <https://doi.org/10.1051/proc:1998014>
- Decreusefond, L., & Üstünel, A. S. (1999). Stochastic analysis of the fractional Brownian motion. *Potential Analysis*, 10, 177–214. <https://doi.org/10.1023/A:1008634027843>
- Duncan, T. E. (2001). Some aspects of fractional Brownian motion. *Nonlinear Analysis: Theory, Methods & Applications*, 47(7), 4775–4782. [https://doi.org/10.1016/S0362-546X\(01\)00589-2](https://doi.org/10.1016/S0362-546X(01)00589-2)
- Duncan, T. E., Pasik-Duncan, B., & Maslowski, B. (2002). Fractional Brownian motion and stochastic equations in Hilbert spaces. *Stochastics and Dynamics*, 2(2), 225–250. <https://doi.org/10.1142/S0219493702000340>
- Dzhaparidze, K., & van Zanten, H. (2004). A series expansion of fractional Brownian motion. *Probability Theory and Related Fields*, 130(1), 39–55. <https://doi.org/10.1007/s00440-003-0310-2>
- Dzhaparidze, K., & van Zanten, H. (2005). Krein's spectral theory and the Paley-Wiener expansion for fractional Brownian motion. *The Annals of Probability*, 33(2), 620–644. <https://doi.org/10.1214/009117904000000955>
- Jost, C. (2006). Transformation formulas for fractional Brownian motion. *Stochastic Processes and their Applications*, 116(8), 1341–1357. <https://doi.org/10.1016/j.spa.2006.02.006>
- Malyarenko, A., Mishura, Y., Ralchenko, K., & Shklyar, S. (2023). Entropy and alternative entropy functionals of fractional Gaussian noise as the functions of Hurst index. *Fractional Calculus and Applied Analysis*, 26, 1052–1081. <https://doi.org/10.1007/s13540-023-00155-2>
- Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4), 422–437. <https://doi.org/10.1137/1010093>
- Mishura, Y., Ralchenko, K., & Shklyar, S. (2020). General Conditions of Weak Convergence of Discrete-Time Multiplicative Scheme to Asset Price with Memory. *Risks*, 8(1), 11. <https://doi.org/10.3390/risks8010011>
- Norros, I., Valkeila, E., & Virtamo, J. (1999). An Elementary Approach to a Girsanov Formula and Other Analytical Results on Fractional Brownian Motions. *Bernoulli*, 5(4), 571–587. <https://doi.org/10.2307/3318691>
- Nourdin, I. (2012). *Selected Aspects of Fractional Brownian Motion*. Springer. <https://doi.org/10.1007/978-88-470-2823-4>
- Novikov, A., & Valkeila, E. (1999). On Some Maximal Inequalities for Fractional Brownian Motions. *Statistics & Probability Letters*, 44(1), 47–54. [https://doi.org/10.1016/S0167-7152\(98\)00290-9](https://doi.org/10.1016/S0167-7152(98)00290-9)
- Nualart, D. (2003). Stochastic integration with respect to fractional Brownian motion and applications. *Contemporary Mathematics*, 336, 3–40.
- Nualart, D., & Ouknine, Y. (2003). Stochastic differential equations with additive fractional noise and locally unbounded drift. In Giné, Houdré, & Nualart (Eds.), *Stochastic Inequalities and Applications* (Vol. 56, pp. 353–365). Birkhäuser. https://doi.org/10.1007/978-3-0348-8069-5_20
- Pipiras, V., & Taqqu, M. S. (2017). *Long-Range Dependence and Self-Similarity*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139600347>
- Samorodnitsky, G. (2016). *Stochastic processes and long range dependence*. Springer Cham. <https://doi.org/10.1007/978-3-319-45575-4>

Отримано редакцією журналу / Received: 05.06.25

Прорецензовано / Revised: 24.08.25

Схвалено до друку / Accepted: 10.10.25

Євгеній БУНДЮК, студ.
ORCID ID: 0009-0006-2130-6705
e-mail: minakami1984@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

Юлія МІШУРА, д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0002-6877-1800
e-mail: yuliyamishura@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

Енріка ПІРОЦЦІ, проф.
ORCID ID: 0000-0001-8692-2759
e-mail: enrica.pirozzi@unicampania.it
Університет Кампанії Луїджі Ванвітеллі, Казерта, Італія

ФОРМУЛИ ПРОГНОЗУ ДЛЯ ІНТЕГРОВАНОГО ДРОБОВОГО БРОУНІВСЬКОГО РУХУ

Досліджено проєкцію (прогноз) приростів інтегрованого дробового броунівського руху (ідБР). Уведено ідБР, обчислено його коваріаційну функцію та встановлено стаціонарність його приростів. Основна мета роботи — визначити коефіцієнти для лінійного прогнозу майбутнього приросту ідБР на основі ряду минулих приростів. Показано, що коефіцієнти прогнозу демонструють складну поведінку, зі зміною знака залежно від показника Хюрста (H). Для приростів на інтервалах, що перетинаються, ця зміна знака відбувається за нетривіальних значень H . На відміну від цього, для приростів на інтервалах, що не перетинаються, зміна знака відбувається за $H=0.5$, що узгоджується з властивостями самого дробового броунівського руху. Дослідження надає явні формули, де це можливо, та велику кількість таблиць із чисельно обрахованими значеннями коефіцієнтів проєкції для різних значень індексу Хюрста, пропонуючи розуміння властивостей і прогнозу ідБР, що закладає основу для подальших досліджень.

Ключові слова: дробовий броунівський рух, інтегрований дробовий броунівський рух, коваріаційна функція, задача проєкції, коефіцієнти проєкції.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

ПРОГНОСТИЧНА МОДЕЛЬ ЛОГІСТИЧНОЇ РЕГРЕСІЇ ДЛЯ ОЦІНЮВАННЯ ПОКАЗАНЬ ДО ЛІКУВАЛЬНОЇ ІМПЛАНТАЦІЇ ПРИ ПРОЛАБУВАННІ ШЛУНКА У СТРАВОХІД АБО КАРДІАЛЬНУ ДІЛЯНКУ ШЛУНКА

Побудовано прогностичну модель на основі логістичної регресії для оцінювання показань до лікувальної імплантації у пацієнтів із пролабуванням шлунка у стравохід або кардіальну ділянку шлунка на основі даних 558 пацієнтів ендоскопічного відділення Кіровоградської обласної лікарні з клінічними симптомами й ендоскопічними характеристиками. Модель ураховує дев'ять предикторів: скарги на печію, регургітацію, тиск за грудиною, осиплість голосу, наявність дефектів слизової за Лос-Анджелівською класифікацією, висоту пролабування шлунка у стравохід або кардіальну ділянку шлунка, положення Z-лінії та наявність ерозій і виразок, укритих гематином. Під час побудови логістичної регресійної моделі особливу увагу приділено медичній інтерпретації отриманих коефіцієнтів, стандартних помилок, статистик Вальда, матриць варіацій, відношень шансів та інших показникам моделі. Це дало змогу не лише оцінити статистичну значущість предикторів, а й зрозуміти їхнє клінічне значення у прогнозуванні можливості імплантації в конкретного пацієнта. Для аналізу залежностей між предикторами були використані таблиці спряженості та коефіцієнти Крамера, що допомогло виявити слабкі й помірні зв'язки між симптомами та зробити висновок про невилучення всіх предикторів з моделі. Для оцінювання пояснювальної здатності моделі використано коефіцієнти детермінації Кокса та Снелла, модифікований коефіцієнт детермінації Нейджелкера, тест Хосмера – Лемешова та класифікаційну таблицю за моделлю. Результати демонструють високу прогностичну здатність моделі (точність 98,6 %) і підтверджують клінічну значущість навіть рідкісних симптомів.

Ключові слова: логістична регресія; прогностична модель, медична імплантація, пролабування шлунка, ендоскопічні предиктори, клінічне рішення.

Класифікація відповідно до AMS 2020: 62J12, 62P10.

Вступ

Задачі медичної діагностики та прогнозування належать до класу задач класифікації, де на основі множини клініко-ендоскопічних змінних необхідно визначити належність пацієнта до однієї з категорій. Математичне моделювання цього процесу вимагає використання апарату ймовірно-статистичних методів, що дає змогу перетворити набір емпіричних даних у формалізовану систему прийняття рішень.

Класичним підходом до таких задач є використання логістичної регресії (Hosmer, Lemeshow, & Sturdivant, 2013), що забезпечує відображення простору предикторів у простір ймовірностей за допомогою логіт-функції. Важливою властивістю логістичної регресії є можливість інтерпретації коефіцієнтів у вигляді відношення шансів, що робить її не лише інструментом прогнозування, але й засобом оцінювання значущості кожного окремого предиктора. Моделі логістичної регресії використовуються в медицині як цінний математико-статистичний апарат для виявлення хвороби Паркінсона (Verma, & Sinha, 2025); прогнозування післяопераційної смертності у хворих на рак легенів (Bernard et al., 2024); прогнозування ймовірності гострого коронарного синдрому (Hua et al., 2025) та ін. Ключові можливості застосування логістичної регресії в медицині наведені в роботі (Schober, & Vetter, 2021).

У пропонованій статті буде побудовано математичну модель, що інтегрує клінічні й ендоскопічні дані для прогнозування показань і протипоказів до імплантації діагностичних та/або терапевтичних пристроїв, зокрема систем Bravo, JSPH-1, BEST Capsule, EndoStim для стимуляції нижнього стравохідного сфінктера, капсульних датчиків Smart papowire та ін. (Lawenko, & Lee, 2016; Li, Liu, & Tao, 2013) у пацієнтів із патологією шлунково-стравохідного з'єднання, з метою підвищення точності й об'єктивності клінічного ухвалення рішень.

1. Постановка задачі

На основі вибірки із 558 пацієнтів ендоскопічного відділення комунального некомерційного підприємства "Кіровоградська обласна лікарня Кіровоградської обласної ради" та їхніх клінічних характеристик, лікарі, на основі детального аналізу їхнього стану, змогли оцінити можливість імплантації / не імплантації лікувального пристрою у пацієнтів із пролабуванням шлунка у стравохід або кардіальну ділянку шлунка. Основними умовами для імплантації були такі: наявність скарг на печію, відчуття регургітації та тиску за грудиною, осиплість голосу (поєднання скарг у контексті дозволу імплантації визначали лікарі); показник висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка має дорівнює або бути меншим за показник Z-лінії; відсутність дефектів типу C і D за Лос-Анджелівською класифікацією оцінки рефлюкс-езофагіту; відсутність ерозій і виразок, укритих гематином у разі дефектів типу A і B й інші менш значущі характеристики, які лікарі визначали у процесі обстеження пацієнтів.

Мета пропонованої статті полягає у знаходженні моделі, яка на основі ряду отриманих клінічних показників й отриманих раніше результатів (заключень) лікарів щодо імплантації / не імплантації лікувального пристрою, зможе на основі клінічних показників нового пацієнта надати відповідь про можливість імплантації пристрою.

Проаналізуємо, наскільки логістична регресія на початкових даних здатна відтворити чіткі клінічні правила класифікації пацієнтів щодо можливості імплантації. Коефіцієнти логістичної регресії покажуть, які початкові ознаки реально впливають на результат, навіть без явного кодування правил.

2. Основні результати

У логістичній регресії предиктори (пояснювальні змінні) позначено як $x_1, x_2, \dots, x_p$, де p – кількість предикторів (у нашій задачі $p = 9$).

Вхідні змінні моделі (предиктори): 1) скарги на печію (бінарна змінна: 1 – наявні; 0 – відсутні) – x_1 ; 2) скарги на відчуття регургітації (бінарна змінна: 1 – наявні; 0 – відсутні) – x_2 ; 3) скарги на відчуття тиску за грудиною (бінарна змінна: 1 – наявні; 0 – відсутні) – x_3 ; 4) скарги на осиплість голосу (бінарна змінна: 1 – наявні; 0 – відсутні) – x_4 ; 5) наявність дефектів слизової, класифікованих за Лос-Анджелівською класифікацією оцінки рефлюкс-езофагіту (порядкова змінна: дефекти відсутні – 0; тип А – 1; тип В – 2; тип С – 3; тип D – 4) – x_5 ; 6) показник висоти пролабування шлунка в стравохід або кардіальну ділянку шлунка (кількісна змінна, см) – x_6 ; 7) показник розташування Z-лінії (кількісна змінна, см) – x_7 ; 8) наявність ерозій, укритих гематином (бінарна змінна: 1 – наявні; 0 – відсутні) – x_8 ; 9) наявність виразок, укритих гематином (бінарна змінна: 1 – наявні; 0 – відсутні) – x_9 . Результуюча змінна Y : можливість імплантації (1 – так; 0 – ні). У нашому випадку $i = \overline{1, n}$, де $n = 558$ – кількість пацієнтів).

Нехай x_i – вектор предикторів для i -го спостереження (пацієнта): $x_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})'$, де перший елемент 1 відповідає вільному члену моделі β_0 . Моделюємо ймовірність позитивного результату (успіху, настання події – імплантації пристрою) як: $p_i = f(x_i; \beta) = \exp \exp(x_i' \beta) / (1 + \exp \exp(x_i' \beta))$, де $p_i = E[Y_i]$ – це очікувана ймовірність успіху (імплантації) для i -го спостереження (пацієнта); β – вектор коефіцієнтів моделі.

Далі логістичну функцію побудуємо як логарифмічну функцію шансів (Nguyen, 2020):

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = x_i' \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip},$$

де величина $p_i/(1-p_i)$ представляє шанси на успіх (співвідношення ймовірності успіху до ймовірності невдачі).

Отже, ймовірність p_i – це ймовірність успіху, яка показує, наскільки ймовірно, що для i -го об'єкта (пацієнта) подія настане (буде показання до імплантації). А величина $p_i/(1-p_i)$ показує у скільки разів ймовірніший успіх за невдачу (від 0 до ∞). Далі натуральний логарифм $\ln(p_i/(1-p_i))$ перетворює шанси на шкалу від $-\infty$ до $+\infty$, що зручно для моделювання.

Оскільки Y_i відповідає розподілу Бернуллі з ймовірністю p_i , то функція правдоподібності для n незалежних спостережень матиме вигляд

$$L(\beta) = L(p_i) = \prod_{i=1}^n p_i^{Y_i} (1-p_i)^{1-Y_i}, \quad p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}, \quad 1-p_i = \frac{1}{1+\exp(\eta_i)}, \quad \eta_i = x_i' \beta.$$

Якщо $y_i = 1$, у добуток іде p_i . Якщо $y_i = 0$, у добуток іде $(1-p_i)$. Отже, функція $L(\beta)$ показує, "наскільки добре" модель з коефіцієнтами β описує спостережувані дані.

Зручніше працювати з логарифмом правдоподібності (лог-правдоподібністю), щоб перетворити добуток у суму:

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^n [y_i \ln p_i + (1-y_i) \ln (1-p_i)].$$

У пакеті IBM SPSS Statistics закладені алгоритми Ньютона – Рафсона й алгоритм Фішера, що дають змогу отримати оцінки максимальної правдоподібності параметрів β . За достатньо великої вибірки розподіл $\hat{\beta}$ (оцінок коефіцієнтів) наближається до багатовимірного нормального розподілу $\hat{\beta} \sim AN(\beta, [I(\beta)]^{-1})$, де $I(\beta)$ – інформаційна матриця за Фішером, що визначає дисперсійно-коваріаційну структуру $\hat{\beta}$. Її визначають як $I(\beta) = E \left[\frac{\partial l(\beta)}{\partial \beta} \frac{\partial l(\beta)}{\partial \beta'} \right]$.

Оцінки коефіцієнтів $\hat{\beta} = l(\beta)$ перебувають в основі ітераційних чисельних методів.

Обчислюємо похідні лог-правдоподібності. Для кожного i : $\frac{\partial l}{\partial \eta_i} = y_i - p_i$.

У матричній формі градієнт записуємо як $U(\beta) = \nabla_{\beta} l(\beta) = X^T (y - p)$, де X – матриця $n \times p$, $p = (p_1, \dots, p_n)$, $y = (y_1, \dots, y_n)$. Матрицю Гессе розраховуємо так: $H(\beta) = \nabla_{\beta}^2 l(\beta) = -X^T W X$, де $W = \text{diag}(\omega_1, \dots, \omega_n)$ і $\omega_i = p_i(1-p_i)$. W – це діагональна матриця, у якій всі елементи поза головною діагоналлю дорівнюють нулю. Зауважимо, що більші ваги в контексті нашої задачі відповідатимуть пацієнтам, для яких модель менш упевнена (тобто p_i близько до 0,5). Менші ваги відповідають пацієнтам, для яких модель дуже впевнена (тобто p_i близько до 0 або 1). Іншими словами, ваги підкреслюють, наскільки "важливим" є конкретне спостереження для уточнення параметрів моделі на поточному кроці ітерації.

Формула Ньютона – Рафсона у нашому випадку матиме вигляд:

$$\beta^{(t+1)} = \beta^{(t)} - H^{-1}(\beta^{(t)}) U(\beta^{(t)}) = \beta^{(t)} - (-X^T W^{(t)} X)^{-1} X^T (y - p^{(t)}) = \beta^{(t)} + (X^T W^{(t)} X)^{-1} X^T (y - p^{(t)}).$$

Уведемо робочий відгук $z^{(t)} = X \beta^{(t)} + [W^{(t)}]^{-1} (y - p^{(t)})$. Тоді крок можна записати як $\beta^{(t+1)} = \arg \max_{\beta} (z^{(t)} - X \beta)^T W^{(t)} (z^{(t)} - X \beta)$, що дає нормальні рівняння $(X^T W^{(t)} X) \beta^{(t+1)} = X^T W^{(t)} z^{(t)}$.

Це показує, що кожна ітерація – це крок зваженого найменших квадратів з вагами ω_i .

Асимптотична матриця варіацій $\text{Var}(\hat{\beta}) \approx I^{-1}(\hat{\beta}) \approx -H^{-1}(\hat{\beta})$. Оскільки $H(\hat{\beta}) = -X^T \hat{W} X$, то $\text{Var}(\hat{\beta}) \approx (X^T \hat{W} X)^{-1}$.

Стандартна помилка j -го коефіцієнту – це квадратний корінь діагонального елементу цієї матриці: $SE(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)}$. Стандартна помилка показує, наскільки точно оцінено вплив певного предиктора на ймовірність імплантації. Низьке значення предиктора вказує на вплив цього предиктора. Велике значення предиктора вказує, що оцінка $\hat{\beta}_j$ нестабільна, тобто даних недостатньо або вони дуже варіабельні для надійного прогнозу. У клінічній практиці навіть нестабільний предиктор може бути важливим для ухвалення рішень, особливо якщо він теоретично обґрунтований (Matsui, Le-Rademacher, & Mandrekar, 2021).

Для будь-якого предиктора x_j коефіцієнт $\hat{\beta}_j$ пов'язаний із відношенням шансів (odds ratio): $OR_j = e^{\hat{\beta}_j}$. У контексті логістичної регресії відношення показує, як зміна предиктора впливає на шанси настання події. Інтерпретація: $OR = 1$ – немає впливу предиктора; $OR > 1$ – збільшення шансу настання події за зростання предиктора; $OR < 1$ – зменшення шансу.

Наведемо критерії збіжності в ітераційному процесі в пакеті IBM SPSS Statistics: 1) $\|\beta^{(t+1)} - \beta^{(t)}\|_{\infty} < \varepsilon$, де $\varepsilon = 0,001$; 2) досягнення максимум ітерацій (maxit), де maxit = 20.

Для аналізу значущості коефіцієнтів регресії (перевірки нульової гіпотези $H_0: \beta_j = 0$) застосовуватиме статистику Вальда $W_j = \hat{\beta}_j^2 / SE(\hat{\beta}_j)^2$. Розподіл χ^2 з 1 ступенем свободи дає змогу обчислити p -значення критерію. Ступені свободи для кожного коефіцієнта β_j $df = 1$, тобто тестуємо лише один параметр. Значущість (p -значення) показує ймовірність отримати таку статистику Вальда, якщо насправді $\beta_i = 0$: $p = 1 - F_{\chi^2}(W_j, df = 1)$, де F_{χ^2} – функція розподілу χ^2 з одним ступенем свободи. Мале p -значення ($< 0,05$) свідчить про те, що наявний предиктор x_j статистично значущо пов'язаний з ймовірністю успішної імплантації. Велике p -значення показує, що недостатньо доказів, щоб стверджувати вплив предиктора, але не означає, що його можна ігнорувати клінічно.

Побудуємо логістичну регресію для перерахованих клінічних показників у статистичному пакеті IBM SPSS Statistics. У табл. 1 наведено показники рівняння логістичної регресії та їхню значущість у моделі регресії. Також подано основну інформацію щодо медичної інтерпретації показників.

Таблиця 1

Показники змінних рівняння логістичної регресії та їхня медична інтерпретація

Клінічні показники	Оцінки коефіцієнтів логістичної регресії β_j	Середньо-квадратичні похибки, $SE(\hat{\beta}_j)$	Статистики Вальда, W_j	Ступені свободи, df	p -значення	Відношення шансів, OR_j	Медична інтерпретація показників
Скарги на печію	1,866	0,841	4,917	1	0,027	6,460	Наявність скарг збільшує шанси показань до імплантації у 6,5 раза. Значущий предиктор
Скарги на відчуття регургітації	-1,784	0,760	5,503	1	0,019	0,168	Наявність регургітації зменшує шанси у 6 разів ($1/0,168 = 5,95$). Змінна "перевертається" у знак "захисного фактора" штучно. Значущий предиктор
Скарги на відчуття тиску за грудиною	-17,030	19381,806	0,000	1	0,999	0,000	Коефіцієнт нестабільний. Предиктор статистично незначущий
Скарги на осиплість голосу	-0,659	1,133	0,338	1	0,561	0,517	Предиктор статистично незначущий
Наявність дефектів слизової, класи фікованих за Лос-Анджелівською класифікацією оцінки рефлюкс-езофагіту	-4,432	0,737	36,188	1	0,000	0,012	Наявність дефектів різко зменшує ймовірність (майже у 83 рази). Дуже сильний і значущий предиктор
Показник висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка	-1,054	0,364	8,408	1	0,004	0,348	Кожна одиниця зростання висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка (см) зменшує шанси імплантації майже у 3 рази ($1/0,348 = 2,87$). Значущий предиктор
Показник розташування Z-лінії	0,748	0,376	3,967	1	0,046	2,114	Збільшення показника Z-лінії підвищує ймовірність імплантації більш як у 2 рази. Значущий предиктор
Наявність ерозій, укритих гематином	-23,762	8127,078	0,000	1	0,998	0,000	Технічно незначущий предиктор через рідкість явища
Наявність виразок, укритих гематином	-23,247	6186,634	0,000	1	0,997	0,000	Технічно незначущий предиктор через рідкість явища
Константа	21,066	11,324	3,461	1	0,063	$1,41 \times 10^9$	Велика, але не значуща

Коефіцієнт регресії $\hat{\beta}_j$ – це оцінка впливу j -ї незалежної змінної на логіт-ймовірності. Для кожної змінної x_j значення $\hat{\beta}_j$ показує, як змінна на одиницю в X_j змінює $\logit(\hat{p}_i)$. Зауважимо, що для додатних значень $\hat{\beta}_j$ шанси на імплантацію зростають у $e^{\hat{\beta}_j}$ -разів. У нашому випадку рівняння логістичної регресії має такий вигляд:

$$\logit(\hat{p}_i) = 21,066 + 1,866x_{i1} - 1,784x_{i2} - 17,030x_{i3} - 0,659x_{i4} - 4,432x_{i5} - 1,054x_{i6} + 0,748x_{i7} - 23,762x_{i8} - 23,247x_{i9}. \quad (1)$$

Проаналізуємо показники якості логістичної моделі на основі коефіцієнтів детермінації Кокса та Снелла, а також коефіцієнта Нейджелкерка. Коефіцієнт детермінації $R^2_{C\&S}$ Кокса та Снелла показує відносно поліпшення лог-правдоподібності моделі з предикторами порівняно з моделлю лише з константою (Cox, & Snell, 1989). Обраховується як

$R_{C\&S}^2 = 1 - (L_0/L_m)^{\frac{2}{n}}$, де L_0 – правдоподібність за нульовою моделлю (з константою); L_m – правдоподібність за оціненою моделлю (з предикторами); n – кількість спостережень. Максимальне значення $R_{C\&S}^2 < 1$ (часто $\approx 0,75$ – $0,9$ навіть для ідеальної моделі). Тому використовуємо модифікований коефіцієнт детермінації Нейджелкера ($R_N^2 = R_{C\&S}^2 / (1 - L_0^{2/n})$), який масштабується так, щоб максимальне значення могло дорівнювати 1.

Наша модель характеризується високою пояснювальною здатністю, що підтверджується високим $R_N^2 = 0,903$, тобто модель відтворює майже всі варіації залежної змінної.

На основі тесту Хосмера – Лемешова проаналізуємо узгодженість спостережуваних (емпіричних) й отриманих теоретичних значень моделі. Для розрахунку статистики Хосмера – Лемешова χ_{HL}^2 спостережувані дані групуються у g -груп за прогнозованою ймовірністю (Hosmer, Lemeshow, & Sturdivant, 2013). Для кожної групи обчислюємо спостережувану кількість подій $\hat{E}_i$ і невідповідностей $(O_i - \hat{E}_i)$. Статистику критерію розраховуємо так: $\chi_{HL}^2 = \sum_{i=1}^g \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i (1 - \hat{E}_i/n_i)}$, де кількість ступенів свободи $df = g - 2$, а p -значення (значущість) розраховуємо через χ^2 -розподіл. У нашому випадку на стандартному рівні $0,05$ за тестом Хосмера – Лемешова модель є прийнятною ($p = 0,034 < 0,05$).

У контексті цього тесту проаналізуємо результати класифікації за отриманою моделлю. Для пацієнтів без показань до імплантації (0) модель правильно передбачила 111 зі 116 випадків, тобто $95,7\%$ правильних передбачень. П'ять пацієнтів були помилково передбачені як 1 (помилка I типу для цієї категорії). Для пацієнтів із показаннями до імплантації (1) модель правильно передбачила 438 із 441 випадку, тобто $99,3\%$ правильних передбачень. Три пацієнти були помилково класифіковані як 0 (помилка II типу). Модель правильно передбачає $98,6\%$ усіх випадків. Це дуже високий показник точності, що свідчить про сильну предиктивну здатність моделі.

Спробуємо проаналізувати, чи доцільно вилучати предиктори з отриманої моделі. Для цього між показниками скарг проаналізуємо залежність. Для оцінювання залежності скористаємося коефіцієнтом спряженості Крамера V (Sheskin, 2000): $V = \sqrt{\chi^2 / (N \cdot (k - 1))}$, де χ^2 – значення критерію Пірсона з таблиці спряженості; N – обсяг вибірки; $k = \min(r, c)$, де r – кількість рядків, c – кількість стовпців таблиці (у нашому випадку $r = c = 2$; $V \in [0,1]$).

Проаналізуємо наповнення таблиць спряженості та коефіцієнти Крамера для кожної пари скарг. 1) Печія та регургітація. Регургітація майже завжди поєднується з печією (лише 14 ізольованих випадків регургітації без печії). $V = 0,596$. Маємо помірний статистично значущий зв'язок. Регургітація є похідним виявом рефлюксних симптомів (печії). 2) Печія та тиск за грудиною. Скарги на тиск за грудиною практично відсутні (лише 3 ізольовані випадки та 196 у поєднанні з печією). У більшості випадків (359) обидві скарги відсутні. $V = 0,099$. Маємо слабкий, але значущий зв'язок. У великих вибірках слабкий зв'язок часто може бути значущим (Van den Berg, 2017). 3) Печія та осиплість голосу. Скарги на осиплість голосу зустрічаються рідко (20 випадків із 558), але переважно в пацієнтів із печією (14 проти 6 випадків без печії). Це свідчить, що осиплість голосу може бути вторинним виявом гастроєзофагеального рефлюксу. $V = 0,138$. Маємо слабкий, але значущий зв'язок. Осиплість голосу можна розглядати як ускладнений вияв печії. 4) Регургітація та тиск за грудиною. Усі випадки тиску за грудиною пов'язані з регургітацією: $V = 0,137$. Маємо слабкий, але значущий зв'язок. Симптоми "співіснують", проте залежність не така сильна, як у пари "печія – регургітація". 5) Регургітація й осиплість голосу. В 11 випадках регургітація поєднується з осиплістю проти 9 ізольованих випадків осиплості: $V = 0,151$. Маємо слабкий, але значущий зв'язок. Осиплість голосу може бути вторинним виявом регургітації. 6) Тиск за грудиною й осиплість голосу. Усі випадки осиплості виникають без тиску (20 ізольованих випадків осиплості): $V = 0,014$. Зв'язок статистично незначущий. Ці симптоми незалежні, можна розглядати їх окремо. Зауважимо, що між жодною парою змінних немає сильного статистичного зв'язку ($V > 0,6$). Отже, виводити якісь показники скарг із загальної моделі регресії не доцільно. Отже, навіть якщо зв'язки між деякими змінними слабкі, ці показники є важливими медичними предикторами і мають залишатися в моделі для збереження клінічної цінності прогнозування (Matsui, Le-Rademacher, & Mandrekar, 2021).

Що стосується предикторів наявності ерозій і виразок, укритих гематином, то у статистичному сенсі предиктори є "незначущими" через надзвичайну рідкість явища (стандартні похибки великі). Це типовий випадок проблеми "розріджених даних" (Greenland, Mansournia, & Altman, 2016). Проте їх не можна вважати "неважливими", бо вони мають високу клінічну специфічність: навіть поодинокі випадки таких уражень є показником тяжкості процесу та ризику ускладнень. У разі вилучення цих змінних модель втрачає клінічну релевантність: вона перестає "бачити" критичні, хоч і рідкісні стани, що одразу знижує її прогностичну цінність у практиці. Крім того, видалення таких змінних призводить до погіршення якості моделі. Це пов'язано з тим, що навіть рідкісні симптоми дають внесок у розмежування груп (напр., "немає патології" чи "важкі форми").

Дискусія і висновки

Отже, отримана модель логістичної регресії (1) найкраще відображає зв'язок між клінічними показниками і результуючою змінною щодо можливості імплантації лікувального пристрою. Зауважимо вплив значущих предикторів, які значно змінюють показання до імплантації пристрою. Наявність скарг на печію в пацієнта збільшує шанси показань до імплантації у $6,5$ разів. Наявність регургітації зменшує шанси у шість разів. Змінна "перевертається" у знак "захисного фактора" штучно. Проте, фактично, скарги на регургітацію значно впливають на показання до імплантації. Наявність дефектів слизової за Лос-Анджелівською класифікацією різко зменшує шанси до імплантації – майже у 83 рази. Дуже сильний і значущий предиктор. Кожен сантиметр зростання висоти пролабування шлунка у стравохід або кардіальну ділянку шлунка зменшує шанси імплантації майже у три рази. Збільшення показника Z-лінії підвищує ймовірність імплантації у понад два рази.

Прогностична модель логістичної регресії є значущою за критерієм Хосмера – Лемешова. Результати демонструють високу прогностичну здатність моделі (точність $98,6\%$) і підтверджують клінічну значущість навіть рідкісних симптомів.

Джерела фінансування. Фінансування дослідження здійснюється власним коштом автора та частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Bernard, A., Cottenet, J., Pages, P. B., & Quantin, C. (2024). Is the Validity of Logistic Regression Models Developed with a National Hospital Database Inferior to Models Developed from Clinical Databases to Analyze Surgical Lung Cancers? *Cancers*, 16(4), 734. <https://doi.org/10.3390/cancers16040734>
- Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). Chapman and Hall/CRC.
- Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ*, 352, i1981. <https://doi.org/10.1136/bmj.i1981>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- Hua, Y., Stead, T. S., George, A., & Ganti, L. (2025). Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. *Academic Medicine & Surgery*. <https://doi.org/10.62186/001c.131964>
- Lawenko, R. M. A., & Lee, Y. Y. (2016). Evaluation of gastroesophageal reflux disease using the Bravo capsule pH system. *Journal of Neurogastroenterology and Motility*, 22(1), 25–30. <https://doi.org/10.5056/jnm15130>
- Li, J. N., Liu, C. L., & Tao, X. H. (2013). Clinical utility and tolerability of JSPH-1 wireless esophageal pH monitoring system. *BMC Gastroenterology*, 13(10). <https://doi.org/10.1186/1471-230X-13-10>
- Matsui, S., Le-Rademacher, J., & Mandrekar, S. J. (2021). Statistical models in clinical studies. *Journal of Thoracic Oncology*, 16(5), 734–739. <https://doi.org/10.1016/j.jtho.2021.02.021>
- Nguyen, M. (2020). *A guide on data analysis*. Bookdown. https://bookdown.org/mike/data_analysis
- Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia & Analgesia*, 132(2), 365–366. <https://doi.org/10.1213/ANE.0000000000005247>
- Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Chapman & Hall/CRC. <https://dl.icdst.org/pdfs/files3/22a131fac452ed75639ed5b0680761ac.pdf>
- Van den Berg, R. G. (2017). *Cramér's V – What and why?* SPSS Tutorials. <https://www.spss-tutorials.com/cramers-v-what-and-why/>
- Verma, V., & Sinha, B. (2025). Logistic regression-based detection of Parkinson's disease. *Preprints*. <https://doi.org/10.20944/preprints202502.0239.v1>

References

- Bernard, A., Cottenet, J., Pages, P. B., & Quantin, C. (2024). Is the Validity of Logistic Regression Models Developed with a National Hospital Database Inferior to Models Developed from Clinical Databases to Analyze Surgical Lung Cancers? *Cancers*, 16(4), 734. <https://doi.org/10.3390/cancers16040734>
- Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). Chapman and Hall/CRC.
- Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ*, 352, i1981. <https://doi.org/10.1136/bmj.i1981>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- Hua, Y., Stead, T. S., George, A., & Ganti, L. (2025). Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. *Academic Medicine & Surgery*. <https://doi.org/10.62186/001c.131964>
- Lawenko, R. M. A., & Lee, Y. Y. (2016). Evaluation of gastroesophageal reflux disease using the Bravo capsule pH system. *Journal of Neurogastroenterology and Motility*, 22(1), 25–30. <https://doi.org/10.5056/jnm15130>
- Li, J. N., Liu, C. L., & Tao, X. H. (2013). Clinical utility and tolerability of JSPH-1 wireless esophageal pH monitoring system. *BMC Gastroenterology*, 13(10). <https://doi.org/10.1186/1471-230X-13-10>
- Matsui, S., Le-Rademacher, J., & Mandrekar, S. J. (2021). Statistical models in clinical studies. *Journal of Thoracic Oncology*, 16(5), 734–739. <https://doi.org/10.1016/j.jtho.2021.02.021>
- Nguyen, M. (2020). *A guide on data analysis*. Bookdown. https://bookdown.org/mike/data_analysis/
- Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia & Analgesia*, 132(2), 365–366. <https://doi.org/10.1213/ANE.0000000000005247>
- Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Chapman & Hall/CRC. <https://dl.icdst.org/pdfs/files3/22a131fac452ed75639ed5b0680761ac.pdf>
- Van den Berg, R. G. (2017). *Cramér's V – What and why?* SPSS Tutorials. <https://www.spss-tutorials.com/cramers-v-what-and-why/>
- Verma, V., & Sinha, B. (2025). Logistic regression-based detection of Parkinson's disease. *Preprints*. <https://doi.org/10.20944/preprints202502.0239.v1>

Отримано редакцією журналу / Received: 05.08.25
Прорецензовано / Revised: 03.10.25
Схвалено до друку / Accepted: 10.10.25

Oleksandr HALUSHKO, PhD Student
ORCID ID: 0009-0009-0275-6542
e-mail: health24@ukr.net
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

PROGNOSTIC LOGISTIC REGRESSION MODEL FOR ASSESSING INDICATIONS FOR THERAPEUTIC IMPLANTATION WITH GASTRIC PROLAPSE INTO THE ESOPHAGUS OR THE CARDIAC PART OF THE STOMACH

In this study, a prognostic model based on logistic regression was developed to assess the indications for therapeutic implantation in patients with gastric prolapse into the esophagus or the gastric cardia. The analysis was performed on clinical and endoscopic data from 558 patients of the Endoscopy Department of the Kirovohrad Regional Hospital. The model incorporated nine predictors: complaints of heartburn, regurgitation, retrosternal pressure, hoarseness, presence of mucosal defects according to the Los Angeles classification, height of gastric prolapse into the esophagus or the gastric cardia, the position of the Z-line, and the presence of hematin-covered erosions or ulcers. Special attention was devoted to the medical interpretation of the estimated coefficients, standard errors, Wald statistics, variance-covariance matrices, odds ratios, and other model diagnostics. This approach enabled not only the evaluation of the statistical significance of the predictors but also the understanding of their clinical relevance in predicting the feasibility of implantation in individual patients. To analyze dependencies among predictors, contingency tables and Cramér's V coefficients were applied, which revealed weak to moderate associations between symptoms and justified retaining all predictors in the model. To assess the significance of the model, the Cox and Snell coefficients of determination, the modified Nagelker coefficient of determination, the Hosmer – Lemeshow test, and the classification table for the model were used. The results demonstrate a high predictive ability of the model (accuracy of 98.6%) and confirm the clinical significance of even rare symptoms.

Key words: logistic regression, prognostic model, medical implantation, gastric prolapse, endoscopic predictors, clinical decision.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 517.5

DOI: <https://doi.org/10.17721/1812-5409.2025/2.5>

Volodymyr YELAHIN, PhD Student

ORCID ID: 0009-0001-6984-5090

e-mail: yelahin@imath.kiev.ua

Institute of Mathematics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

CONTINUOUS TRANSFORMATIONS PRESERVING TAILS OF NEGABINARY REPRESENTATIONS OF NUMBERS FROM UNIT INTERVAL

For the negabinary representation of numbers from the $[0; 1]$ interval, we consider continuous transformations of this interval (i.e., bijective mappings of the set onto itself) that preserve the tails of number representations. These transformations are constructed using left and right shift operators. It is proven that the set of such transformations, under the operation of composition, forms a non-commutative group, which contains a continuous subgroup of increasing transformations.

Key words: Negabinary representation of numbers, cylinder, left and right shift operators, tail set, continuous transformations of the interval, group of interval transformations.

AMS 2020 classification: 28D05, 28D15.

Introduction

The negabinary numeral system has long been used in electronics and programming (Rao, Rao, & Krishnamurthy, 1974; Grover et al., 1983; Cherri, & Kamal, 2004). The negabinary representation of numbers in the interval $[0; 1]$:

$$x = \frac{2}{3} + \sum_{n=1}^{\infty} \frac{\alpha_n}{(-2)^n} = \frac{2}{3} - \left(\frac{\alpha_1}{2} - \frac{\alpha_2}{2^2} + \frac{\alpha_3}{2^3} - \dots \right) \equiv \Delta_{\alpha_1 \alpha_2 \alpha_3 \dots}^{-2}$$

is metrically equivalent to the classical binary representation (Pratsiovytyi, 2017). Moreover, it is topologically equivalent to the continued fraction A_2 representation of numbers (Pratsiovytyi, 2022). These are the examples of two-symbol encoding, which has multiple applications in number metric theory, probability theory, dynamical systems theory, fractal analysis, etc (Galambos, 2006; Schweiger, 1995; Albeverio, Pratsiovytyi, & Torbin, 2004). The efficiency of using the negabinary representation has been convincingly demonstrated multiple times in applications within the constructive theory of locally complex functions (Pratsiovytyi, 2022).

It is known that the group of continuous transformations of the unit interval (bijections of the interval onto itself) contains many interesting subgroups, including those related to different coding systems for real numbers and fractals (Pratsiovytyi, 2022). The cardinality of such subgroups remained mysterious for some time. In works (Osaulenko, 2016; Pratsiovytyi, Lysenko, & Maslova, 2020; Isaieva, & Pratsiovytyi, 2016; Pratsiovytyi, Chuikov, & Skrypnyk, 2019), only their infinitude was established. However, in (Pratsiovytyi, Lysenko, & Ratushniak, 2024), the continuity (cardinality of the continuum) of the group of transformations preserving the tails of Q_2 -representations was proven. In this paper, we obtain analogous results for transformations that preserve the tails of the negabinary representation of numbers.

1. Negabinary representation of an integer

The negabinary representation of an integer x is (Pratsiovytyi, Goncharenko, & Lysenko, 2015; Kasatkin, 1982) its expression in the form

$$\alpha_1 \cdot (-2)^0 + \alpha_2 \cdot (-2)^1 + \dots + \alpha_n \cdot (-2)^{n-1} + \dots \equiv (\alpha_1 \alpha_2 \dots \alpha_n \dots)_{-2},$$

where α_i take values of 0 or 1. For any integer x , there exists a unique sequence $(\alpha_1, \dots, \alpha_{2n})$ such that

$$x = 1 \cdot (-2)^{2n} + \alpha_{2n} \cdot (-2)^{2n-1} + \dots + \alpha_1 \cdot (-2)^0.$$

In this representation, the negabinary encoding of any natural number always contains an odd number of digits.

Example. : $9_{10} = 11001_{-2}$, $-3_{10} = 1101_{-2}$

Let $A \equiv \{0, 1\}$ be the alphabet of the binary system, and let $L \equiv A \times A \times \dots$ denote the space of sequences of elements from this alphabet (the space of sequences consisting of 0 and 1). It is known (Pratsiovytyi, 2022) that for any $x \in [0, 1]$, there exists a sequence $(\alpha_n) \in L$ such that

$$x = \frac{2}{3} + \sum_{n=1}^{\infty} \frac{\alpha_n}{(-2)^n} = \frac{2}{3} - \left(\frac{\alpha_1}{2} - \frac{\alpha_2}{2^2} + \frac{\alpha_3}{2^3} - \dots \right) \equiv \Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2} \quad (1)$$

The representation of x by the series (1) is called the negabinary representation, and the symbolic notation $\Delta_{\alpha_1 \alpha_2 \alpha_3 \dots \alpha_n \dots}^{-2}$ is referred to as the negabinary encoding.

Remark 1. In (Pratsiovytyi, 2017) the relationship between the negabinary and the classic binary representation of the same number $x = \Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2} = \Delta_{\omega_1 \omega_2 \dots \omega_n \dots}^2$ is given by the formulas:

$$\alpha_n = \begin{cases} 1 - \omega_n & \text{if } n \text{ is odd,} \\ \omega_n & \text{if } n \text{ is even.} \end{cases}$$

Thus,

$$\Delta_{\omega_1 \omega_2 \omega_3 \dots}^{-2} = \Delta_{[1-\omega_1] \omega_2 [1-\omega_3] \dots}^{-2}.$$

Lemma 1. (Pratsiovytyi, Goncharenko, & Lysenko, 2015) Numbers x and x' that have the following negabinary representations:

$$x = \Delta_{c_1 \dots c_n 0(01)}^{-2} \quad \text{and} \quad x' = \Delta_{c_1 \dots c_n 1(10)}^{-2}$$

are equal.

Numbers that have exactly two negabinary representations are called (Pratsiovytyi, Goncharenko, & Lysenko, 2015) negabinary-rational (NDR) numbers. The set of negabinary-rational numbers forms a dense subset of the rational numbers in $[0, 1]$, but not every rational number is negabinary-rational. The set of rational numbers that are not negabinary-rational is also dense in $[0, 1]$. Numbers from the interval $[0, 1]$ that are not negabinary-rational are called negabinary-irrational. Each digit in the negabinary representation of a negabinary-irrational number is a correctly defined function of the represented number.

A cylinder set of rank n with base $\alpha_1 \alpha_2 \dots \alpha_n$, corresponding to the negabinary representation of numbers in the interval $[0; 1]$, is defined (Pratsiovytyi, 2017) as the set

$$\Delta_{\alpha_1 \alpha_2 \dots \alpha_n}^{-2} = \{x = \Delta_{\alpha_1 \alpha_2 \dots \alpha_n \beta_1 \beta_2 \dots}^{-2}, (\beta_n) \in L\}.$$

The cylinder set is given by the interval

$$\Delta_{\alpha_1 \alpha_2 \dots \alpha_n}^{-2} = [u - v; u + w],$$

where

$$u = \frac{2}{3} + \sum_{i=1}^n \frac{\alpha_i}{(-2)^i},$$

and

$$v = \begin{cases} \frac{1}{3 \cdot 2^n}, & \text{if } n \text{ is odd,} \\ \frac{1}{3 \cdot 2^{n-1}}, & \text{if } n \text{ is even;} \end{cases} \quad w = \begin{cases} \frac{1}{3 \cdot 2^{n-1}}, & \text{if } n \text{ is odd,} \\ \frac{1}{3 \cdot 2^n}, & \text{if } n \text{ is even.} \end{cases}$$

2. Special functions of the negabinary representation of the fractional part of a real number

The left shift operator for negabinary digit sequences is defined as:

$$\eta(\Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2}) = \Delta_{\alpha_2 \alpha_3 \dots \alpha_n \dots}^{-2}$$

This operator is well-defined only when a convention is adopted to use one of the two representations of negabinary-rational (NDR) numbers (e.g., the representation with period (01)). The analytical expression of the left shift operator is:

$$\eta(x) = -2x + 2 - \alpha_1(x);$$

$$\eta(x) = \begin{cases} -2x + 2, & \text{if } x \in \Delta_0^{-2}, \\ -2x + 1, & \text{if } x \in \Delta_1^{-2}. \end{cases}$$

The operator is linear on each first-rank cylinder.

The right shift operator with parameter i , where $i \in \{0, 1\}$, is defined as:

$$\tau_i(x = \Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2}) = \Delta_{i \alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2}.$$

The function has the following form:

$$\tau_i(x) = \begin{cases} -\frac{x}{2} + 1, & \text{if } i = 1, \\ -\frac{x}{2} + \frac{1}{2}, & \text{if } i = 0. \end{cases}$$

The n -multiple left shift operator for negabinary representation is defined as:

$$\eta^n(\Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2}) = \eta(\eta^{n-1}(\Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2})) = \Delta_{\alpha_{n+1} \alpha_{n+2} \dots}^{-2}.$$

The analytical expression is:

$$\eta^n(x) = (-2)^n \cdot x + a,$$

where

$$a = 2 - 2^{n-1} \cdot \alpha_1 + 2^{n-2} \cdot \alpha_2 + \dots + (-1)^n \cdot \alpha_n.$$

The n -multiple right shift operator with the parameter set $(i_1, i_2, \dots, i_n)$ is defined as:

$$\tau_{i_1 i_2 \dots i_n}^n(\Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2}) = \tau_{i_1}(\tau_{i_2 \dots i_n}^{n-1}(\Delta_{\alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2})) = \Delta_{i_1 i_2 \dots i_n \alpha_1 \alpha_2 \dots \alpha_n \dots}^{-2}.$$

The analytical expression is:

$$\tau_{i_1 i_2 \dots i_n}^n(x) = \frac{x}{(-2)^n} + a,$$

where

$$a = \frac{2}{3} - \frac{2}{3} \left(\frac{1}{-2} \right)^n - \frac{i_1}{2} + \frac{i_2}{4} + \dots + \frac{i_n}{(-2)^n}.$$

3. Continuous transformations of the interval $[0; 1]$ that preserve tails

Recall that a transformation of a set is a bijective mapping (i.e., both injective and surjective) of that set onto itself. A continuous transformation of the interval $[0; 1]$ is a continuous increasing or decreasing function such that either $f(0) = 0$, $f(1) = 1$, or $f(0) = 1$, $f(1) = 0$.

The negabinary representations of numbers $\Delta_{\alpha_1 \alpha_2 \dots \alpha_n}^{-2}$ and $\Delta_{\beta_1 \beta_2 \dots \beta_n}^{-2}$ have the same tails if there exist natural numbers n and m such that $\alpha_{n+j} = \beta_{m+j}$ for all $j \in \mathbb{N}$.

A function f preserves the tails of the negabinary representation of numbers if there exist natural numbers n and m such that $a_{n+j}(x) = a_{m+j}(f(x))$ for all $j \in \mathbb{N}$.

Theorem 1. The set G of all continuous transformations of the interval $[0; 1]$ that preserve the tails of negabinary representations of numbers, under the operation of composition, forms a continuous non-commutative group, whose continuous subgroup consists of increasing transformations.

Proof. It is known that all continuous transformations of the interval $[0; 1]$ form a group under the operation of composition, where the identity transformation is the neutral element, and the inverse transformation serves as the symmetric (inverse) element for a given transformation. Clearly, the set G is closed under the operation of composition, and if $f \in G$, then $f^{-1} \in G$. Therefore, G forms a group by the subgroup criterion.

To prove that the group G is non-commutative, consider the transformations $f_1(x)$ and $f_2(x)$, defined by the following expressions:

$$f_1(x) = \begin{cases} \tau_1(x), & \text{when } 0 \leq x \leq \frac{2}{3} = \Delta_{(0)}^{-2}, \\ \eta(x), & \text{when } \frac{2}{3} \leq x \leq 1. \end{cases}$$

$$f_2(x) = \begin{cases} \tau_{10}(x), & \text{when } 0 \leq x \leq \frac{4}{5} = \Delta_{(0110)}^{-2}, \\ \eta^2(x), & \text{when } \frac{4}{5} \leq x \leq 1. \end{cases}$$

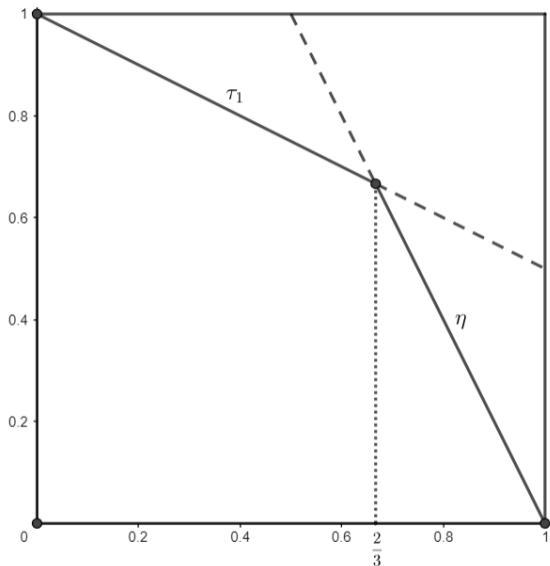


Fig. 1. $f_1(x)$

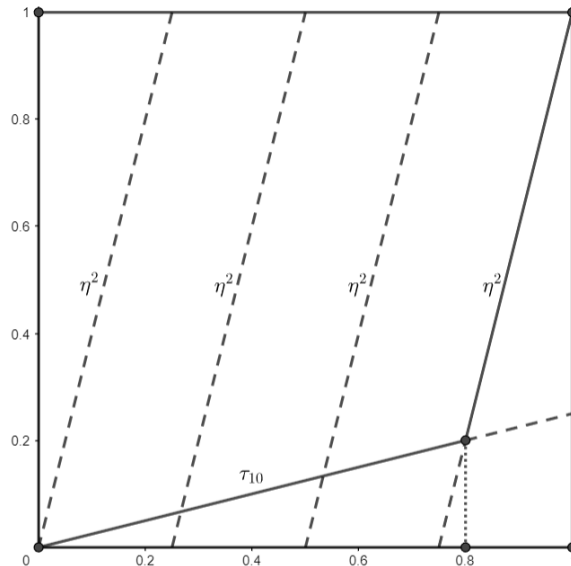


Fig. 2. $f_2(x)$

It is easy to see that $f_1(x)$ is a continuous strictly decreasing transformation, since τ_0 is a continuous and decreasing function, $\eta(x)$ is continuous and decreasing on the cylinder Δ_0^{-2} , and $\frac{2}{3}$ is a solution of the equation $\tau_0(x) = \eta(x)$. Similarly, $f_2(x)$ is a continuous strictly increasing transformation, since τ_{10} is continuous and increasing, $\eta^2(x)$ is continuous and increasing on the cylinder Δ_{01}^{-2} , and $\frac{4}{5}$ is a solution of the equation $\tau_{10}(x) = \eta^2(x)$.

Let $x_0 = \Delta_{0111(0)}^{-2} = \frac{41}{48}$ then:

$$f_1(f_2(x_0)) = f_1(\Delta_{11(0)}^{-2}) = f_1\left(\frac{5}{12}\right) = \Delta_{111(0)}^{-2} = \frac{7}{24},$$

$$f_2(f_1(x_0)) = f_2(\Delta_{111(0)}^{-2}) = f_2\left(\frac{7}{24}\right) = \Delta_{10111(0)}^{-2} = \frac{7}{96}.$$

Thus, $f_1(f_2(x_0)) \neq f_2(f_1(x_0))$.

Let us prove that the group is infinite. Consider an infinite sequence of functions f_3 :

$$f_3(x) = \begin{cases} \tau_{\overbrace{1010\dots10}^{2k}}, & \text{when } 0 \leq x \leq x_k, \\ \eta^{2k}(x), & \text{when } x_k \leq x \leq 1, \end{cases}$$

where $x_k = \Delta_{\overbrace{0101\dots01}^{2k}\overbrace{1010\dots10}^{2k}}$, depending on a natural parameter k .

Each function in this sequence is continuous, strictly increasing, and preserves the tails of the negabinary representation of numbers. Therefore, the group G is infinite. $\square$

Lemma 2. The equation $\eta^n(x) = \tau_{i_1\dots i_n}(x)$ has 2^n solutions $\Delta_{(\alpha_1\dots\alpha_n i_1\dots i_n)}^{-2}$, where $\alpha_j \in \{0, 1\}$, $j = \overline{1, n}$, and the equation $\eta^n(x) = \tau_{i_1\dots i_n}(x) = x$ has exactly one solution $\Delta_{(i_1 i_2 \dots i_n)}^{-2}$.

Proof. Let us consider the equation $\eta^n(x) = \tau_{i_1\dots i_n}(x)$. A number $x = \Delta_{\alpha_1\alpha_2\dots\alpha_n\dots}^{-2}$ is a solution of this equation if and only if the following system holds:

$$\begin{cases} i_1 = \alpha_{n+1} = \alpha_{3n+1} = \dots \\ i_2 = \alpha_{n+2} = \alpha_{3n+2} = \dots \\ \dots \\ i_n = \alpha_{2n} = \alpha_{4n} = \dots \\ \alpha_1 = \alpha_{2n+1} = \alpha_{4n+1} = \dots \\ \alpha_2 = \alpha_{2n+2} = \alpha_{4n+2} = \dots \\ \dots \\ \alpha_n = \alpha_{3n} = \alpha_{5n} = \dots \end{cases}$$

The solutions to this system are the numbers $\Delta_{(\alpha_1\dots\alpha_n i_1\dots i_n)}^{-2}$, where $(\alpha_1, \alpha_2, \dots, \alpha_n) \in A^n$. Therefore, in each cylinder of rank n , there is exactly one point that belongs to the set of solutions of the equation $\eta^n(x) = \tau_{i_1\dots i_n}(x)$. Among all the solutions of the given equation, only one point $x = \Delta_{(i_1 i_2 \dots i_n)}^{-2}$ solves the equation $\tau_{i_1\dots i_n}(x) = x$. $\square$

Introduce a sequence of points:

$$\{0; \Delta_{(11)}^{-2}; \Delta_{(00)}^{-2}; \Delta_{(0111)}^{-2}; \Delta_{(0100)}^{-2}; \dots \Delta_{(01\dots0111)}^{-2}; \Delta_{(01\dots0100)}^{-2}; \dots; 1\}$$

and denote them by $w_0, w_1, \dots, w_j, \dots$

The first point of this sequence is 0, the second point will be $\frac{1}{3}$, the third point will be $\frac{2}{3}$, and the next following points will also be the solutions of the equations $\tau_{\overbrace{0101\dots0111}^{2n}}(x) = x = \eta^{2n}(x) = w_{2n}$, $\tau_{\overbrace{0101\dots0100}^{2n}}(x) = x = \eta^{2n}(x) = w_{2n+1}$, where $n \in \mathbb{N}$.

Define the functions $f_T(x)$, $f_{00n}(x)$, $f_{11n}(x)$ on the interval $[w_{2n-1}; w_{2n}]$ as follows:

$$\begin{aligned} f_T(x) &= x; \\ f_{00n}(x) &= \begin{cases} \eta^{2n}(x), & \text{when } w_{2n-1} \leq x \leq t_{00}, \\ \tau_{\overbrace{01\dots0100}^{2n}}(x), & \text{when } t_{00} \leq x \leq w_{2n}; \end{cases} \\ f_{11n}(x) &= \begin{cases} \tau_{\overbrace{01\dots0111}^{2n}}(x), & \text{when } w_{2n-1} \leq x \leq t_{11}, \\ \eta^{2n}(x), & \text{when } t_{11} \leq x \leq w_{2n}; \end{cases} \end{aligned}$$

where t_{00}, t_{11} are the intersection points of the $2n$ -multiple left shift and right shift operators, which have the following representations:

$$t_{00} \equiv \Delta_{\overbrace{(01\dots0111)}^{2n}\overbrace{01\dots0100}^{2n}}^{-2}, t_{11} \equiv \Delta_{\overbrace{(01\dots0100)}^{2n}\overbrace{01\dots0111}^{2n}}^{-2}.$$

Define a sequence (a_n) , where $a_i \in \{0, 1, 2\}$. Then, define function $f_{a_n}(x)$, that is a piecewise-defined function and on each interval $[w_{2n-1}; w_{2n}]$ is defined as follows:

$$f_{a_n}(x) = \begin{cases} f_{00n}(x), & \text{if the } n\text{-th element of the sequence } (a_n) \text{ is equal to } 0, \\ f_{11n}(x), & \text{if the } n\text{-th element of the sequence } (a_n) \text{ is equal to } 1, \\ f_T(x), & \text{if the } n\text{-th element of the sequence } (a_n) \text{ is equal to } 2. \end{cases}$$

The set of all numbers in the interval $[0; 1]$ can be put into a one-to-one correspondence with the constructed class of transformations V by the following rule: by writing a number c from the interval $[0; 1]$ in the ternary numeral system,

$c = \sum_{n=1}^{\infty} c_n 3^{-n} \equiv \Delta_{c_1 c_2 \dots c_n \dots}^3$, we obtain its ternary code (the sequence c_n), which belongs to the alphabet $\{0, 1, 2\}$. This corresponds to the transformation $f_{a_n}(x) \in V$. Therefore, V is a continuous set.

Discussion and conclusions

The structure of the set G and its invariants require separate investigation. The question remains open whether there exist continuous transformations of the interval $[0; 1]$, different from piecewise linear ones, that preserve the tails of the negabinary representation of numbers.

Sources of funding. This work was supported by a grant from the Simons Foundation (SFI-PD-Ukraine00014586, Yelahn V.O.)

References

- Albeverio, S., Pratsiovytyi, M. V., & Torbin, G. M. (2004). Fractal probability distributions and transformations preserving the Hausdorff–Besicovitch dimension. *Ergodic Theory and Dynamical Systems*, 24(1), 1–16. <https://doi.org/10.1017/S0143385703000397>
- Cherri, A., & Kamal, H. (2004). Parallel negabinary signed-digit arithmetic operations: One-step negabinary, one-step trinary, and one-step quaternary addition algorithms. *Proceedings of SPIE*, 5484, 35–45. <https://doi.org/10.1117/12.568845>
- Galambos, J. (2006). *Representations of real numbers by infinite series*. Springer.
- Grover, D., Madan, V. K., Nanda, N. K., & Singh, H. (1983). Some hardware realizations of negabinary arithmetic. *International Journal of Electronics*, 55(2), 235–241. <https://doi.org/10.1080/00207218308961584>
- Isaieva, T. M., & Pratsiovytyi, M. V. (2016). Transformations of $(0,1)$ preserving tails of Δ^μ -representation of numbers. *Algebra and Discrete Mathematics*, 22(1), 102–115.
- Kasatkin, V. N. (1982). *New insights into number systems*. Vyshcha Shkola [in Russian]. [Касаткін, В. Н. (1982). *Новое о системах счисления*. Вища школа].
- Osaulenko, R. (2016). The group of continuous transformations of the interval $[0,1]$ preserving the digit frequencies of Q_s -expansions of numbers. *Collection of Works of the Institute of Mathematics of the NAS of Ukraine*, 13(3), 191–204 [in Ukrainian]. [Осауленко, Р. (2016). Група неперервних перетворень відрізка $[0,1]$, що зберігають частоти цифр у Q_s -розкладеннях чисел. *Збірник праць Інституту математики НАН України*, 13(3), 191–204].
- Pratsiovytyi, M. V. (2017). Nega-cantor representations of real numbers as trivial recodings of Cantor representations (nega-s-adic as recodings of s-adic representations). *Proceedings of the Institute of Mathematics of the NAS of Ukraine*, 14(4), 167–177 [in Ukrainian]. [Працьовитий, М. В. (2017). Нега-канторові зображення дійсних чисел як тривіальні перекодування канторових зображень (нега-s-кові як перекодування s-кових). *Труди Інституту математики НАН України*, 14(4), 167–177].
- Pratsiovytyi, M. V. (2022). *Two-symbol encoding systems of real numbers and their application*. Naukova dumka [in Ukrainian]. [Працьовитий, М. В. (2022). *Двосимвольні системи кодування дійсних чисел та їх застосування*. Наукова думка].
- Pratsiovytyi, M. V., Chuikov, A. S., & Skrypnyk, S. V. (2019). The chain D_2 -representation of real numbers and some related functions. *Proceedings of the Institute of Mathematics of the NAS of Ukraine*, 16(3), 101–114 [in Ukrainian]. [Працьовитий, М. В., Чуйков, А. С., & Скрипник, С. В. (2019). Ланцюгове D_2 -зображення дійсних чисел та деякі пов'язані функції. *Труди Інституту математики НАН України*, 16(3), 101–114].
- Pratsiovytyi, M. V., Goncharenko, Y. V., & Lysenko, I. M. (2015). Negabinary representation of real numbers and its applications. *Collection of scientific papers of the National Pedagogical Dragomanov University. Series 1: Physical and Mathematical Sciences*, 17, 83–106 [in Ukrainian]. [Працьовитий, М. В., Гончаренко, Я. В., & Лисенко, І. М. (2015). Нега-двійкове зображення дійсних чисел та його застосування. *Збірник наукових праць Національного педагогічного університету імені М. П. Драгоманова. Серія 1: Фізико-математичні науки*, 17, 83–106].
- Pratsiovytyi, M. V., Lysenko, I. M., & Maslova, Y. P. (2020). Group of continuous transformations of real interval preserving tails of G_2 -representation of numbers. *Algebra and Discrete Mathematics*, 29(1), 99–108. <https://doi.org/10.12958/adm1498>
- Pratsiovytyi, M. V., Lysenko, I. M., & Ratushniak, S. P. (2024). Uncountable group of continuous transformations of unit segment preserving tails of Q_2 -representation of numbers. *Proceedings of the International Geometry Center*, 17(2), 133–142. <https://doi.org/10.15673/pigc.v17i2.2755>
- Rao, G. S., Rao, M. N., & Krishnamurthy, M. N. (1974). A variable-shift nega-binary multiplier. *International Journal of Electronics*, 36(6), 749–751. <https://doi.org/10.1080/00207217408900472>
- Schweiger, F. (1995). *Ergodic theory of fibred systems and metric number theory*. Oxford University Press.

Отримано редакцією журналу / Received: 18.07.25

Прорецензовано / Revised: 06.09.25

Схвалено до друку / Accepted: 10.10.25

Володимир СЛАГІН, асп.

ORCID ID: 0009-0001-6984-5090

e-mail: yelahn@imath.kiev.ua

Інститут математики Національної академії наук України, Київ, Україна

НЕПЕРЕРВНІ ПЕРЕТВОРЕННЯ ВІДРІЗКА, ЩО ЗБЕРІГАЮТЬ ХВОСТИ НЕГА-ДВІЙКОВОГО ЗОБРАЖЕННЯ ЧИСЕЛ З ВІДРІЗКА $[0;1]$

Для нега-двійкового зображення чисел відрізка $[0; 1]$ розглянуто неперервні перетворення цього відрізка (тобто бієктивні відображення множини на себе), які зберігають хвости зображення чисел. Вони конструюються за допомогою операторів ліво- та правостороннього зсувів. Доведено, що множина таких перетворень відносно операції "композиція" утворює некомутативну групу, що має континуальну підгрупу зростаючих перетворень.

Ключові слова: нега-двійкове зображення чисел, циліндр, оператори ліво- та правостороннього зсувів, хвостова множина, неперервні перетворення відрізка, група перетворень відрізка.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження(у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 519.22

DOI: <https://doi.org/10.17721/1812-5409.2025/2.6>

Rostyslav MAIBORODA, DSc (Phys. & Math.), Prof.

ORCID ID: 0000-0002-3899-558X

e-mail: rostmaiboroda@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Olena SUGAKOVA, DSc (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0002-6529-0788

e-mail: olenasugakova@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

BANDWIDTH SELECTION FOR DENSITY ESTIMATION BY MIXTURE WITH VARYING CONCENTRATIONS

Finite mixture models arise in statistics of biological and medical data when the investigated subjects belong to sub-populations with different distributions of observed variable. In the model of mixture with varying concentrations (MVC) the concentrations of the mixture components can vary from observation to observation. We consider estimation of a probability density for a mixture component in MVC by a modification of the kernel density estimator (KDE). To apply KDE one needs to select a tuning parameter named the bandwidth. Two approaches to the bandwidth selection are considered. The first one is a modification of the Silverman's rule of thumb. The second one is a version of the leave-one-out cross-validation algorithm. We present results of simulation which show that both algorithms demonstrate similar behavior for nearly Gaussian densities and the cross-validation outperforms the Silverman's rule of thumb on highly non-Gaussian densities.

Key words: *finite mixture model, varying concentrations, kernel density estimator, bandwidth selection, Silverman's rule of thumb, leave-one-out cross-validation.*

AMS 2020 classification: 62G07, 62G09.

Introduction

Nonparametric density estimators are widely used in applied statistics. One of the most popular such estimators is the kernel density estimator (KDE) (Hastie, Tibshirani, & Friedman, 2009, Section 6.6). To apply KDE one needs to select a tuning parameter called the bandwidth. Among different methods of bandwidth selection the Silverman's rule of thumb is maybe the simplest one, and the cross-validation technique is one of most advanced ones (Silverman, 2018, Section 3.4), (Stone, 1984). Standard versions of these techniques were developed for homogeneous samples of independent identically distributed observations.

In this paper we discuss KDE and bandwidth selection rules in the case when observed subjects belong to M different sub-populations with different distributions of the observed variable ξ . If the sub-population which the subject belongs to is unknown, then the observed distribution of ξ is a mixture of its distributions for different components. In classical finite mixture models the mixing probabilities are the same for all observed subjects (Titterton, Smith, & Makov, 1985), (McLachlan, & Peel, 2000). We consider more flexible model of mixture with varying concentrations (MVC) at which the mixing probabilities are different for different subjects (see (Maiboroda, & Sugakova, 2012) for theory and applications to genetical data analysis, (Pidnebesna et al., 2023) for applications in neuroscience). A modification of KDE and Silverman's rule of thumb for MVC is considered in (Sugakova, 1999).

The main goal of this paper is to introduce a modification of the cross-validation technique for MVC KDE and to compare its performance with the performance of KDE with the bandwidth selected by the Silverman's rule of thumb. We performed a simulation study to analyze accuracy of these estimators. Note that the jackknife technique modification for MVC was considered in (Maiboroda, Miroshnichenko, & Sugakova, 2022). In the case of independent identically distributed observations the jackknife is rather similar to the cross-validation. But their modifications for MVC data differ significantly.

1. Mixtures with varying concentrations and components estimation

Assume that the observed subjects $O_1, \dots, O_n$ are obtained from M different sub-populations (components of the mixture). The true number $\kappa(O)$ of the sub-population which the subject O belongs to is unknown, but the probabilities $p_j^i = P\{\kappa(O) = i\}$ are given. These probabilities are called the concentrations of components or the mixing probabilities.

For each subject we observe a numerical random variable $\xi_j = \xi(O_j)$. The distribution of $\xi(O)$ depends on $\kappa(O)$:

$$F_i(x) = \Pr\{\xi(O) < x \mid \kappa(O) = i\}.$$

So, the cumulative distribution function of ξ_j is

$$P\{\xi_j < x\} = \sum_{i=1}^M p_j^i F_i(x). \quad (1)$$

Formula (1) is called the model of mixture with varying concentrations (MVC). It is assumed that ξ_j are independent. The CDFs F_i are unknown. We suppose that they have probability densities f_i :

$$F_i(x) = \int_{-\infty}^x f_i(t) dt.$$

Then the probability density of ξ_j is $\sum_{i=1}^m p_j^i f_i(x)$. Our aim is to estimate f_k by the data $\Xi = (\xi_1, \dots, \xi_n)$. Let us start from estimation of F_k . The usual way to estimate CDF by a homogeneous sample is to use the empirical CDF

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n 1\{\xi_j < x\}.$$

But in the MVC case the empirical CDF would estimate some mixture of different F_i , $i = 1, \dots, M$, not the CDF F_k of a chosen mixture component. To estimate F_k one can use some weighted version of empirical CDF with weights which suppress the influence of nuisance components. In (Maiboroda, & Sugakova, 2012) weighted CDF with minimax weights is proposed for this purpose:

$$\hat{F}_{k;n}(x) = \frac{1}{n} \sum_{j=1}^n a_j^k 1\{\xi_j < x\}.$$

Here the matrix of minimax weights $\mathbf{a} = (a_j^i)_{j=1, \dots, n, i=1, \dots, M} \in \mathbb{R}^{n \times M}$ is defined as $\mathbf{a} = \mathbf{p} \Gamma_n^{-1}$, where

$$\mathbf{p} = (p_j^i)_{j=1, \dots, n, i=1, \dots, M} \in \mathbb{R}^{n \times M}, \quad \Gamma_n = \frac{1}{n} \mathbf{p}^T \mathbf{p}.$$

As discussed in (Maiboroda, & Sugakova, 2012), if $\det \Gamma_n \neq 0$, then $\hat{F}_{k;n}(x)$ is a minimax estimator to $F_k(x)$ with respect to the quadratic loss function. With the weighted CDF $\hat{F}_{k;n}$ at hand, one can readily construct estimators for numerical characteristics of the k -th component distribution. Say, to estimate the m -th moment of F_k ,

$$\mu_k^m = \int_{-\infty}^{\infty} x^m F_k(dx),$$

one can use

$$\hat{\mu}_{k;n}^m = \int_{-\infty}^{\infty} x^m \hat{F}_{k;n}(dx) = \frac{1}{n} \sum_{j=1}^n a_j^k (\xi_j)^m.$$

The variance

$$\sigma^2(F_k) = \int_{-\infty}^{\infty} (x - \mu_k^1)^2 F_k(dx)$$

can be estimated by

$$\hat{\sigma}_k^2 = \int_{-\infty}^{\infty} (x - \hat{\mu}_{k;n}^1)^2 \hat{F}_{k;n}(dx) = \frac{1}{n} \sum_{j=1}^n a_j^k (\xi_j - \hat{\mu}_{k;n}^1)^2. \quad (2)$$

Consistency and asymptotic normality of these estimators is discussed in (Maiboroda, & Sugakova, 2012).

In what follows we will also need an estimator for the interquartile range (IQR) of F_k . Recall that IQR of a CDF F_k can be defined as

$$\text{IQR}(F_k) = \sup\{x : F_k(x) < 3/4\} - \inf\{x : F_k(x) > 1/4\}.$$

(If the function F_k has an inverse F_k^{-1} , then $\text{IQR}(F_k) = F_k^{-1}(3/4) - F_k^{-1}(1/4)$). So, we define the estimator for $\text{IQR}(F_k)$ as

$$\widehat{\text{IQR}}_k = \sup\{x : \hat{F}_{k;n}(x) < 3/4\} - \inf\{x : \hat{F}_{k;n}(x) > 1/4\}. \quad (3)$$

Asymptotic behavior of such estimators is discussed in (Maiboroda, & Sugakova, 2020), (Maiboroda, Miroshnichenko, & Sugakova, 2024).

2. Modified kernel density estimator for mixture components

Let us now consider estimation of the k th component density f_k . In (Sugakova, 1999) a modified kernel density estimator (MKDE) for $f_k(x)$ is considered of the form

$$\hat{f}_{k;n}(x) = \frac{1}{n} \sum_{j=1}^n a_j^k K\left(\frac{x - \xi_j}{h}\right), \quad (4)$$

where $K : \mathbb{R} \rightarrow \mathbb{R}$ is a kernel, i.e. a function satisfying $\int f(z) dz = 1$, $h > 0$ is the bandwidth. The bandwidth is a tuning parameter of the estimator which must be selected carefully to ensure a good accuracy. The integrated squared error

$$\text{ISE}(h) = \text{ISE}(\hat{f}_{k;n}) = \int_{-\infty}^{\infty} (\hat{f}_{k;n}(x) - f_k(x))^2 dx, \quad (5)$$

and the mean integrated squared error

$$\text{MISE}(h) = \text{MISE}(\hat{f}_{k;n}) = \int_{-\infty}^{\infty} \mathbb{E}(\hat{f}_{k;n}(x) - f_k(x))^2 dx = \mathbb{E}[\text{ISE}(\hat{f}_{k;n})] \quad (6)$$

are the most useful measures of estimator's accuracy.

The following theorem on the lower bound of asymptotic MISE of kernel density estimators is shown in (Sugakova, 1999).

Theorem 1. Assume that

- (i) There exists a finite interval $[b_1, b_2] \subset \mathbb{R}$, such that $f_k(x) = 0$ for all $x \notin [b_1, b_2]$.
- (ii) $f_k''(x) = d^2 f_k(x)/dx^2$ exists for all $x \in [b_1, b_2]$ and

$$\phi_k = \int_{-\infty}^{\infty} (f_k''(x))^2 dx < \infty.$$

(iii)

$$\int_{-\infty}^{\infty} zK(z)dz = 0, \quad D = \int_{-\infty}^{\infty} z^2 K(z)dz < \infty, \quad d^2 = \int_{-\infty}^{\infty} K^2(z)dz < \infty.$$

(iv) $A_n = \frac{1}{n} \sum_{j=1}^n (a_j^k)^2 \rightarrow A$ as $n \rightarrow \infty$, where $0 < A < \infty$.

(v) There exists $c > 0$, such that $\det \Gamma_n > c$ for all n .

Then

$$\liminf_{n \rightarrow \infty} n^{4/5} \text{MISE}(\hat{f}_{k;n}) \geq \frac{5}{4} (D A d^2)^{4/5} \quad (7)$$

and the equality in (7) is attained if

$$h = h_n^{\text{opt}} = \left(\frac{A d^2}{D^4 \phi_k} \right)^{1/5}. \quad (8)$$

The bandwidth h_n^{opt} defined by (8) is called the theoretically optimal bandwidth. It corresponds to asymptotically least possible value of MISE as $n \rightarrow \infty$. But it is impossible to use h_n^{opt} in the density estimation by real data, since the value of ϕ_k depends from the unknown density. So one needs some practical sub-optimal bandwidth selection algorithm.

3. Modified Silverman's rule of thumb

The simplest bandwidth selection algorithm is the Silverman's rule of thumb. It is based on the idea to approximate ϕ_k by its estimator by the data. This needs some additional assumptions on f_k . If f_k was a Gaussian density with variance σ_k^2 , then its ϕ_k would be

$$\phi_k = \frac{3}{8\sqrt{\pi}\sigma^5}$$

and the optimal bandwidth would be

$$h_n^{\text{opt}} = \left(\frac{8\sqrt{\pi} A d^2}{3 D^4} \right)^{1/5} \sigma_k.$$

Replacing here σ_k by its estimator $\hat{\sigma}_k$ defined by (2), we obtain a simple bandwidth selection rule. To make this rule more robust, one can utilize the IQR for estimation of σ_k . Namely, since $\sigma_k = \text{IQR}(F_k)/(2\Phi^{-1}(0.75))$ for Gaussian F_k , we obtain the following modification of the Silverman's rule of thumb for MVC data:

$$h_n^{\text{Silv}} = \left(\frac{8\sqrt{\pi} A d^2}{3 D^4} \right)^{1/5} \min(\hat{\sigma}_k, \widehat{\text{IQR}}_k / (2\Phi^{-1}(0.75))).$$

Surely, the true f_k to be estimated, is not Gaussian. So this rule of thumb will not be optimal even asymptotically. But one expects that it should lead to appropriate estimators if f_k is a Gaussian-like probability density.

4. Modified cross-validation bandwidth selector

Another way to chose the bandwidth is based on the leave-one-out cross-validation technique (LOO-CV). It is based on the approximate minimization of ISE given by (5). Observe that

$$\text{ISE}(h) = J_1(h) - 2J_2(h) + J_3,$$

where

$$J_1(h) = \int_{-\infty}^{\infty} (\hat{f}_{k;n}(x))^2 dx, \quad J_2(h) = \int_{-\infty}^{\infty} \hat{f}_{k;n}(x) f_k(x) dx, \\ J_3 = \int_{-\infty}^{\infty} (f_k(x))^2 dx.$$

The term J_3 does not depend on h and can be dropped. The term $J_1(h)$ can be calculated by the data directly. But $J_2(h)$ depends on h and on the unknown f_k . To obtain a feasible bandwidth selector we need to estimate $J_2(h)$. Let us do it by a LOO-CV type estimator. Observe that

$$J_2(h) = E[\hat{f}_{k;n}(\eta) \mid \Xi],$$

where η is a random variable with CDF F_k independent of the data Ξ by which $\hat{f}_{k;n}$ is calculated. To estimate $J_2(h)$ we use the weighted mean

$$\hat{J}_2(h) = \frac{1}{n} \sum_{j=1}^n a_j^k \hat{f}_{k;n}^{(j-)}(\xi_j),$$

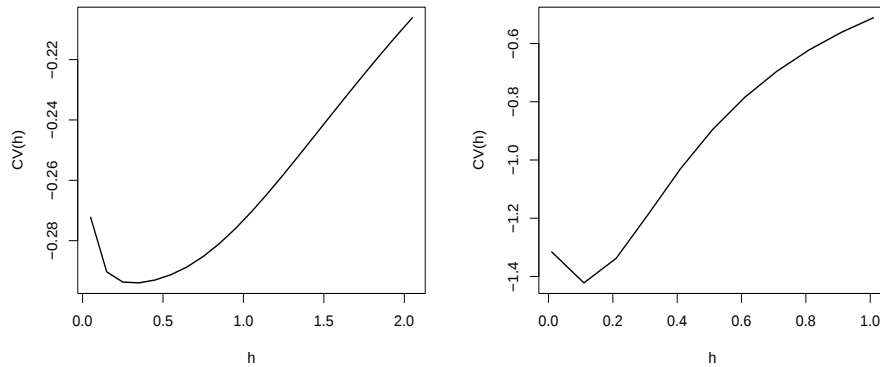


Fig. 1. Graph of $CV(h)$ in the Experiment 1. Component 1 on the left panel, Component 2 on the right panel

where $\hat{f}_{k;n}^{(j-)}(x)$ is the MKDE, calculated by all observations except the j th one. More precisely, $\hat{f}_{k;n}^{(j-)}(x)$ is defined as follows. Let

$$p_j^{k,i-} = \begin{cases} p_j^k & \text{if } j \neq i, \\ 0 & \text{if } j = k, \end{cases}$$

$$\mathbf{p}^{(i-)} = (p_j^{l,i-})_{j=1,\dots,n,l=1,\dots,M} \in \mathbb{R}^{n \times M}, \quad \mathbf{\Gamma}_n^{(i-)} = \frac{1}{n} (\mathbf{p}^{(i-)})^T \mathbf{p}^{(i-)}.$$

$$\mathbf{a}^{(i-)} = \mathbf{p}^{(i-)} (\mathbf{\Gamma}_n^{(i-)})^{-1}.$$

Then

$$\hat{f}_{k,n}^{(i-)}(x) = \frac{1}{n} \sum_{j=1}^n a_j^{k,i-} K\left(\frac{x - \xi_j}{h}\right).$$

Observe that $a_i^{k,i-} = 0$, so ξ_i and $\hat{f}_{k,n}^{(i-)}(x)$ are independent.

Now the cross-validation functional is defined as

$$CV(h) = J_1(h) - 2\hat{J}_2(h).$$

It is an estimator of $ISE(h) - J_3$. The cross-validation bandwidth is

$$h_n^{CV} = \operatorname{argmin}_{h>0} CV(h). \quad (9)$$

In the simulation experiments we considered two-component mixtures ($M = 2$) with the concentrations defined as

$$p_j^1 = \frac{j}{n}; \quad p_j^2 = 1 - \frac{j}{n}; \quad j = 1, \dots, n.$$

The Gaussian kernel $K(t) = \exp(-t^2/2)/\sqrt{2\pi}$ was used in the density estimator.

Experiment 1. In the first experiment we simulate the sample of size $n = 1000$. Both components are normal: the first one has distribution $N(10, 1)$ and the second one $N(0, 0.2)$. Plot of $CV(h_n)$ for the first and second components (Fig. 1) show presence of the global minimum. So, calculation of the optimal h_n by formula (9) is reasonable.

5. Simulation results

In all the following experiments we generated samples of the size $n = 100, 250, 500, 750, 1000$. For each n , $T = 1000$ samples were generated. Firstly, we calculate for every case h_n^{Silv} and h_n^{CV} . Then we derive means and variations for $ISE(h_n)$ defined by (5) for $h_n = h_n^{Silv}$ and $h_n = h_n^{CV}$.

Experiment 2. In second experiment both components are normal too, but with another parameters, then in the Experiment 1: the first component $\sim N(-1, 1)$, the second one $\sim N(1, 1)$. The results are presented in the Tab. 1 and 2.

The results are practically identical for the Silverman's and CV bandwidths, but the estimators based on the Silverman's rule perform slightly better. In the next experiment we work with heavy-tailed distributions.

Table 1

Experiment 2, Component 1				
n	$\operatorname{mean}(ISE(h_n^{CV}))$	$\operatorname{mean}(ISE(h_n^{Silv}))$	$\operatorname{var}(ISE(h_n^{CV}))$	$\operatorname{var}(ISE(h_n^{Silv}))$
100	0.024065	0.018527	7.23E-04	1.93E-04
250	0.011508	0.008928	9.95E-05	3.48E-05
500	0.006502	0.005176	2.41E-05	1.08E-05
750	0.004907	0.003936	1.50E-05	5.63E-06
1000	0.003885	0.003137	8.67E-06	3.69E-06

Table 2

Experiment 2, Component 1				
n	$\text{mean}(\text{ISE}(h_n^{CV}))$	$\text{mean}(\text{ISE}(h_n^{Silv}))$	$\text{var}(\text{ISE}(h_n^{CV}))$	$\text{var}(\text{ISE}(h_n^{Silv}))$
100	0.025176	0.018637	6.59E-04	1.82E-04
250	0.011968	0.009080	1.09E-04	3.38E-05
500	0.006694	0.005290	3.42E-05	1.16E-05
750	0.005156	0.004025	1.71E-05	6.66E-06
1000	0.003967	0.003087	9.07E-06	3.01E-06

Experiment 3. In this experiment both components have Student's t distribution: the first component with $df = 5$ degrees of freedom, the second one with $df = 20$ degrees of freedom.

The results of this experiment presented in Tab. 3, 4 are rather similar to Experiment 1. The Silverman's rule performs a bit better. So the heavy tails of the estimated density does not deteriorate the estimators performance.

Table 3

Experiment 3, Component 1				
n	$\text{mean}(\text{ISE}(h_n^{CV}))$	$\text{mean}(\text{ISE}(h_n^{Silv}))$	$\text{var}(\text{ISE}(h_n^{CV}))$	$\text{var}(\text{ISE}(h_n^{Silv}))$
100	0.024022	0.016803	6.46E-04	1.70E-04
250	0.011216	0.007846	1.47E-04	3.35E-05
500	0.007206	0.005233	4.34E-05	1.21E-05
750	0.005117	0.003769	2.48E-05	6.25E-06
1000	0.003963	0.003115	9.52E-06	4.30E-06

Table 4

Experiment 3, Component 2				
n	$\text{mean}(\text{ISE}(h_n^{CV}))$	$\text{mean}(\text{ISE}(h_n^{Silv}))$	$\text{var}(\text{ISE}(h_n^{CV}))$	$\text{var}(\text{ISE}(h_n^{Silv}))$
100	0.025452	0.017431	9.55E-04	1.76E-04
250	0.011807	0.008621	1.10E-04	3.60E-05
500	0.007048	0.005089	4.78E-05	1.19E-05
750	0.004811	0.003675	1.57E-05	6.12E-06
1000	0.004145	0.003010	1.86E-05	3.68E-06

Experiment 4. In this experiment the first component is normal $\sim N(0, 1)$ but the second one is a mixture of two normal distributions $N(3, 1)$ and $N(-3, 1)$ with concentrations equal 1/2. The results are presented in Tab. 5 and 6.

Here the distribution of the second component significantly differs from the normal one, so one expects that the Silverman's rule would fall. Accordingly, in this case the CV-based estimator outperform the Silverman's one.

Table 5

Experiment 4, Component 1				
n	$\text{mean}(\text{ISE}(h_n^{CV}))$	$\text{mean}(\text{ISE}(h_n^{Silv}))$	$\text{var}(\text{ISE}(h_n^{CV}))$	$\text{var}(\text{ISE}(h_n^{Silv}))$
100	0.024363	0.019493	4.24E-04	1.76E-04
250	0.011588	0.009500	8.18E-05	3.24E-05
500	0.006695	0.005624	2.37E-05	1.11E-05
750	0.004867	0.004023	1.14E-05	5.13E-06
1000	0.003905	0.003246	6.97E-06	3.26E-06

Table 6

Experiment 4, Component 2				
n	$\text{mean}(\text{ISE}(h_n^{CV}))$	$\text{mean}(\text{ISE}(h_n^{Silv}))$	$\text{var}(\text{ISE}(h_n^{CV}))$	$\text{var}(\text{ISE}(h_n^{Silv}))$
100	0.020952	0.038626	2.07E-04	2.17E-05
250	0.009808	0.027813	3.65E-05	1.19E-05
500	0.005848	0.020735	1.20E-05	6.20E-06
750	0.004155	0.017020	6.17E-06	4.13E-06
1000	0.003320	0.014971	2.87E-06	3.40E-06

Discussion and conclusion

We have introduced a new modification of CV-approach to the bandwidth selection for the kernel density estimators. This approach allows to chose the bandwidth for the estimation by data from a mixture with varying concentrations. Results of simulations demonstrate that the CV-based estimators perform nearly as well as the modified Silverman's rule of thumb based estimators when the estimated density is similar to the Gaussian one. But in the highly non-Gaussian case the CV bandwidth selection outperforms the Silverman's rule significantly even for small sample sizes. Further experiments are needed for investigation of small sample behavior of CV-based estimators.

Authors' contribution: Rostyslav Maiboroda — conceptualization; methodology; Olena Sugakova — methodology; software.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning data mining, inference, and prediction*. Springer.
- Maiboroda, R., Miroshnichenko, V., & Sugakova, O. (2022). Jackknife for nonlinear estimating equations. *Modern Stochastics: Theory and Applications*, 9(4), 377–399. <https://doi.org/10.15559/22-VMSTA208>
- Maiboroda, R., Miroshnichenko, V., & Sugakova, O. (2024). Quantile estimators for regression errors in mixture models with varying concentrations. *Bulletin of Taras Shevchenko National University of Kyiv. Physical and Mathematical Sciences*, 1(78), 45–50. <https://doi.org/10.17721/1812-5409.2024/1.8>
- Maiboroda, R., & Sugakova, O. (2012). Statistics of mixtures with varying concentrations with application to dna microarray data analysis. *Nonparametric statistics*, 24(1), 201–215. <https://doi.org/10.1080/10485252.2011.630076>

- Maiboroda, R., & Sugakova, O. (2020). Tests of hypotheses on quantiles of distributions of components in a mixture. *Theory of Probability and Mathematical Statistics*, 101, 179–191. <https://doi.org/10.1090/tpms/1120>
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley. <https://doi.org/10.1002/0471721182>
- Pidnebesna, A., Fajnerová, I., Horáček, J., & Hlinka, J. (2023). Mixture components inference for sparse regression: Introduction and application for estimation of neuronal signal from fmri bold. *Applied Mathematical Modelling*, 116, 735–748. <https://doi.org/10.1016/j.apm.2022.11.034>
- Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.
- Stone, C. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, 12(4), 1285–1297. <https://doi.org/10.1002/047172118210.1214/aos/1176346792>
- Sugakova, O. (1999). Asymptotics of a kernel estimate for the density of a distribution constructed from observations of a mixture with varying concentration. *Theory of Probability and Mathematical Statistics*, 59, 161–171.
- Titterton, D., Smith, A., & Makov, O. (1985). *Analysis of finite mixture distributions*. Wiley.

Отримано редакцією журналу / Received: 30.04.25

Прорецензовано / Revised: 02.09.25

Схвалено до друку / Accepted: 10.10.25

Ростислав МАЙБОРОДА, д-р фіз.-мат. наук, проф.

ORCID ID: 0000-0002-3899-558X

e-mail: rostmaiboroda@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Олена СУГАКОВА, д-р фіз.-мат. наук, доц.

ORCID ID: 0000-0002-6529-0788

e-mail: olenasugakova@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

ВИБІР ПАРАМЕТРА ЗГЛАДЖУВАННЯ ДЛЯ ЯДЕРНОЇ ОЦІНКИ ЗА СПОСТЕРЕЖЕННЯМИ ІЗ СУМІШІ

Статистичні дані біологічних і медичних досліджень часто являють собою суміш спостережень об'єктів, що належать різним підпопуляціям з різними статистичними властивостями спостережуваних характеристик. Такі дані зручно описувати математичними моделями скінченних сумішей. У пропонованій роботі для опису даних використано модель суміші зі змінними концентраціями (СЗК), у якій концентрації компонентів можуть змінюватись від спостереження до спостереження. Розглянуто задачу непараметричного оцінювання щільності розподілу окремої компоненти суміші в межах моделі СЗК. Для цього використовується модифікована ядерна оцінка щільності (ЯОЩ). Практичне застосування ЯОЩ потребує вибору параметра згладжування. У статті запропоновано два підходи до такого вибору: на основі модифікованого правила Сілвермана і за допомогою алгоритму кросвалідації, адаптованого до СЗК. Якість отриманих оцінок порівнюється за допомогою імітаційного моделювання. Результати моделювання показують, що у випадку щільностей, близьких до гауссових, обидва підходи дають майже однакові результати. Для щільностей, які відрізняються від гауссових, вибір параметра згладжування на основі модифікованої кросвалідації дає значно кращі результати, ніж на основі правила Сілвермана.

Ключові слова: модель скінченної суміші, змінні концентрації, ядерна оцінка щільності, вибір параметра згладжування, правило Сілвермана, кросвалідація.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 519.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.7>

Oleksandr MASYUTKA, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-7301-8813
e-mail: omasutka@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Mikhail MOKLYACHUK, DSc (Phys. & Math.), Prof.
ORCID ID: 0000-0002-6173-0280
e-mail: moklyachuk@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

INTERPOLATION PROBLEM FOR MULTIDIMENSIONAL HARMONIZABLE STABLE SEQUENCES

The problem of estimation of unobserved values of stochastic processes is of constant interest in the theory and applications of stochastic processes. The problem of forecasting future values of economic and physical processes, the problem of restoring lost information, cleaning signals, or other data from observations with noise is magnified in an information-laden world. Therefore, the development of estimation methods is one of the main tasks of the modern theory of stochastic processes. In this paper, we consider the problem of optimal linear interpolation of a functional that depends on the unknown values of a vector-valued harmonizable symmetric α -stable random sequence from observations of the sequence with noise. We use the classical approach to derive formulas for computing values of the mean-square error and the spectral characteristic of the optimal linear estimate of the functional. The crucial assumption of this approach is that the spectral densities of the involved stochastic sequences are exactly known. However, in practice, complete information on the spectral densities is impossible in most cases. In this situation, one finds a parametric or non-parametric estimate of the unknown spectral density and then applies one of the traditional estimation methods provided that the selected density is the true one. This procedure can result in a significant increase in the value of the estimation error. To avoid this effect, one can search for the estimates that are optimal for all densities from a certain class of admissible spectral densities. These estimates are called minimax because they minimize the maximal values of the errors of estimates for all densities from a given class. Therefore, in the case of spectral uncertainty, we use the minimax approach and propose formulas that determine the least favorable spectral densities and the minimax spectral characteristics of the optimal estimates of the functional for some classes of admissible spectral densities.

Key words: *harmonizable stable random sequence, periodically harmonizable stable random sequence, optimal linear estimate, minimax-robust estimate, minimax spectral characteristic, least favorable spectral density.*

AMS 2020 classification: 60G10, 60G25, 60G35, 62M20, 93E10, 93E11.

Introduction

The problem of estimation of the unknown values of harmonizable random sequences and processes were investigated in papers (Cambanis, 1983; Cambanis, & Soltani, 1984; Hosoya, 1982; Masani, & Wiener, 1957). The interpolation problem for harmonizable symmetric α -stable random sequences were investigated in papers (Weron, 1985; Pourahmadi, 1984).

Basic results concerning estimation of the unknown (missed) values of stochastic processes are based on the assumption that spectral densities of processes are exactly known. In practice, however, complete information on the spectral densities is impossible in most cases. In such situations one finds parametric or nonparametric estimates of the unknown spectral densities. Then the classical estimation method is applied under the assumption that the estimated densities are true. This procedure can result in significant increasing of the value of error as Vastola and Poor have demonstrated with the help of some examples (Vastola, & Poor, 1983). This is a reason to search estimates which are optimal for all densities from a class of admissible spectral densities. These estimates are called minimax since they minimize the maximal value of the error. A survey of results (till 1985) in minimax (robust) methods of data processing can be found in the paper (Kassam, & Poor, 1985). The paper (Grenander, 1957) should be marked as the first one where the minimax extrapolation problem for stationary processes was formulated and solved. Later Franke and Poor (Franke, & Poor, 1984; Franke, 1984, 1985) applied the convex optimization methods for investigation the minimax-robust extrapolation and interpolation problems. The book (Moklyachuk, & Masyutka, 2012) is dedicated to minimax-robust extrapolation, interpolation and filtering problems for vector-valued stationary processes and sequences. In the book (Moklyachuk, & Golichenko, 2016) the minimax-robust extrapolation, interpolation and filtering problems for periodically correlated processes are investigated. In the paper (Moklyachuk, & Ostapenko, 2016) minimax-robust interpolation problems are studied for harmonizable stable sequences.

In this paper the problem of optimal estimation is investigated for the linear functional

$$A_N \vec{\xi} = \sum_{j=0}^N (\vec{a}(j))^T \vec{\xi}(j)$$

that depends on the unknown values $\vec{\xi}(j), j = 0, 1, \dots, N$, of a vector-valued harmonizable symmetric α -stable random sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ where $\vec{\xi}(j)$ and $\vec{\eta}(j)$ are mutually independent harmonizable symmetric α -stable random sequences which have the spectral densities $f(\theta)$ and $g(\theta)$ satisfying the minimality condition.

The problem is investigated under the condition of spectral certainty as well as under the condition of spectral uncertainty. Formulas for calculation the value of the error and spectral characteristic of the optimal linear estimate of the

functional are derived under the condition of spectral certainty where spectral density of the sequence is exactly known. In the case where spectral density of the sequence is not exactly known sets of admissible spectral densities is available, relations which determine least favorable densities and the minimax-robust spectral characteristics for different classes of spectral densities are found.

1. Harmonizable symmetric α -stable random sequence

Definition 1. (symmetric α -stable random variable) A real-valued random variable ξ is said to be symmetric α -stable, $S_{\alpha}S$, if its characteristic function has the form $E \exp(it\xi) = \exp(-c|t|^{\alpha})$ for some $c \geq 0$ and $0 < \alpha \leq 2$. The real random variables $\xi_1, \xi_2, \dots, \xi_n$ are jointly $S_{\alpha}S$ if all linear combinations $\sum_{k=1}^n a_k \xi_k$ are $S_{\alpha}S$, or, equivalently, if the characteristic function of the random vector $\vec{\xi} = (\xi_1, \dots, \xi_n)$ is of the form

$$\phi_{\vec{\xi}}(\vec{t}) = E \exp \left(i \sum_{k=1}^n t_k \xi_k \right) = \exp \left\{ - \int_{S_n} \left| \sum_{k=1}^n t_k x_k \right|^{\alpha} d\Gamma_{\vec{\xi}}(\vec{x}) \right\},$$

where $\vec{t} = (t_1, \dots, t_n)$, $t_1, \dots, t_n$ are real numbers and $\Gamma_{\vec{\xi}}(\vec{x})$ is a symmetric measure defined on the unit sphere $S_n \in R^n$, called the spectral measure of the random vector $\vec{\xi} = (\xi_1, \dots, \xi_n)$. There is a one-to-one correspondence between the distribution of $\vec{\xi}$ and its spectral measure $\Gamma_{\vec{\xi}}(\vec{x})$ (Cambanis, 1983).

For real-valued jointly $S_{\alpha}S$ random variables ξ, η with $1 < \alpha \leq 2$ the covariation of ξ, η is defined by

$$[\xi, \eta]_{\alpha} = \int_{S_2} (x)(y)^{<\alpha-1>} d\Gamma_{\xi, \eta}(x, y),$$

where $(y)^{<\beta>} = |y|^{\beta-1}y$.

For jointly $S_{\alpha}S$ random variables $\xi = \xi_1 + i\xi_2$ and $\eta = \eta_1 + i\eta_2$ the covariation of ξ with η is defined as (Cambanis, 1983)

$$[\xi, \eta]_{\alpha} = \int_{S_4} (x_1 + ix_2)(y_1 + iy_2)^{<\alpha-1>} d\Gamma_{\xi_1, \xi_2, \eta_1, \eta_2}(x_1, x_2, y_1, y_2),$$

where $z^{<\beta>} = |z|^{\beta-1}\bar{z}$ for a complex number z and $\beta > 0$.

The covariation in general is not symmetric and linear on second argument and for ξ, ξ_1, ξ_2, η jointly $S_{\alpha}S$ has the following properties (Cambanis, 1983; Weron, 1985)

$$[\xi_1 + \xi_2, \eta]_{\alpha} = [\xi_1, \eta]_{\alpha} + [\xi_2, \eta]_{\alpha}, \quad (1)$$

$$[a\xi, b\eta]_{\alpha} = a(b)^{\alpha-1} [\xi, \eta]_{\alpha}, \quad (2)$$

$$[\xi, \eta]_{\alpha} = 0 \quad \text{if } \xi \text{ and } \eta \text{ are independent}, \quad (3)$$

$$[\xi, \eta_1 + \eta_2]_{\alpha} = [\xi, \eta_1]_{\alpha} + [\xi, \eta_2]_{\alpha} \quad \text{if } \eta_1 \text{ and } \eta_2 \text{ are independent}, \quad (4)$$

$$|[\xi, \eta]_{\alpha}| \leq \|\xi\|_{\alpha} \|\eta\|_{\alpha}^{\alpha-1}, \quad (5)$$

the functional

$$\|\xi\|_{\alpha} = [\xi, \xi]_{\alpha}^{1/\alpha} \quad (6)$$

is a norm in a linear space of $S_{\alpha}S$ random variables which is equivalent to convergence in probability

$$\left\| \sum_{k=1}^n \xi_k \right\|_{\alpha}^{\alpha} = \sum_{k=1}^n \|\xi_k\|_{\alpha}^{\alpha} \quad (7)$$

when $\xi_1, \dots, \xi_n$ are independent, the mapping

$$\xi \rightarrow [\xi, \eta]_{\alpha} \quad (8)$$

is a bounded linear functional with the norm $\|\eta\|_{\alpha}^{\alpha-1}$ on the linear space of $S_{\alpha}S$ random variables, and every bounded linear functional on such a space is of this form for some η .

It should be noted that $\|\cdot\|_{\alpha}$ is not necessarily the usual L^{α} norm.

Definition 2 (symmetric α -stable stochastic sequence). A stochastic sequence $\{\xi(n), n \in \mathbb{Z}\}$ is called symmetric α -stable, $S_{\alpha}S$, if all linear combinations $\sum_{m=1}^l a_m \xi(n_m)$ are $S_{\alpha}S$ random variables.

A vector-valued stochastic sequence $\vec{\xi}(n) = \{\xi_k(n)\}_{k=1}^T$, $n \in \mathbb{Z}$, is called symmetric α -stable, $S_{\alpha}S$, stochastic sequence, if all linear combinations $\sum_{k=1}^T \sum_{m=1}^l a_{mk} \xi_k(n_m)$ are $S_{\alpha}S$ random variables.

Let $Z = \{Z(t) : -\infty < t < \infty\}$ be a complex $S_{\alpha}S$ process with independent increments. The spectral measure of the process Z is defined as $\mu\{(s, t]\} = \|Z(t) - Z(s)\|_{\alpha}^{\alpha}$.

The integrals $\int a(t)dZ(t)$ can be defined for all $a(t) \in L^\alpha(\mu)$ with properties for all $a \in L^\alpha(\mu), b \in L^\alpha(\mu)$ (Cambanis, 1983; Cambanis, & Soltani, 1984; Hosoya, 1982):

$$\left\| \int a(t)dZ(t) \right\|_\alpha^\alpha = \int |a(t)|^\alpha d\mu(t), \quad (9)$$

$$\left[\int a(t)dZ(t), \int b(t)dZ(t) \right]_\alpha = \int a(t)(b(t))^{\langle \alpha-1 \rangle} d\mu(t), \quad (10)$$

and for vector-valued functions $\vec{a}(t) = \{a_k(t)\}_{k=1}^T, a_k(t) \in L^\alpha(\mu), \vec{b}(t) = \{b_k(t)\}_{k=1}^T, b_k(t) \in L^\alpha(\mu)$, and $\vec{Z}(t) = \{Z_k(t)\}_{k=1}^T$

$$\left[\int (\vec{a}(t))^\top d\vec{Z}(t), \int (\vec{b}(t))^\top d\vec{Z}(t) \right]_\alpha = \int (\vec{a}(t))^\top d\mu(t) (\vec{b}(t))^{\langle \alpha-1 \rangle}, \quad (11)$$

where μ is the matrix-valued spectral measure corresponding to the process $\vec{Z}(t)$.

Definition 3 (harmonizable symmetric α -stable sequence). A symmetric α -stable, $S\alpha S$, stochastic sequence $\{\xi(n), n \in \mathbb{Z}\}$ is said to be harmonizable, $HS\alpha S$, if there exists a $S\alpha S$ process $Z = \{Z(\theta) : \theta \in [-\pi, \pi]\}$ with independent increments and a finite spectral measure μ such that sequence $\xi(n)$ has the spectral representation

$$\xi(n) = \int_{-\pi}^{\pi} e^{in\theta} dZ(\theta), \quad n \in \mathbb{Z},$$

and the covariation has the representation

$$[\xi(n), \xi(m)]_\alpha = \int_{-\pi}^{\pi} e^{i(n-m)\theta} d\mu(\theta), \quad m, n \in \mathbb{Z}.$$

A vector-valued symmetric α -stable, $S\alpha S$, stochastic sequence $\vec{\xi}(n) = \{\xi_k(n)\}_{k=0}^{T-1}, n \in \mathbb{Z}$, is said to be harmonizable, $HS\alpha S$, if there exists a vector-valued $S\alpha S$ process $\vec{Z}(\theta) = \{Z_k(\theta)\}_{k=0}^{T-1}, \theta \in [-\pi, \pi]$ with independent increments and a finite matrix-valued spectral measure μ such that sequence $\vec{\xi}(n)$ has the spectral representation

$$\vec{\xi}(n) = \int_{-\pi}^{\pi} e^{in\theta} d\vec{Z}(\theta), \quad n \in \mathbb{Z},$$

and the covariation has the representation

$$[\vec{\xi}(n), \vec{\xi}(m)]_\alpha = \int_{-\pi}^{\pi} e^{i(n-m)\theta} d\mu(\theta), \quad m, n \in \mathbb{Z}.$$

Definition 4 (periodically harmonizable $S\alpha S$ sequence). A symmetric α -stable, $S\alpha S$, stochastic sequence $\{\xi(n), n \in \mathbb{Z}\}$ is said to be periodically harmonizable, $PHS\alpha S$, if the vector-valued stochastic sequence $\vec{\xi}(n) = \{\xi(nT + k)\}_{k=0}^{T-1}, n \in \mathbb{Z}$, is harmonizable symmetric α -stable, $HS\alpha S$, stochastic sequence.

Note that a $HS\alpha S$ stochastic sequence is not necessarily stationary even second order stationary, but for $\alpha = 2$ the $HS\alpha S$ sequences are stationary with Gaussian distribution.

In this article we consider the case where $1 < \alpha \leq 2$.

Denote by $H(\xi)$ the time domain of the $HS\alpha S$ sequence $\{\xi(n), n \in \mathbb{Z}\}$, which is a closed in the norm $\|\cdot\|_\alpha$ linear manifold generated by all values of the $HS\alpha S$ sequence $\{\xi(n), n \in \mathbb{Z}\}$. It follows from the spectral representation of the $HS\alpha S$ sequence $\{\xi(n), n \in \mathbb{Z}\}$ that the mapping $\xi(n) \leftrightarrow e^{in\theta}, n \in \mathbb{Z}$, extends to an isomorphism between the spaces $H(\xi)$ and $L^\alpha(\mu)$. Under this isomorphism to each $\eta \in H(\xi)$ corresponds a unique $f \in L^\alpha(\mu)$ such that $\eta = \int_{-\pi}^{\pi} f(\theta) dZ(\theta)$.

For a closed linear subspace $M \subseteq L^\alpha(\mu)$ and $f \in L^\alpha(\mu)$, there exists a unique element from M which minimizes the distance to f . This element is called projection of f onto M or the best approximation of f in M . This projection is denoted by $P_M f$ and is uniquely determined by the condition (Singer, 1970)

$$\int_{-\pi}^{\pi} g(f - P_M f)^{\langle \alpha-1 \rangle} d\mu = 0, \quad g \in M. \quad (12)$$

Similarly, for $HS\alpha S$ stochastic sequence $\{\xi(n), n \in \mathbb{Z}\}$ and a closed linear subspace $H^-(\xi)$ of the space $H(\xi)$ there is a uniquely determined element $\hat{\xi}(n) \in H^-(\xi)$ which minimizes the distance to $\xi(n)$ and is uniquely determined from the condition

$$[\eta, \xi(n) - \hat{\xi}(n)]_\alpha = 0, \quad \eta \in H^-(\xi). \quad (13)$$

From linearity of the covariation with respect to the first argument from this relation we have that

$$\|\xi(n) - \hat{\xi}(n)\|_\alpha^\alpha = [\xi(n), \xi(n) - \hat{\xi}(n)]_\alpha - [\hat{\xi}(n), \xi(n) - \hat{\xi}(n)]_\alpha = [\xi(n), \xi(n) - \hat{\xi}(n)]_\alpha. \quad (14)$$

This relation plays a fundamental role in the characterization of minimal $HS\alpha S$ stochastic sequences $\{\xi(n), n \in \mathbb{Z}\}$.

2. Interpolation problem. Observations with noise. Projection approach

Consider the problem of the optimal estimation of the linear functional

$$A_N \vec{\xi} = \sum_{j=0}^N (\vec{a}(j))^T \vec{\xi}(j) = \int_{-\pi}^{\pi} (A_N(e^{i\theta}))^T d\vec{Z}^{\xi}(\theta),$$

where

$$A_N(e^{i\theta}) = \sum_{j=0}^N \vec{a}(j) e^{ij\theta},$$

that depends on the unknown values $\vec{\xi}(j), j = 0, 1, \dots, N$, of a vector-valued harmonizable symmetric α -stable random sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$.

We consider the problem for mutually independent vector-valued harmonizable symmetric α -stable random sequences $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$ and $\vec{\eta}(j) = \{\eta_k(j)\}_{k=1}^T$ which have absolutely continuous spectral measures and the spectral densities $f(\theta)$ and $g(\theta)$, satisfying the minimality condition (Pourahmadi, 1984; Weron, 1985)

$$\int_{-\pi}^{\pi} \text{Tr} \left[(f(\theta) + g(\theta))^{-1/(\alpha-1)} \right] d\theta < \infty. \quad (15)$$

Denote by $H^N(\xi + \eta)$ the closed in the $\|\cdot\|_{\alpha}$ norm linear manifold generated by values of the harmonizable symmetric α -stable stochastic sequence $\vec{\xi}(j) + \vec{\eta}(j), j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ in the space $H(\xi + \eta)$ generated by all values of the harmonizable symmetric α -stable, $HS\alpha S$, stochastic sequence $\vec{\xi}(j) + \vec{\eta}(j), j \in \mathbb{Z}$.

The optimal estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ is a projection of $A_N \vec{\xi}$ on the subspace $H^N(\xi + \eta)$ which is determined by the relations

$$[\eta, A_N \vec{\xi} - \hat{A}_N \vec{\xi}]_{\alpha} = 0, \quad \forall \eta \in H^N(\xi + \eta),$$

or, equivalently, by relations

$$[\xi_k(j) + \eta_k(j), A_N \vec{\xi} - \hat{A}_N \vec{\xi}]_{\alpha} = 0, \quad j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}, \quad k = \overline{1, T}. \quad (16)$$

It follows from the isomorphism between the spaces $H(\xi + \eta)$ and $L^{\alpha}(F + G)$ that the optimal estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ is of the form

$$\hat{A}_N \vec{\xi} = \int_{-\pi}^{\pi} (h(\theta))^T \left(d\vec{Z}^{\xi}(\theta) + d\vec{Z}^{\eta}(\theta) \right). \quad (17)$$

It is determined by the spectral characteristic $h(\theta)$ of the estimate which is from the subspace $L_N^{\alpha}(F + G)$ of the $L^{\alpha}(F + G)$ space generated by functions

$$e^{ij\theta} \delta_k, \quad \delta_k = \{\delta_{kl}\}_{l=1}^T, \quad k = 1, \dots, T, \quad j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}.$$

The spectral characteristic $h(\theta) = \{h_k(\theta)\}_{k=1}^T$ of the optimal estimate satisfies the following equations

$$\int_{-\pi}^{\pi} e^{ij\theta} \left[f(\theta) \left(A(e^{i\theta}) - h(\theta) \right)^{<\alpha-1>} - g(\theta) (h(\theta))^{<\alpha-1>} \right] d\theta = 0, \quad j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}. \quad (18)$$

It follows from these equations that the spectral characteristic $h(\theta)$ of the estimate is determined by the equation

$$\left[f(\theta) \left(A(e^{i\theta}) - h(\theta) \right)^{<\alpha-1>} - g(\theta) (h(\theta))^{<\alpha-1>} \right] = C_N(e^{i\theta}), \quad (19)$$

$$C_N(e^{i\theta}) = \sum_{j=0}^N \vec{c}(j) e^{-ij\theta},$$

where $\vec{c}(j) = \{c_k(j)\}_{k=1}^T, j = 0, 1, \dots, N$ are unknown coefficients.

The unknown coefficients $\vec{c}(j), j = 0, 1, \dots, N$, are determined from the condition $h(\theta) \in L_N^{\alpha}(f + g)$, which gives us the system of equations

$$\int_{-\pi}^{\pi} e^{-i\theta k} h(\theta) d\theta = 0, \quad k = 0, 1, \dots, N. \quad (20)$$

The variance of the optimal estimate of the functional is calculated by the formula

$$\begin{aligned} \|A_N \xi - \hat{A}_N \xi\|_{\alpha}^{\alpha} &= \int_{-\pi}^{\pi} \left(A_N(e^{i\theta}) - h(\theta) \right)^T f(\theta) \left(A_N(e^{i\theta}) - h(\theta) \right)^{<\alpha-1>} d\theta + \\ &+ \int_{-\pi}^{\pi} (h(\theta))^T g(\theta) (h(\theta))^{<\alpha-1>} d\theta. \end{aligned} \quad (21)$$

In the case where the dimension of the sequence $T = 1$, this formula can be written in the form

$$\|A_N \xi - \hat{A}_N \xi\|_\alpha^\alpha = \int_{-\pi}^{\pi} |A_N(e^{i\theta}) - h(\theta)|^\alpha f(\theta) d\theta + \int_{-\pi}^{\pi} |h(\theta)|^\alpha g(\theta) d\theta. \quad (22)$$

We can conclude that the following theorems hold true.

Theorem 1. Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$ and $\vec{\eta}(j) = \{\eta_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$ be mutually independent harmonizable symmetric α -stable $HS\alpha S$, stochastic sequences which have absolutely continuous spectral measures and the spectral densities $f(\theta)$ and $g(\theta)$ satisfying the minimality condition (15). The optimal linear estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi} = \sum_{j=0}^N (\vec{a}(j))^\top \vec{\xi}(j)$, that depends on the unknown values $\vec{\xi}(j)$, $j = 0, 1, \dots, N$, of the sequence $\vec{\xi}(j)$, from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ is calculated by formula (17). The spectral characteristic $h(\theta)$ of the estimate is determined by equation (19), where the unknown coefficients $\vec{c}(j)$, $j = 0, 1, \dots, N$, are determined from the system of equations (20). The variance of the optimal estimate of the functional is calculated by formula (21) (by formula (22) in the case where dimension $T = 1$).

Theorem 2. Let the vector-valued harmonizable symmetric α -stable, $PHS\alpha S$, stochastic sequences

$$\vec{\xi}(n) = \{\xi(nT + k)\}_{k=0}^{T-1}, \quad \vec{\eta}(n) = \{\eta(nT + k)\}_{k=0}^{T-1}, \quad n \in \mathbb{Z},$$

that correspond to periodically harmonizable symmetric α -stable, $PHS\alpha S$, stochastic sequences $\xi(n)$ and $\eta(n)$, $n \in \mathbb{Z}$, have absolutely continuous spectral measures and the spectral densities $f(\theta)$, $g(\theta)$, satisfying the minimality condition (15). The optimal linear estimate $\hat{A}_N \vec{\xi}$ of the functional

$$A_N \vec{\xi} = \sum_{j=0}^N (\vec{a}(j))^\top \vec{\xi}(j) = \sum_{j=0}^N \sum_{k=0}^{T-1} a_k(j) \xi(jT + k),$$

where $\vec{a}(j) = \{a_k(j)\}_{k=0}^{T-1}$, $\vec{\xi}(j) = \{\xi(jT + k)\}_{k=0}^{T-1}$, that depends on the unknown values $\xi(j)$, $j = 0, 1, \dots, T(N+1) - 1$, of the sequence $\xi(j)$ from observations of the sequence $\xi(j) + \eta(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, T(N+1) - 1\}$ is calculated by the formula (17). The spectral characteristic $h(\theta)$ of the estimate is determined by equation (19), where the unknown coefficients $\vec{c}(j) = \{c_k(j)\}_{k=0}^{T-1}$, $j = 0, 1, \dots, T(N+1) - 1$, are determined from the system of equations (20). The variance of the optimal estimate of the functional is calculated by the formula (21) (by formula (22) in the case where period $T = 1$).

3. Interpolation problem. Observations without noise. Projection approach

Consider the problem of the optimal estimation of the linear functional

$$A_N \vec{\xi} = \sum_{j=0}^N (\vec{a}(j))^\top \vec{\xi}(j) = \int_{-\pi}^{\pi} (A_N(e^{i\theta}))^\top d\vec{Z}^\xi(\theta),$$

where

$$A_N(e^{i\theta}) = \sum_{j=0}^N \vec{a}(j) e^{ij\theta},$$

that depends on the unknown values $\vec{\xi}(j)$, $j = 0, 1, \dots, N$, of a vector-valued harmonizable symmetric α -stable random sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, from observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$.

We consider the problem for vector-valued harmonizable symmetric α -stable random sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$ which have absolutely continuous spectral measure μ and the spectral density $f(\theta)$, satisfying the minimality condition (Pourahmadi, 1984; Weron, 1985)

$$\int_{-\pi}^{\pi} \text{Tr} \left[(f(\theta))^{-1/(\alpha-1)} \right] d\theta < \infty. \quad (23)$$

Denote by $H^N(\xi)$ the closed in the $\|\cdot\|_\alpha$ norm linear manifold generated by values of the harmonizable symmetric α -stable random sequence $\vec{\xi}(j)$, $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ in the space $H(\xi)$ generated by all values of the $HS\alpha S$ sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$.

The optimal estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ is a projection of $A_N \vec{\xi}$ on the subspace $H^N(\xi)$ which is determined by the relations

$$[\eta, A_N \vec{\xi} - \hat{A}_N \vec{\xi}]_\alpha = 0, \quad \forall \eta \in H^N(\xi),$$

or, equivalently, by relations

$$[\xi_k(j), A_N \vec{\xi} - \hat{A}_N \vec{\xi}]_\alpha = 0, \quad j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}, \quad k = \overline{1, T}. \quad (24)$$

It follows from the isomorphism between the spaces $H(\xi)$ and $L^\alpha(\mu)$ that the optimal estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ is of the form

$$\hat{A}_N \vec{\xi} = \int_{-\pi}^{\pi} (h(\theta))^\top d\vec{Z}^\xi(\theta). \quad (25)$$

It is determined by the spectral characteristic $h(\theta)$ of the estimate which is from the subspace $L_N^\alpha(\mu)$ of the $L^\alpha(\mu)$ space generated by functions

$$e^{ij\theta} \delta_k, \delta_k = \{\delta_{kl}\}_{l=1}^T, k = 1, \dots, T, j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}.$$

The spectral characteristic $h(\theta) = \{h_k(\theta)\}_{k=1}^T$ of the optimal estimate satisfies the following equations

$$\int_{-\pi}^{\pi} e^{ij\theta} f(\theta) \left(A_N(e^{i\theta}) - h(\theta) \right)^{<\alpha-1>} d\theta = 0, \quad j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}. \quad (26)$$

It follows from these equations that the spectral characteristic $h(\theta)$ of the estimate is determined by the equation

$$f(\theta) \left(A_N(e^{i\theta}) - h(\theta) \right)^{<\alpha-1>} = C_N(e^{i\theta}), \quad C_N(e^{i\theta}) = \sum_{j=0}^N \tilde{c}(j) e^{-ij\theta},$$

where $\tilde{c}(j) = \{c_k(j)\}_{k=1}^T, j = 0, 1, \dots, N$ are unknown coefficients. It follows from the last relation that the spectral characteristic $h(\theta)$ of the estimate is of the form

$$h(\theta) = A_N(e^{i\theta}) - \left(f^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>}. \quad (27)$$

The unknown coefficients $\tilde{c}(j), j = 0, 1, \dots, N$, are determined from the condition $h(\theta) \in L_N^\alpha(\mu)$, which gives us the system of equations

$$\int_{-\pi}^{\pi} e^{-i\theta k} h(\theta) d\theta = 0, \quad k = 0, 1, \dots, N. \quad (28)$$

These equations are of the form

$$\int_{-\pi}^{\pi} e^{-i\theta k} \left[\left(\sum_{j=0}^N \tilde{a}(j) e^{ij\theta} \right) - \left(f(\theta)^{-1} \left(\sum_{j=0}^N \tilde{c}(j) e^{-ij\theta} \right) \right)^{<\frac{1}{\alpha-1}>} \right] d\theta = 0, \quad k = 0, 1, \dots, N. \quad (29)$$

The variance of the optimal estimate of the functional is calculated by the formula

$$\left\| A_N \vec{\xi} - \hat{A}_N \vec{\xi} \right\|_\alpha^\alpha = \int_{-\pi}^{\pi} \left[\left(f^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \right]^\top f(\theta) \left(f^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{\alpha-1}{\alpha}>} d\theta. \quad (30)$$

In the case where the dimension of the sequence $T = 1$, this formula can be written in the form

$$\left\| A_N \xi - \hat{A}_N \xi \right\|_\alpha^\alpha = \int_{-\pi}^{\pi} \left| \left(f^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \right|^\alpha f(\theta) d\theta. \quad (31)$$

We can conclude that the following theorems hold true.

Theorem 3. Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T, j \in \mathbb{Z}$ be a vector-valued harmonizable symmetric α -stable, $HS\alpha S$, stochastic sequence which has absolutely continuous spectral measure $\mu(\theta)$ and the spectral density $f(\theta)$, satisfying the minimality condition (23). The optimal linear estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi} = \sum_{j=0}^N (\tilde{a}(j))^\top \vec{\xi}(j)$, that depends on the unknown values $\vec{\xi}(j), j = 0, 1, \dots, N$, of the sequence $\vec{\xi}(j)$, from observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ is calculated by the formula (25). The spectral characteristic $h(\theta)$ of the optimal estimate is determined by the equation (27), where the unknown coefficients $\tilde{c}(j), j = 0, \dots, N$, are determined from the system of equations (29). The variance of the optimal estimate of the functional is calculated by the formula (30) (by formula (31) in the case where dimension of the sequence $T = 1$).

Theorem 4. Let the vector-valued harmonizable symmetric α -stable, $HS\alpha S$, stochastic sequence $\vec{\xi}(n) = \{\xi(nT + k)\}_{k=0}^{T-1}, n \in \mathbb{Z}$, that corresponds to periodically harmonizable symmetric α -stable, $PHS\alpha S$, stochastic sequence $\{\xi(n), n \in \mathbb{Z}\}$, has absolutely continuous spectral measure $\mu(\theta)$ and the spectral density $f(\theta)$, satisfying the minimality condition (23). The optimal linear estimate $\hat{A}_N \vec{\xi}$ of the functional

$$A_N \vec{\xi} = \sum_{j=0}^N (\tilde{a}(j))^\top \vec{\xi}(j) = \sum_{j=0}^N \sum_{k=0}^{T-1} a_k(j) \xi(jT + k),$$

$\tilde{a}(j) = \{a_k(j)\}_{k=0}^{T-1}, \vec{\xi}(j) = \{\xi(jT + k)\}_{k=0}^{T-1}$, that depends on the unknown values $\xi(j), j = 0, 1, \dots, T(N+1) - 1$, of the sequence $\xi(j)$ from observations of the sequence $\xi(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, T(N+1) - 1\}$ is calculated by the formula (25). The spectral characteristic $h(\theta)$ of the optimal estimate is of the form (27), where the unknown coefficients $\tilde{c}(j) = \{c_k(j)\}_{k=0}^{T-1}, j = 0, 1, \dots, T(N+1) - 1$, are determined from the system of equations (29). The variance of the optimal estimate of the functional is calculated by the formula (30) (by formula (31) in the case where period $T = 1$).

Example. Consider the problem of the optimal linear estimation of the functional $A_0 \xi = a\xi(0)$, that depends on the unknown value $\xi(0)$ of a harmonizable symmetric α -stable, $HS\alpha S$, stochastic sequence $\xi(n)$ that has absolutely continuous spectral

measure $\mu(\theta)$ and the spectral density $f(\theta)$, satisfying the minimality condition (23), from observations of the sequence $\xi(n)$ at points $n \in \mathbb{Z} \setminus \{0\}$ (Moklyachuk, & Ostapenko, 2016).

In this case the spectral characteristic $h(\theta)$ of the optimal estimate of the functional is of the form

$$h(\theta) = a - c^{\langle \frac{1}{\alpha-1} \rangle} (f(\theta))^{\frac{-1}{\alpha-1}}.$$

The variance of the optimal estimate of the functional is calculated by the formula

$$\|\hat{A}_0\xi - A_0\xi\|_\alpha^\alpha = \int_{-\pi}^{\pi} \left| c^{\langle \frac{1}{\alpha-1} \rangle} (f(\theta))^{\frac{-1}{\alpha-1}} \right|^\alpha f(\theta) d\theta,$$

where the constant c is a solution of the equation

$$\int_{-\pi}^{\pi} \left(a - c^{\langle \frac{1}{\alpha-1} \rangle} (f(\theta))^{\frac{-1}{\alpha-1}} \right) d\theta = 0,$$

$$c = \frac{(2\pi a)^{\langle \alpha-1 \rangle}}{\left(\int_{-\pi}^{\pi} (f(\theta))^{\frac{-1}{\alpha-1}} d\theta \right)^{\langle \alpha-1 \rangle}}.$$

In the case of $\alpha = 2$ we have the Kolmogorov (Kolmogorov, 1992) results

$$h(\theta) = a - c(f(\theta))^{-1},$$

$$c = a \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} ((f(\theta))^{-1}) d\theta \right)^{-1},$$

$$\|\hat{A}\xi - A\xi\|_2^2 = 2\pi|a|^2 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} ((f(\theta))^{-1}) d\theta \right)^{-1}.$$

4. Interpolation problem. Minimax approach

The value of the error

$$\Delta(h(f_0, g_0); f, g) = \left\| A_N \vec{\xi} - \hat{A}_N \vec{\xi} \right\|_\alpha^\alpha$$

and the spectral characteristic $h(f, g) = h(\theta)$ of the optimal estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi} = \sum_{j=0}^N (\vec{a}(j))^\top \vec{\xi}(j)$ that depends on the unknown values $\vec{\xi}(j)$, $j = 0, 1, \dots, N$, of the sequence $\vec{\xi}(j)$, from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ can be calculated by the proposed formulas only in the case where we know the spectral densities $f(\theta)$ and $g(\theta)$ of the mutually independent harmonizable symmetric α -stable $HS_\alpha S$, stochastic sequences $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$ and $\vec{\eta}(j) = \{\eta_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$.

However, in practice we can't exactly evaluate the spectral densities of stochastic sequences, but, instead, we often can have a set $D = D_f \times D_g$ of admissible spectral densities. In this case we can apply the minimax-robust method of estimation to the interpolation problem. This method let us find an estimate that minimizes the maximum of the errors for all spectral densities from the given set $D = D_f \times D_g$ of admissible spectral densities simultaneously (Moklyachuk, & Masyutka, 2012; Moklyachuk, & Golichenko, 2016).

Definition 5. For a given class of spectral densities $D = D_f \times D_g$ the spectral densities $f_0(\theta) \in D_f$, $g_0(\theta) \in D_g$ are called the least favorable in $D = D_f \times D_g$ for the optimal linear estimation $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$, if the following relation hold true

$$\Delta(f_0, g_0) = \Delta(h(f_0, g_0); f_0, g_0) = \max_{(f, g) \in D_f \times D_g} \Delta(h(f, g); f, g).$$

Definition 6. For a given class of spectral densities $D = D_f \times D_g$ the spectral characteristic $h^0 = h(f_0)$ of the optimal estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ is called minimax (robust) for the optimal linear estimation $A_N \vec{\xi}$, if the following relations hold true

$$h^0(\theta) \in H_D = \bigcap_{(f, g) \in D_f \times D_g} L^\alpha(f + g),$$

$$\min_{h \in H_D} \max_{(f, g) \in D} \Delta(h; f, g) = \max_{(f, g) \in D} \Delta(h^0; f, g).$$

The least favorable spectral densities $f_0(\theta)$, $g_0(\theta)$ and the minimax spectral characteristic $h^0 = h(f_0, g_0)$ form a saddle point of the function $\Delta(h; f, g)$ on the set $H_D \times D$. The saddle point inequalities

$$\Delta(h; f_0, g_0) \geq \Delta(h^0; f_0, g_0) \geq \Delta(h^0; f, g), \quad \forall h \in H_D, \quad \forall f \in D_f, \quad \forall g \in D_g$$

hold true if $h^0 = h(f_0, g_0)$ and $h(f_0, g_0) \in H_D$, where (f_0, g_0) is a solution to the conditional extremum problem

$$\max_{(f, g) \in D_f \times D_g} \Delta(h(f_0, g_0); f, g) = \Delta(h(f_0, g_0); f_0, g_0), \quad (32)$$

$$\Delta(h(f_0, g_0); f, g) = \left\| A_N \vec{\xi} - \hat{A}_N \vec{\xi} \right\|_{\alpha}^{\alpha} = \int_{-\pi}^{\pi} \left(A_N(e^{i\theta}) - h^0(\theta) \right)^{\top} f(\theta) \left(A_N(e^{i\theta}) - h^0(\theta) \right)^{<\alpha-1>} d\theta + \int_{-\pi}^{\pi} \left(h^0(\theta) \right)^{\top} g(\theta) \left(h^0(\theta) \right)^{<\alpha-1>} d\theta. \quad (33)$$

The constrained optimization problem (32) is equivalent to the unconditional extremum problem

$$\Delta_D(f, g) = -\Delta(h(f_0, g_0); f, g) + \delta(f, g | D_f \times D_g) \rightarrow \inf, \quad (34)$$

where $\delta(f, g | D_f \times D_g)$ is the indicator function of the set $D = D_f \times D_g$. Solution (f_0, g_0) to the problem (34) is characterized by the condition $0 \in \partial \Delta_D(f_0, g_0)$, where $\partial \Delta_D(f_0)$ is the subdifferential of the convex functional $\Delta_D(f, g)$ at point (f_0, g_0) . This condition makes it possible to find the least favorable spectral densities in some special classes of spectral densities $D = D_f \times D_g$ (Ioffe, & Tihomirov, 1979; Pshenichnyj, 1971; Rockafellar, 1970).

Note, that the form of the functional $\Delta(h(f_0, g_0); f, g)$ is convenient for application the Lagrange method of indefinite multipliers for finding solution to the problem (32). Making use the method of Lagrange multipliers and the form of subdifferentials of the indicator functions we describe relations that determine least favourable spectral densities in some special classes of spectral densities.

Summing up the derived formulas and the introduced definitions we come to conclusion that the following lemmas hold true

Lemma 1. Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$ and $\vec{\eta}(j) = \{\eta_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$ be mutually independent harmonizable symmetric α -stable random sequences which have absolutely continuous spectral measures and the spectral densities $f_0(\theta)$ and $g_0(\theta)$ satisfying the minimality condition (15). Let the spectral densities $(f_0, g_0) \in D_f \times D_g$ give a solution to the constrained optimization problem (32). The spectral densities (f_0, g_0) are the least favorable ones in $D_f \times D_g$ and $h^0 = h(f_0, g_0)$ is the minimax spectral characteristic of the optimal linear estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ that depends on the unknown values $\vec{\xi}(j)$, $j = 0, 1, \dots, N$, of the sequence $\vec{\xi}(j)$, from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$, if $h^0 = h(f_0, g_0) \in H_D$.

Lemma 2. Let $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, $j \in \mathbb{Z}$ be a harmonizable symmetric α -stable random sequence which has absolutely continuous spectral measure and the spectral density $f_0(\theta)$ satisfying the minimality condition (23). Let the spectral density $f_0 \in D_f$ gives a solution to the constrained optimization problem

$$\max_{f \in D_f} \Delta(h(f_0); f) = \Delta(h(f_0); f_0), \quad (35)$$

where

$$\Delta(h(f_0); f) = \left\| A_N \vec{\xi} - \hat{A}_N \vec{\xi} \right\|_{\alpha}^{\alpha} = \int_{-\pi}^{\pi} \left[\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \right]^{\top} f(\theta) \left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{\alpha-1}{\alpha}>} d\theta. \quad (36)$$

The spectral density f_0 is the least favorable spectral density in D_f and $h^0 = h(f_0)$ is the minimax spectral characteristic of the optimal linear estimate $\hat{A}_N \vec{\xi}$ of the functional $A_N \vec{\xi}$ that depends on the unknown values $\vec{\xi}(j)$, $j = 0, 1, \dots, N$, of the sequence $\vec{\xi}(j)$, from observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$, if $h^0 = h(f_0) \in H_D$.

5. Least favorable spectral density in classes $D = D_0 \times D_{\varepsilon}$

Consider the problem of the mean-square optimal interpolation of the functional $A_N \vec{\xi}$ in the case when spectral densities of sequences belong to the class of admissible spectral densities $D = D_0^k \times D_{\varepsilon}^k$, $k = 1, 2, 3, 4$, where

$$\begin{aligned} D_0^1 &= \left\{ f(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr } f(\theta) d\theta = p \right. \right\}, \\ D_{\varepsilon}^1 &= \left\{ g(\theta) \left| \text{Tr } g(\theta) = (1 - \varepsilon) \text{Tr } g_1(\theta) + \varepsilon \text{Tr } w(\theta), \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr } g(\theta) d\theta = q \right. \right\}; \\ D_0^2 &= \left\{ f(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{kk}(\theta) d\theta = p_k, k = \overline{1, T} \right. \right\}, \\ D_{\varepsilon}^2 &= \left\{ g(\theta) \left| g_{kk}(\theta) = (1 - \varepsilon) g_{kk}^1(\theta) + \varepsilon w_{kk}(\theta), \frac{1}{2\pi} \int_{-\pi}^{\pi} g_{kk}(\theta) d\theta = q_k, k = \overline{1, T} \right. \right\}; \\ D_0^3 &= \left\{ f(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \langle B_1, f(\theta) \rangle d\theta = p \right. \right\}, \\ D_{\varepsilon}^3 &= \left\{ g(\theta) \left| \langle B_2, g(\theta) \rangle = (1 - \varepsilon) \langle B_2, g_1(\theta) \rangle + \varepsilon \langle B_2, w(\theta) \rangle, \frac{1}{2\pi} \int_{-\pi}^{\pi} \langle B_2, g(\theta) \rangle d\theta = q \right. \right\}; \\ D_0^4 &= \left\{ f(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) d\theta = P \right. \right\}, \\ D_{\varepsilon}^4 &= \left\{ g(\theta) \left| g(\theta) = (1 - \varepsilon) g_1(\theta) + \varepsilon w(\theta), \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta) d\theta = Q \right. \right\}, \end{aligned}$$

where $g_1(\theta)$ is a known and fixed spectral density matrix while $w(\theta)$ is an unknown spectral density matrix, $p, q, p_k, q_k, k = \overline{1, T}$ are given numbers, P, Q, B_1, B_2 are given matrices.

From the condition $0 \in \partial \Delta_D(f_0, g_0)$ we find the following equations which determine the least favourable spectral densities for these given sets of admissible spectral densities.

For the first pair $D_0^1 \times D_\varepsilon^1$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta) \right) \left((A_N(e^{i\theta}) - h^0(\theta))^T \right)^{<\alpha-1>} = \alpha^2 E, \quad (37)$$

$$h^0(\theta) \left((h^0(\theta))^T \right)^{<\alpha-1>} = \beta^2 + \gamma(\theta), \quad (38)$$

where $\gamma(\theta) \leq 0$ and $\gamma(\theta) = 0$ if $\text{Tr } g_0(\theta) > (1 - \varepsilon) \text{Tr } g_1(\theta)$.

For the second pair $D_0^2 \times D_\varepsilon^2$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta) \right) \left((A_N(e^{i\theta}) - h^0(\theta))^T \right)^{<\alpha-1>} = \{ \alpha_k^2 \delta_{kl} \}_{k,l=1}^T, \quad (39)$$

$$h^0(\theta) \left((h^0(\theta))^T \right)^{<\alpha-1>} = \{ (\beta_k^2 + \gamma_k(\theta)) \delta_{kl} \}_{k,l=1}^T, \quad (40)$$

where $\gamma_k(\theta) \leq 0$ and $\gamma_k(\theta) = 0$ if $g_{kk}^0(\theta) > (1 - \varepsilon) g_{kk}^1(\theta)$.

For the third pair $D_0^3 \times D_\varepsilon^3$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta) \right) \left((A_N(e^{i\theta}) - h^0(\theta))^T \right)^{<\alpha-1>} = \alpha^2 B_1, \quad (41)$$

$$h^0(\theta) \left((h^0(\theta))^T \right)^{<\alpha-1>} = (\beta^2 + \gamma'(\theta)) B_2, \quad (42)$$

where $\gamma'(\theta) \leq 0$ and $\gamma'(\theta) = 0$ if $\langle B_2, g_0(\theta) \rangle > (1 - \varepsilon) \langle B_2, g_1(\theta) \rangle$.

For the fourth pair $D_0^4 \times D_\varepsilon^4$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta) \right) \left((A_N(e^{i\theta}) - h^0(\theta))^T \right)^{<\alpha-1>} = \vec{\alpha} \cdot \vec{\alpha}^*, \quad (43)$$

$$h^0(\theta) \left((h^0(\theta))^T \right)^{<\alpha-1>} = \vec{\beta} \cdot \vec{\beta}^* + \Gamma(\theta), \quad (44)$$

where $\Gamma(\theta) \leq 0$ and $\Gamma(\theta) = 0$ if $g_0(\theta) > (1 - \varepsilon) g_1(\theta)$.

Thus, the following statements hold true.

Theorem 5. Let the minimality condition (15) hold true. The least favorable spectral densities $f_0(\theta), g_0(\theta)$ in the classes $D_0 \times D_\varepsilon$ for the optimal linear interpolation of the functional $A_N \vec{\xi}$ are determined by relations (37), (38) for the first pair $D_0^1 \times D_\varepsilon^1$ of sets of admissible spectral densities; (39), (40) for the second pair $D_0^2 \times D_\varepsilon^2$ of sets of admissible spectral densities; (41), (42) for the third pair $D_0^3 \times D_\varepsilon^3$ of sets of admissible spectral densities; (43), (44) for the fourth pair $D_0^4 \times D_\varepsilon^4$ of sets of admissible spectral densities; constrained optimization problem (32) and restrictions on densities from the corresponding classes $D_0 \times D_\varepsilon$. The minimax-robust spectral characteristic of the optimal estimate of the functional $A_N \vec{\xi}$ is determined by the formula (19).

Corollary 1. Let the minimality condition (23) hold true. The least favorable spectral densities $f_0(\theta)$ in the classes D_0^k , $k = 1, 2, 3, 4$, for the optimal linear interpolation of the functional $A_N \vec{\xi}$, which depends on the unknown values of the sequence $\vec{\xi}(j)$ based on observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$, are determined by the following equations, respectively,

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))^T} \right)^{<\frac{\alpha-1}{\alpha-1}>} = \alpha^2 E, \quad (45)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))^T} \right)^{<\frac{\alpha-1}{\alpha-1}>} = \{ \alpha_k^2 \delta_{kl} \}_{k,l=1}^T, \quad (46)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))^T} \right)^{<\frac{\alpha-1}{\alpha-1}>} = \alpha^2 B_2, \quad (47)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))^T} \right)^{<\frac{\alpha-1}{\alpha-1}>} = \vec{\alpha} \cdot \vec{\alpha}^*, \quad (48)$$

constrained optimization problem (35) and restrictions on densities from the corresponding classes D_0^k , $k = 1, 2, 3, 4$. The minimax spectral characteristic of the optimal estimate of the functional $A_N \vec{\xi}$ is determined by the formula (27).

Corollary 2. Let the minimality condition (23) hold true. The least favorable spectral densities $f_0(\theta)$ in the classes D_ε^k , $k = 1, 2, 3, 4$, for the optimal linear interpolation of the functional $A_N \vec{\xi}$, which depends on the unknown values of the sequence $\vec{\xi}(j)$ based on observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$, are determined by the following equations, respectively,

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))^T} \right)^{<\frac{\alpha-1}{\alpha-1}>} = (\beta^2 + \gamma(\theta)) E, \quad (49)$$

$$\left(f_0^{-1}(\theta)C_N(e^{i\theta})\right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta)C_N(e^{i\theta}))^\top}\right)^{<\frac{\alpha-1}{\alpha-1}>} = \{(\beta_k^2 + \gamma_k(\theta))\delta_{kl}\}_{k,l=1}^T, \quad (50)$$

$$\left(f_0^{-1}(\theta)C_N(e^{i\theta})\right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta)C_N(e^{i\theta}))^\top}\right)^{<\frac{\alpha-1}{\alpha-1}>} = (\beta^2 + \gamma'(\theta))B_2, \quad (51)$$

$$\left(f_0^{-1}(\theta)C_N(e^{i\theta})\right)^{<\frac{1}{\alpha-1}>} \left(\overline{(f_0^{-1}(\theta)C_N(e^{i\theta}))^\top}\right)^{<\frac{\alpha-1}{\alpha-1}>} = \vec{\beta} \cdot \vec{\beta}^* + \Gamma(\theta), \quad (52)$$

constrained optimization problem (35) and restrictions on densities from the corresponding classes D_ε^k , $k = 1, 2, 3, 4$. The minimax spectral characteristic of the optimal estimate of the functional $A_N \vec{\xi}$ is determined by the formula (27).

6. Least favorable spectral density in classes $D = D_V^U \times D_{2\delta}$

Consider the problem of mean square optimal interpolation of the functional $A_N \vec{\xi}$ in the case when spectral densities of the sequences belong to the class of admissible spectral densities $D = D_V^U \times D_{2\delta}$,

$$\begin{aligned} D_V^{U1} &= \left\{ f(\theta) \left| \text{Tr } v(\theta) \leq \text{Tr } f(\theta) \leq \text{Tr } u(\theta), \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr } f(\theta) d\theta = p \right. \right\}, \\ D_{2\delta}^1 &= \left\{ g(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} |\text{Tr}(g(\theta) - g_1(\theta))|^2 d\theta \leq \delta \right. \right\}; \\ D_V^{U2} &= \left\{ f(\theta) \left| v_{kk}(\theta) \leq f_{kk}(\theta) \leq u_{kk}(\theta), \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{kk}(\theta) d\theta = p_k, k = \overline{1, T} \right. \right\}, \\ D_{2\delta}^2 &= \left\{ g(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} |g_{kk}(\theta) - g_{kk}^1(\theta)|^2 d\theta \leq \delta_k, k = \overline{1, T} \right. \right\}; \\ D_V^{U3} &= \left\{ f(\theta) \left| \langle B_1, v(\theta) \rangle \leq \langle B_1, f(\theta) \rangle \leq \langle B_1, u(\theta) \rangle, \frac{1}{2\pi} \int_{-\pi}^{\pi} \langle B_1, f(\theta) \rangle d\theta = p \right. \right\}, \\ D_{2\delta}^3 &= \left\{ g(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} |\langle B_2, g(\theta) - g_1(\theta) \rangle|^2 d\theta \leq \delta \right. \right\}; \\ D_V^{U4} &= \left\{ f(\theta) \left| v(\theta) \leq f(\theta) \leq u(\theta), \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) d\theta = P \right. \right\}, \\ D_{2\delta}^4 &= \left\{ g(\theta) \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} |g_{ij}(\theta) - g_{ij}^1(\theta)|^2 d\theta \leq \delta_{ij}, i, j = \overline{1, T} \right. \right\}, \end{aligned}$$

where spectral densities $v(\theta), u(\theta), g_1(\theta)$ are known and fixed. The class D_V^U describes the "strip" model of stochastic processes, $D_{2\delta}$ describes "δ-district" in the space L_2 of the given bounded spectral density $g_1(\theta)$.

From the condition $0 \in \partial \Delta_D(f_0, g_0)$ we find the following equations which determine the least favourable spectral densities for these given sets of admissible spectral densities.

For the first pair $D_V^{U1} \times D_{2\delta}^1$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta)\right) \left(\overline{(A_N(e^{i\theta}) - h^0(\theta))^\top}\right)^{<\alpha-1>} = (\alpha^2 + \gamma_1(\lambda) + \gamma_2(\lambda))E, \quad (53)$$

$$h^0(\theta) \left(\overline{(h^0(\theta))^\top}\right)^{<\alpha-1>} = \beta^2 \text{Tr}(g_0(\theta) - g_1(\theta))E, \quad (54)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\text{Tr}(g_0(\theta) - g_1(\theta))|^2 d\theta = \delta, \quad (55)$$

where $\gamma_1(\theta) \leq 0$ and $\gamma_1(\theta) = 0$ if $\text{Tr } f_0(\theta) > \text{Tr } v(\theta)$, $\gamma_2(\theta) \geq 0$ and $\gamma_2(\theta) = 0$ if $\text{Tr } f_0(\lambda) < \text{Tr } u(\theta)$.

For the second pair $D_V^{U2} \times D_{2\delta}^2$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta)\right) \left(\overline{(A_N(e^{i\theta}) - h^0(\theta))^\top}\right)^{<\alpha-1>} = \{(\alpha_k^2 + \gamma_{1k}(\theta) + \gamma_{2k}(\theta))\delta_{kl}\}_{k,l=1}^T, \quad (56)$$

$$h^0(\theta) \left(\overline{(h^0(\theta))^\top}\right)^{<\alpha-1>} = \{\beta_k^2(g_{kk}^0(\theta) - g_{kk}^1(\theta))\delta_{kl}\}_{k,l=1}^T, \quad (57)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_{kk}^0(\theta) - g_{kk}^1(\theta)|^2 d\theta = \delta_k, k = \overline{1, T}, \quad (58)$$

where $\gamma_{1k}(\theta) \leq 0$ and $\gamma_{1k}(\theta) = 0$ if $f_{kk}^0(\theta) > v_{kk}(\theta)$, $\gamma_{2k}(\theta) \geq 0$ and $\gamma_{2k}(\theta) = 0$ if $f_{kk}^0(\theta) < u_{kk}(\theta)$.

For the third pair $D_V^{U3} \times D_{2\delta}^3$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta)\right) \left(\overline{(A_N(e^{i\theta}) - h^0(\theta))^\top}\right)^{<\alpha-1>} = (\alpha^2 + \gamma'_1(\theta) + \gamma'_2(\theta))B_1, \quad (59)$$

$$h^0(\theta) \left(\overline{(h^0(\theta))^\top}\right)^{<\alpha-1>} = \beta^2 \langle B_2, g_0(\theta) - g_1(\theta) \rangle I, \quad (60)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\langle B_2, g_0(\theta) - g_1(\theta) \rangle|^2 d\theta = \delta, \quad (61)$$

where $\gamma'_1(\theta) \leq 0$ and $\gamma'_1(\theta) = 0$ if $\langle B_1, f_0(\theta) \rangle > \langle B_1, v(\theta) \rangle$, $\gamma'_2(\theta) \geq 0$ and $\gamma'_2(\theta) = 0$ if $\langle B_1, f_0(\theta) \rangle < \langle B_1, u(\theta) \rangle$, I is a matrix of ones.

For the fourth pair $D_V^{U^4} \times D_{2\delta}^4$ we have equations

$$\left(A_N(e^{i\theta}) - h^0(\theta) \right) \left(A_N(e^{i\theta}) - h^0(\theta) \right)^\top <^{\alpha-1} = \vec{\alpha} \cdot \vec{\alpha}^* + \Gamma_1(\theta) + \Gamma_2(\theta) \quad (62)$$

$$h^0(\theta) \left(h^0(\theta) \right)^\top <^{\alpha-1} = \{ \beta_{ij}(g_{ij}^0(\theta) - g_{ij}^1(\theta)) \}_{i,j=1}^T, \quad (63)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |g_{ij}^0(\theta) - g_{ij}^1(\theta)|^2 d\theta = \delta_{ij}, \quad i, j = \overline{1, T}, \quad (64)$$

where $\Gamma_1(\theta) \leq 0$ and $\Gamma_1(\theta) = 0$ if $f_0(\theta) > v(\theta)$, $\Gamma_2(\theta) \geq 0$ and $\Gamma_2(\theta) = 0$ if $f_0(\theta) < u(\theta)$.

The following theorem and corollaries hold true.

Theorem 6. Let the minimality condition (15) hold true. The least favorable spectral densities $f_0(\theta)$, $g_0(\theta)$ in the classes $D = D_V^U \times D_{2\delta}$ for the optimal linear interpolation of the functional $A_N \vec{\xi}$ are determined by relations (53) – (55) for the first pair $D_V^{U^1} \times D_{2\delta}^1$ of sets of admissible spectral densities; (56) – (58) for the second pair $D_V^{U^2} \times D_{2\delta}^2$ of sets of admissible spectral densities; (59) – (61) for the third pair $D_V^{U^3} \times D_{2\delta}^3$ of sets of admissible spectral densities; (62) – (64) for the fourth pair $D_V^{U^4} \times D_{2\delta}^4$ of sets of admissible spectral densities; constrained optimization problem (32) and restrictions on densities from the corresponding classes $D = D_V^U \times D_{2\delta}$. The minimax-robust spectral characteristic of the optimal estimate of the functional $A_N \vec{\xi}$ is determined by the formula (19).

Corollary 3. Let the minimality condition (23) hold true. The least favorable spectral densities $f_0(\theta)$ in the classes $D_V^{U^k}$, $k = 1, 2, 3, 4$, for the optimal linear interpolation of the functional $A_N \vec{\xi}$, which depends on the unknown values of the sequence $\vec{\xi}(j)$ based on observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$, are determined by the following equations, respectively,

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = (\alpha^2 + \gamma_1(\theta) + \gamma_2(\theta))E, \quad (65)$$

$$\begin{aligned} \left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = \\ = \{ (\alpha_k^2 + \gamma_{1k}(\theta) + \gamma_{2k}(\theta)) \delta_{kl} \}_{k,l=1}^T, \end{aligned} \quad (66)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = (\alpha^2 + \gamma'_1(\theta) + \gamma'_2(\theta))B_1, \quad (67)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = \vec{\alpha} \cdot \vec{\alpha}^* + \Gamma_1(\theta) + \Gamma_2(\theta), \quad (68)$$

constrained optimization problem (35) and restrictions on densities from the corresponding classes $D_V^{U^k}$. The minimax spectral characteristic of the optimal estimate of the functional $A_N \vec{\xi}$ is determined by the formula (27).

Corollary 4. Let the minimality condition (23) hold true. The least favorable spectral densities $f_0(\theta)$ in the classes $D_{2\delta}^k$, $k = 1, 2, 3, 4$, for the optimal linear interpolation of the functional $A_N \vec{\xi}$, which depends on the unknown values of the sequence $\vec{\xi}(j)$ based on observations of the sequence $\vec{\xi}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$, are determined by the following equations, respectively,

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = \beta^2 \text{Tr}(f_0(\theta) - g_1(\theta))E, \quad (69)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = \{ \beta_k^2 (f_{kk}^0(\theta) - g_{kk}^1(\theta)) \delta_{kl} \}_{k,l=1}^T, \quad (70)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = \beta^2 \langle B_2, f_0(\theta) - g_1(\theta) \rangle I, \quad (71)$$

$$\left(f_0^{-1}(\theta) C_N(e^{i\theta}) \right) <^{\frac{1}{\alpha-1}} \left(\overline{(f_0^{-1}(\theta) C_N(e^{i\theta}))}^\top \right) <^{\frac{\alpha-1}{\alpha-1}} = \{ \beta_{ij}(f_{ij}^0(\theta) - g_{ij}^1(\theta)) \}_{i,j=1}^T \quad (72)$$

constrained optimization problem (35) and the following restrictions on densities from the classes $D_{2\delta}^k$, $k = 1, 2, 3, 4$, respectively,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\text{Tr}(f_0(\theta) - g_1(\theta))|^2 d\theta = \delta, \quad (73)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f_{kk}^0(\theta) - g_{kk}^1(\theta)|^2 d\theta = \delta_k, \quad k = \overline{1, T}, \quad (74)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\langle B_2, f_0(\theta) - g_1(\theta) \rangle|^2 d\theta = \delta, \quad (75)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f_{ij}^0(\theta) - g_{ij}^1(\theta)|^2 d\theta = \delta_{ij}, \quad i, j = \overline{1, T}. \quad (76)$$

The minimax spectral characteristic of the optimal estimate of the functional $A_N \vec{\xi}$ is determined by the formula (27).

Discussion and conclusions

We propose methods for solving the optimal linear estimation problem for the linear functionals that depend on the unknown values $\vec{\xi}(j)$, $j = 0, 1, \dots, N$, of a vector-valued harmonizable symmetric α -stable random sequence $\vec{\xi}(j) = \{\xi_k(j)\}_{k=1}^T$, from observations of the sequence $\vec{\xi}(j) + \vec{\eta}(j)$ at points $j \in \mathbb{Z} \setminus \{0, 1, \dots, N\}$ where $\vec{\xi}(j)$ and $\vec{\eta}(j)$ are mutually independent harmonizable symmetric α -stable random sequences which have the spectral densities $f(\theta)$ and $g(\theta)$ satisfying the minimality condition.

The problem is investigated under the condition of spectral certainty as well as under the condition of spectral uncertainty. Formulas for calculating the value of the error and the spectral characteristic of the optimal linear estimate of the functional are derived under the condition of spectral certainty where spectral density of the sequence is exactly known. In the case where spectral density of the sequence is not exactly known, but a set of admissible spectral densities is available, relations which determine least favorable densities and the minimax-robust spectral characteristics for different classes of spectral densities are found.

Authors' contribution: Mikhail Moklyachuk – methodology, derivation of the main results and preparation of the paper; Oleksandr Masyutka – derivation of the main results and preparation of the paper.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Cambanis, S. (1983). Complex symmetric stable variables and processes. In Sen (Ed.), *Contributions to statistics: Essays in honour of Norman L. Johnson* (pp. 63–79). North-Holland, New York.
- Cambanis, S., & Soltani, R. (1984). Prediction of stable processes: Spectral and moving average representations. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 66, 593–612. <https://doi.org/10.1007/BF00531892>
- Franke, J. (1984). On the robust prediction and interpolation of time series in the presence of correlated noise. *Journal of Time Series Analysis*, 5(4), 227–244. <https://doi.org/10.1111/J.1467-9892.1984.TB00389.X>
- Franke, J. (1985). Minimax robust prediction of discrete time series. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 68, 337–364. <https://doi.org/10.1007/BF00532645>
- Franke, J., & Poor, H. V. (1984). Minimax-robust filtering and finite-length robust predictors. In Franke, Härdle, & Martin (Eds.), *Robust and nonlinear time series analysis* (pp. 87–126). Springer-Verlag. https://doi.org/10.1007/978-1-4615-7821-5_6
- Grenander, U. (1957). A prediction problem in game theory. *Arkiv för Matematik*, 3(4), 371–379. <https://doi.org/10.1007/BF02589429>
- Hosoya, Y. (1982). Harmonizable stable processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 60, 517–533. <https://doi.org/10.1007/BF00535714>
- Ioffe, A. D., & Tihomirov, V. M. (1979). *Theory of extremal problems*. North-Holland Publishing Company.
- Kassam, S., & Poor, H. (1985). Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, 73(3), 433–481. <https://doi.org/10.1109/PROC.1985.13167>
- Kolmogorov, A. (1992). *Selected works of A. N. Kolmogorov. Vol. II: Probability theory and mathematical statistics*. Dordrecht etc.: Kluwer Academic Publishers.
- Masani, P., & Wiener, N. (1957). The prediction theory of multivariate stochastic processes. *Acta Mathematica*, 98, 111–150. <https://doi.org/10.1007/BF02404472>
- Moklyachuk, M., & Golichenko, I. (2016). *Periodically correlated processes estimates*. LAP Lambert Academic Publishing.
- Moklyachuk, M., & Masyutka, O. (2012). *Minimax-robust estimation technique for stationary stochastic processes*. LAP Lambert Academic Publishing.
- Moklyachuk, M., & Ostapenko, V. (2016). Minimax interpolation of harmonizable sequences. *Theory of Probability and Mathematical Statistics*, 92, 135–146. <https://doi.org/10.1090/TPMS%2F988>
- Pourahmadi, M. (1984). On minimality and interpolation of harmonizable stable processes. *Siam Journal on Applied Mathematics*, 44(5), 1023–1030. <https://doi.org/10.1137/0144072>
- Pshenichnyj, B. (1971). *Necessary conditions for an extremum*. Marcel Dekker.
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press.
- Singer, I. (1970). *Best approximation in normed linear spaces by elements of linear subspaces*. Springer-Verlag.
- Vastola, K. S., & Poor, H. V. (1983). An analysis of the effects of spectral uncertainty on Wiener filtering. *Automatica*, 19(3), 289–293. <https://doi.org/10.1016/0005-1098%2883%2990105-X>
- Weron, A. (1985). Harmonizable stable processes on groups: spectral, ergodic and interpolation properties. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 68, 473–491. <https://doi.org/10.1007/BF00535340>

Отримано редакцією журналу / Received: 06.04.25

Прорецензовано / Revised: 24.09.25

Схвалено до друку / Accepted: 10.10.25

Олександр МАСЮТКА, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0002-7301-8813
e-mail: omasyutka@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

Михайло МОКЛЯЧУК, д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0002-6173-0280
e-mail: moklyachuk@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

ЗАДАЧА ИНТЕРПОЛЯЦИИ ВЕКТОРНЫХ ГАРМОНИЗОВАННЫХ СТИЙКИХ ПОСЛЕДОВАТЕЛЕЙ

Задача оцінювання невідомих значень стохастичних процесів є актуальною проблемою як у теорії, так і в прикладних застосуваннях стохастичних процесів. Проблеми прогнозування майбутніх значень економічних і фізичних процесів, відновлення втраченої інформації, очищення сигналу або інших даних зі спостережень із шумом гостро постають в насиченому інформацією світі. З огляду на це розроблення методів оцінювання є одним з основних завдань сучасної теорії стохастичних процесів. У пропонованій статті розглянуто задачу оптимальної лінійної інтерполяції функціонала від невідомих значень векторної гармонізованої симетричної альфа-стійкої випадкової послідовності за спостереженнями послідовності із шумом. Використано класичний підхід для виведення формул для обчислення значень середньоквадратичної похибки та спектральної характеристики оптимальної лінійної оцінки функціонала. Основним припущенням цього підходу є те, що спектральні щільності наявних стохастичних послідовностей точно відомі. Однак на практиці отримати повну інформацію про спектральні щільності в більшості випадків неможливо. У цьому випадку знаходять параметричну або непараметричну оцінку невідомої спектральної щільності, а потім застосовують один із традиційних методів оцінювання за умови, що обрана щільність є істинною. Цей підхід може привести до значного зростання значення похибки оцінювання. Для подолання цього ефекту можна шукати оцінки, які є оптимальними для всіх щільностей з певного класу допустимих спектральних щільностей. Ці оцінки називають мінімаксними, оскільки вони мінімізують максимальні значення похибок оцінок для всіх щільностей із заданого класу. Тому у випадку спектральної невизначеності ми використовуємо мінімаксний підхід і пропонуємо формули, що визначають найменш сприятливі спектральні щільності та мінімаксні спектральні характеристики оптимальних оцінок функціонала для деяких класів допустимих спектральних щільностей.

Ключові слова: гармонізована стійка випадкова послідовність, періодично гармонізована стійка випадкова послідовність, оптимальна лінійна оцінка, мінімаксна (робастна) оцінка, мінімаксна спектральна характеристика, найменш сприятлива спектральна щільність.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 517.91

DOI: <https://doi.org/10.17721/1812-5409.2025/2.8>

Alexandr PRISHLYAK, DSc (Phys. & Math.), Prof.

ORCID ID: 0000-0002-7164-807X

e-mail: prishlyak@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Kateryna BUTOK, PhD, Researcher

ORCID ID: 0009-0003-3000-4416

e-mail: katro.butok@gmail.com

Otto von Guericke Universität, Magdeburg, Germany

TOPOLOGICAL STRUCTURE OF SIMPLE HAMILTONIAN FLOWS ON THE PROJECTIVE PLANE AND THE KLEIN BOTTLE

For every non-orientable surface, there exists an oriented double cover. We consider flows whose lift to the double cover are Hamiltonian flows with a Hamiltonian being a Morse function. By analogy with oriented manifolds, a flow is simple if there are no saddle connections between different saddles within it.

To describe the topological properties of functions on two-dimensional manifolds, an invariant known as the Reeb graph is used. In the case of simple Morse functions on closed oriented two-dimensional manifolds, this is a complete topological invariant for stratified equivalence and topological equivalence of Hamiltonian flows. However, in our case, on a non-orientable manifold, there is no function that generates a Hamiltonian flow. Therefore, the concept of the Reeb graph requires redefinition.

To investigate the topological properties of simple Hamiltonian flows on non-orientable surfaces, we describe the structure of all possible atoms, which are analogous to the Bolsinov-Fomenko atoms for Morse functions and Hamiltonian flows on oriented surfaces. In particular, there are five types of such atoms on non-orientable surfaces. We construct a complete topological invariant, which is the Reeb graph with marked vertices and half-edges incident to the vertices, describe its properties, and prove a criterion for the topological equivalence of flows. For the projective plane and the Klein bottle, we prove theorems about the existence of flows with a given Reeb graph. For cyclic Reeb graphs on the Klein bottle, we describe an algorithm for their enumeration. For all other types of Reeb graphs on the projective plane and the Klein bottle, we obtain recursive formulas for determining their numbers. The results provide calculations of the number of topological structures of flows with no more than eight saddles. The obtained results open up the possibility of studying the structures of Hamiltonian flows of greater complexity on compact non-orientable surfaces.

Key words: topological equivalence, Reeb graph, classification, non-orientable surface.

AMS 2020 classification: 37C10, 37C15, 37C20.

Introduction

For oriented closed surfaces, the trajectories of Hamiltonian flows coincide with the level sets (components of the level curve) of the Hamiltonian function, which means that the topological structure of Hamiltonian flows is determined by the topological structure of the functions. The functions in general position are simple Morse functions, where each critical level contains one critical point, and this point is non-degenerate. For flows, simplicity means the absence of trajectories connecting saddle points.

For every non-oriented surface, there exists a unique (up to homeomorphism) oriented double cover. Therefore, functions and flows on non-oriented surfaces correspond to $\mathbb{Z}_2$ -invariant functions and flows on oriented surfaces. For example, functions on the projective plane $\mathbb{R}P^2$ (the Klein bottle Kl) can be defined as even functions on the sphere S^2 (the torus T^2). However, such functions do not define a flow on a non-oriented surface, as they do not allow for the alignment of movement directions along the level sets on the non-oriented surface. Conversely, if we use odd functions instead of even ones, then the directions of movement along the trajectories of their Hamiltonian flows are consistent with the symmetry with respect to the origin. It is precisely such flows, defined by odd Morse functions, that we consider in this paper.

If we collapse each component of the function level on a closed 2-manifold to a point, we obtain a graph known as the Reeb graph (Reeb, 1946) or the Liapunov graph for the Liapunov function (de Rezende, & Franzosa, 1993). In the case of simple Morse functions on closed oriented two-dimensional manifolds, this is a complete topological invariant for fibre equivalence. If a linear order is given on the set of vertices, it becomes a complete topological invariant with respect to topological equivalence. Automorphisms of Reeb graphs was described in (Kravchenko, & Maksymenko, 2019).

In the case of non-orientable two-dimensional manifolds, as well as manifolds with boundaries, additional information is needed to construct a complete topological invariant. Another way to determine the topological structure of Morse functions is to use Morse vector fields with specified function values at critical points. Therefore, the topological classification of Morse vector fields is closely related to the classification of functions and Hamiltonian flows.

The topological classification of Morse vector fields for closed surfaces was obtained in (Cobo, Gutierrez, & Llibre, 2010; Oshemkov, & Sharko, 1998; Maksymenko, 2009; Neumann, & O'Brien, 1976; Nikolaev, & Zhuzhoma, 1999; Prishlyak, & Loseva, 2020; Wang, 1990; Yokoyama, 2022).

Moreover, topological invariants of Morse functions on surfaces were constructed in (Bolsinov, & Fomenko, 2004; Hladys, & Prishlyak, 2019; Khohlyk, & Maksymenko, 2020; Maksymenko, 2006, 2020; Prishlyak, 2000; Sharko, 2003). The object of research is flows on non-orientable surfaces that do not have connections between different saddles, and whose preimages on a two-sided orientable cover are Hamiltonian flows for Morse functions.

The aim and objectives of the research is to investigate the topological properties of Reeb graphs of simple Hamiltonian flows on the projective plane and the Klein bottle, and to count the number of their different structures for flows with no more than 8 saddle points.

The methods of the research are following: 1) Reeb and Liapunov graph methods, 2) atoms and molecules method according to Bolsinov-Fomenko, 3) topological graph theory, 4) combinatorial methods for enumerating trees.

The structure of the paper. In Section 1, we describe methods for defining flows on the projective plane and the Klein bottle. In Section 2, we study the local structure of simple Hamiltonian flows on non-orientable surfaces. To do this, we describe atoms that are analogous to the Bolsinov-Fomenko atoms for Morse functions. All atoms are divided into five types, and the structure of flows on each type is described. In Section 3, we investigate a global invariant that is analogous to the Reeb graphs of simple Morse functions. The properties of this graph are outlined in Theorem 1, while Theorem 2 provides a criterion for the topological equivalence of flows using Reeb graphs. Section 4 describes all possible structures of simple Hamiltonian flows on the projective plane (Theorem 3) and derives a formula for the number of Reeb trees (Lemma 1), which is used to prove a recursive formula for the number of flows on the projective plane (Theorem 4). In Section 5, a recursive formula for Reeb trees with two fixed leaves is derived. Section 6 describes all possible structures of Reeb graphs on the Klein bottle. Theorem 5 proves the realization of the Reeb graph by a flow, and Theorem 6 addresses the number of flow structures with a given number of saddles on the Klein bottle. Section 7 outlines a method for counting the number of Reeb graphs with a cycle (the last type of flows on the Klein bottle). Section 8 presents examples and computational results using the derived formulas.

1. Flows on the Projective Plane and the Klein Bottle

The real projective plane $\mathbb{RP}^2$ is the set of one-dimensional subspaces of three-dimensional Euclidean space. A one-dimensional subspace is a line that passes through the origin. The distance between points in projective space is defined as the sector between the lines that represent them. The projective plane can also be viewed as a quotient space of the sphere S^2 under an equivalence relation where each equivalence class consists of a pair of antipodal points. In this case, the sphere serves as a double-sheeted oriented cover over the projective plane. The projection of the sphere onto the projective plane allows us to construct a flow on the sphere from a flow on the projective plane. The corresponding vector field at symmetric points on the sphere is defined by opposite vectors.

If we cut out a regular neighborhood of any point from the projective plane, we obtain a set that is homeomorphic to the Möbius strip. Additionally, the projective plane can conveniently be represented as a two-dimensional disk D^2 where diametrically opposite points on the boundary are identified. In this case, the flow V at opposite points must satisfy the property in polar coordinates (ρ, φ) such that $V_\rho(1, \varphi) = -V_\rho(1, \varphi + \pi)$, $V_\varphi(1, \varphi) = V_\varphi(1, \varphi + \pi)$. In particular, if the flow is tangent to the boundary, then at opposite points it defines the same orientation of the boundary.

The Klein bottle Kl can conveniently be defined as a quotient space of the square $[0, 1] \times [0, 1]$ under an equivalence relation generated by the following relations: $(0, v) \sim (1, v)$, $v \in [0, 1]$, $(u, 0) \sim (1 - u, 1)$, $u \in [0, 1]$. Then, the vector field $V = \{V_1, V_2\}$ on the Klein bottle is defined by a vector field on the square with coordinates (u, v) under the conditions $V(0, v) = V(1, v)$, $V_1(u, 0) = -V_1(1 - u, 1)$, $V_2(u, 0) = V_2(1 - u, 1)$, $u, v \in [0, 1]$.

If we consider the torus as a surface of revolution in $\mathbb{R}^3$, then by identifying points that are symmetric with respect to the center of symmetry of the torus, we obtain a Klein bottle. This allows us to view the torus as a double cover over the Klein bottle. Therefore, vector fields on the Klein bottle correspond to vector fields on the torus that are symmetric with respect to the center of symmetry of the torus.

Note that oriented double covers over non-oriented surfaces are uniquely defined up to homeomorphism.

By analogy with flows on the projective plane and the Klein bottle, we can define flows on other non-oriented closed surfaces. For a non-oriented surface of odd genus, we attach empty handles to a sphere (taking the connected sum with an even number of tori at pairs of symmetric points). On the resulting surface, we define a flow that is symmetric with respect to its center of symmetry. After factoring by the action of the group $\mathbb{Z}/2\mathbb{Z}$ as central symmetry, we obtain a flow on a non-oriented surface of odd genus. Flows on non-oriented surfaces of even genus can be defined using a similar construction, starting with a torus instead of a sphere.

2. Atoms of Simple Hamiltonian Flows on Non-Oriented Surfaces

Definition. A flow on a non-orientable surface is called *Hamiltonian* if its lift to the orientable double cover is a Hamiltonian flow. A Hamiltonian flow on a non-orientable surface is called *simple* if each of its separatrices forms a loop, and the Hamiltonian of the lift to the orientable double cover is a Morse function.

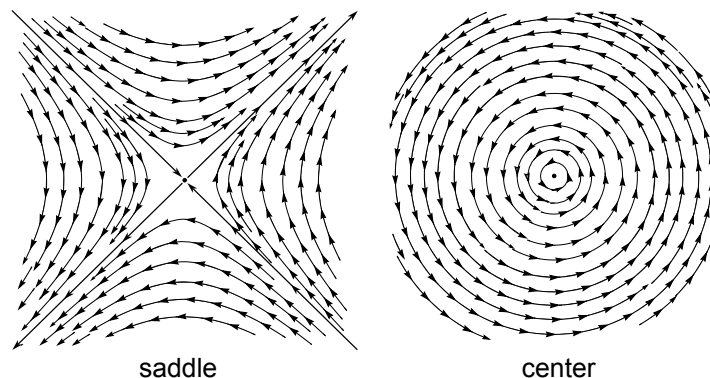


Fig. 1. Singular points of simple Hamiltonian flows

Since in a Morse function the critical points are non-degenerate, they can be of three types: local minima (there exists a coordinate system x, y such that in the neighborhood of this point the function $f(x, y) = x^2 + y^2$), saddles ($f(x, y) = x^2 - y^2$), and local maxima ($f(x, y) = -x^2 - y^2$). For a saddle, the singular point of the Hamiltonian vector field is also a saddle, $X(x, y) = \{y, x\}$, while for local extrema it is a center, $X(x, y) = \{-y, x\}$ (see Fig. 1).

Under the condition of simplicity of the flow, the trajectories entering and exiting a saddle point form loops. Therefore, they create a graph that is homeomorphic to 8. The neighborhood of this graph is called the atom of the saddle point if its boundary consists of closed trajectories of the flow and the neighborhood does not contain other singular points (except for the point belonging to 8). If this neighborhood is oriented, then it is homeomorphic to a three-holed sphere (a sphere with three holes), and in it, the flow is a Hamiltonian flow of a Morse function with one saddle point and constant on the boundary components. We denote this atom as A_0 (Fig. 2).

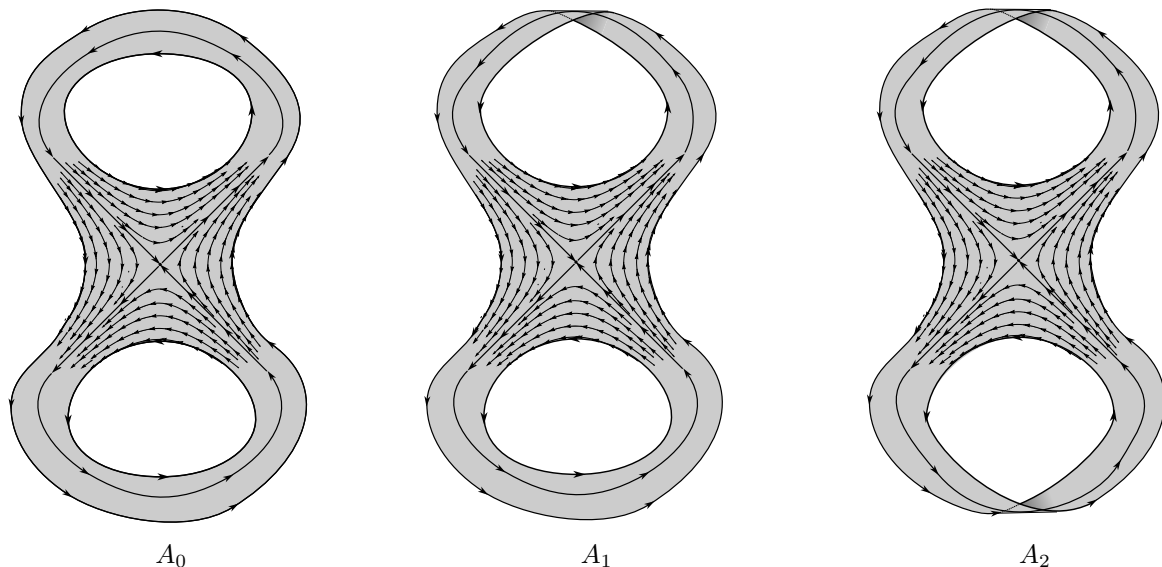


Fig. 2. Atoms of Saddle Points

Example 1. On the torus, as a surface of revolution ($x = (2 + \cos u) \cos v$, $y = (2 + \cos u) \sin v$, $z = \sin u$), we consider a Hamiltonian flow generated by the function $f(x, y, z) = x$. Then, the atom A_0 can be taken as the set $A_0 = f^{-1}([-3/2, -1/2])$. If, on the torus, we identify points that are symmetric with respect to the origin of coordinates, we obtain the atom A_0 on the Klein bottle.

In other cases, we have unoriented atoms. There are two possible cases depending on the number of boundary components of the atom. Each of these cases can be obtained from an oriented atom through the following procedure. Consider a regular point on one of the loops. We draw a transversal to the flow that intersects the boundary of the atom at two points. We cut the atom along this transversal and glue it back together, using a reflection that reverses the orientation and is constant at the point of the loop. In the first variant, we apply this operation to one loop, and in the second – to each of the loops.

Atom A_1 . If the neighborhood of one of the two loops of the atom is homeomorphic to a Möbius strip, and the other loop is homeomorphic to a cylinder, then the atom has two boundary components. Thus, the atom is homeomorphic to the projective plane with two holes (or, equivalently, a Möbius strip with a hole). The double cover of this atom is a sphere with four holes.

Example 2. An example of such a cover is the neighborhood U of the critical level (preimage of 0) of the function $f(x, y, z) = x^3 - y^3 + xyz$ on the unit sphere $x^2 + y^2 + z^2 = 1$, $U = f^{-1}([-0.01, 0.01])$ (Fig. 3). If this function is extended to the sphere so that each component of the complement to the atom contains one critical point that is a local extremum, then it serves as a Hamiltonian for the Hamiltonian flow on the sphere. Upon identifying symmetric points, we obtain a Hamiltonian flow on the projective plane with one saddle and two centers.

Atom A_2 . If the neighborhood of each of the two loops is homeomorphic to a Möbius strip, then the atom has one boundary component. Therefore, the atom is an unoriented surface of genus 2 with one boundary component – the Klein bottle with a hole. The double cover of this atom is a torus with two holes.

Example 3. In local coordinates u, v on the torus, the double cover is the neighborhood of the preimage of 0 of the function $f(u, v) = \sin u + \sin v$, $u \in [0, 2\pi]$, $v \in [0, 2\pi]$. The atom A_2 is obtained from $[0, 2\pi] \times [0, \pi] / \sim$, where $(u, 0) \sim (2\pi - u, \pi)$, $u \in [0, 2\pi]$, by removing a neighborhood of the point $(\pi/2, \pi/2)$. The field of the skew gradient of the function f defines a Hamiltonian flow on A_2 . By extending it to the flow on the removed neighborhood, we obtain a Hamiltonian flow on the Klein bottle with one saddle and one center.

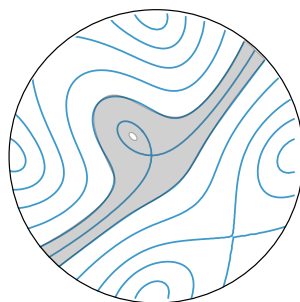


Fig. 3. Atom A_1 . The gray region is the image of U by the orthogonal projection of the hemisphere

Definition 1. A closed trajectory is called *regular* if it has a neighborhood homeomorphic to $S^1 \times [0, 1]$. A closed trajectory is called *singular* if its regular neighborhood is homeomorphic to a Möbius strip.

The atom of the center is called its regular neighborhood, the boundary of which is a closed trajectory. We denote it as C . Thus, C is homeomorphic to D^2 (Fig. 4).

An atom B (Fig. 4) be called a regular neighborhood of a singular closed trajectory, the boundary of which is a regular closed trajectory of the flow. Thus, B is homeomorphic to a Möbius strip.

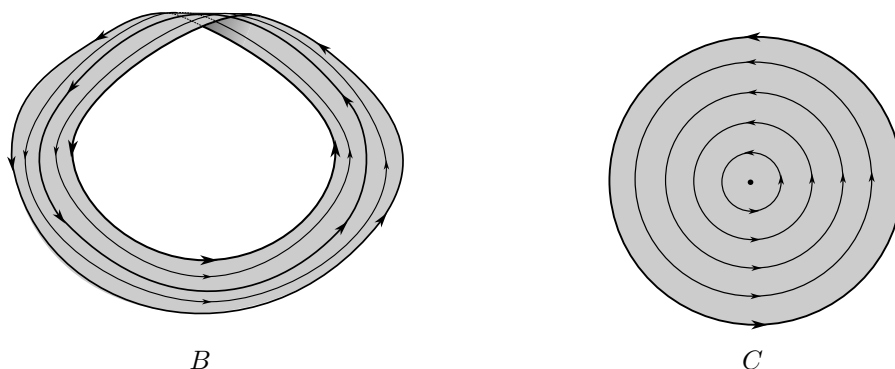


Fig. 4. Atoms B and C

Example 4. If we consider the function $f(x, y, z) = z$ on the sphere $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$, then its Hamiltonian vector field is symmetric with respect to the origin. After identifying diametrically opposite points on the sphere, we can distinguish two atoms:

$$B = \{(x, y, z) \in S^2 : |z| \leq \frac{1}{3}\} / \sim, \quad C = \{(x, y, z) \in S^2 : |z| \geq \frac{2}{3}\} / \sim.$$

As the last example shows, there exists a Hamiltonian flow on the projective plane without saddles. From the Poincaré-Hopf theorem, it follows that on the projective plane, the number of centers is one more than the number of saddles, while on the Klein bottle, the number of centers equals the number of saddles. Since every Morse function on a closed manifold has extremum points, the simple Hamiltonian flow contains at least one center. Therefore, on the Klein bottle, every simple Hamiltonian flow contains a saddle.

3. Reeb Graph

Definition 2. A *fiber* of a simple Hamiltonian flow be one of the following sets: 1) a *regular fiber* is a regular closed trajectory; 2) a *singular fiber* is either a singular closed trajectory or the union of a singular point along with its separatrices.

Definition 3. The factor space $M / \sim$ is called a *Reeb graph*, where $x_1 \sim x_2$ if x_1 and x_2 belong to the same fiber. In this graph, for each vertex of degree 1, the type of the corresponding atom (A_2 , B , or C , meaning that vertices of degree 1 are colored in three colors) is indicated, and for each vertex of degree greater than 1, one of the incident edges is marked. *Marked edge* for a vertex v of the Reeb graph is the edge, which corresponds to two or three hyperbolic sectors of a singular point, that corresponds to v . Reeb graphs are equivalent if there exists an isomorphism between them that preserves the type of vertices of degree 1 and the marked incident edges for vertices of degree greater than 1.

Let $N(A_i)$ ($N(B)$, $N(C)$) denote the number of vertices in the Reeb graph corresponding to A_i (B , C) atoms, V be the total number of vertices, and E be the total number of edges.

Theorem 1. The Reeb graph of a simple Hamiltonian flow on a connected non-orientable surface has the following properties:

- 1) the graph is connected, and the degree of the vertices does not exceed 3;
- 2) the non-orientable genus g of the surface can be computed using the formula:

$$g = N(A_1) + 2N(A_2) + N(B) + 2(E - V + 1). \quad (1)$$

Proof. 1) The Reeb graph is a factor space of a connected manifold, and therefore it is connected. Since the vertices of the Reeb graph correspond to singular fibers, their degrees are the same as those of the corresponding atoms in the Reeb graphs: one for A_2, B, C , two for A_1 , and three for A_0 .

2) If all atoms are removed from the surface, we obtain a disconnected union of cylinders. Therefore, the surface can be viewed as a union of these sets glued together. Each Möbius strip increases the genus by 1. Thus, gluing each of the atoms A_1 or B increases the genus by 1. The atom A_2 is a union of two Möbius strips, so its gluing increases the genus by 2. Hence, if the Reeb graph is a tree, the genus of the surface is determined by the number of atoms A_1, A_2, B . Any graph can be obtained from a tree by connecting pairs of vertices with edges. Adding such an edge to the Reeb graph is equivalent to adding a blank handle to the surface (removing two disks and gluing a cylinder to the boundaries of the holes). This operation increases the oriented genus of the surface by 1 or the non-oriented genus by 2. This accounts for the presence of the last term in the sum. $\square$

Remark 1. For Morse-Smale fields and Lyapunov graphs, analog of Theorems 1 was proven in (de Rezende, & Franzosa, 1993).

Theorem 2. Two simple Hamiltonian flows on non-oriented closed connected surfaces are topologically equivalent if and only if their Reeb graphs are equivalent.

Proof. The proof scheme is analogous to the proof for the oriented case (Bolsinov, & Fomenko, 2004). First, we prove the topological equivalence of flows on the atoms, and then we extend the obtained homeomorphisms to the complements of the atoms (cylinders).

The isomorphism of Reeb graphs defines a bijection between atoms and their boundary components. We fix a Riemannian metric on the manifold, which in the neighborhoods of singular points coincides with the Euclidean metric of the coordinate system in which the Hamiltonian $H(x, y) = \pm x^2 \pm y^2$. Then, the trajectories orthogonal to the flow, along with the flow trajectories, can be used to construct a local coordinate system, and the topological equivalence of flows can be defined through the convergence of coordinates.

Let π be a double covering, H be a Hamiltonian, p be a saddle point, x, y be local coordinates in a neighborhood of the point p , in which $H(x, y) = x^2 - y^2$, the atom A is a component of $\pi(H^{-1}(-\varepsilon, \varepsilon))$, $\varepsilon > 0$, containing $p = (0, 0)$; $U = \{(x, y) : |x^2 - y^2| \leq \varepsilon, |xy| \leq \varepsilon\}$ is a neighborhood of the point p ; $b \subset A$ is the singular fiber; β_i are the components of $|x^2 - y^2| \leq \varepsilon, |xy| = \varepsilon$; α_i are the components of $b \setminus \text{Int}(U)$ (in Fig. 2, U is the region filled with trajectories, β_i is an arc perpendicular to the flow at the boundary of this region, and α_i is an arc of the critical level in the region not filled with trajectories). Note that the curves $xy = \text{const}$ are orthogonal to the trajectories of the Hamiltonian field lying on the curves $x^2 - y^2 = \text{const}$. In U , the homeomorphism is defined by the equality of the corresponding coordinates of points. We then extend this homeomorphism to the atom A . For this, we define homeomorphisms for α_i . Since the components $A \setminus U$ are homeomorphic to $\alpha_i \times \beta_i$, and homeomorphisms are constructed on each of α_i and β_i , they define a homeomorphism for $\alpha_i \times \beta_i$ and for A . Similarly, the homeomorphism from the atoms extends to the cylinders. The consistency of the constructed homeomorphisms on each of the segments follows from the isomorphisms of Reeb graphs. $\square$

4. Simple Hamiltonian Flows on the Projective Plane

To obtain the number of topologically non-equivalent Hamiltonian flows, we calculate the number of Reeb graphs with a given number of vertices that correspond to saddle critical points.

We begin calculations for graphs that are trees, having vertices of degree one (type C) and three (type A_0), with one of the degree-one vertices serving as the root. Such graphs are Reeb graphs of Morse functions (and their Hamiltonian flows) on the 2-disk (the root is correspond to the boundary fiber). We orient all edges according to the following rules: 1) if the edge is incident to the root, then it is directed away from it, 2) if the marked edge for a saddle vertex is entering it, then two other edges are outgoing from it, 3) if the marked edge is exiting from a saddle, then two other edges are incoming to the saddle.

Since a vertex of degree 2 is not a source or sink on the graph, the orientation of one of the edges incident to it determines the orientation of the second edge. Similarly, at a vertex of degree 3, the condition that the marked edge has the opposite orientation to the other two edges allows the orientation of one edge to determine the orientation of the other two incident edges. Since the tree is a connected graph, the orientation of one edge sequentially determines the orientations of the edges adjacent to it, then the orientations of the edges adjacent to the adjacent ones, and so on, until the orientation of all edges of the graph is determined.

Lemma 1. The number K_k of Reeb graphs that are rooted trees with k saddle points (vertices of type A_0) and other vertices of type C can be computed using the following recursive formulas: $K_0 = 1$,

$$K_{2n} = 3 \sum_{i=0}^{n-1} K_i K_{2n-1-i} \quad \text{for even } k = 2n; \quad (2)$$

$$K_{2n+1} = \frac{3K_n^2 + K_n}{2} + 3 \sum_{i=0}^{n-1} K_i K_{2n-i} \quad \text{for odd } k = 2n + 1. \quad (3)$$

Proof. If $k=0$, then the tree has no saddles. Such a graph is unique. It has two vertices and one edge connecting them. From now on, we assume that $k > 0$.

Consider the saddle v closest to the root of a rooted tree G with k saddle points (vertices of type A_0) and other vertices of type C . It is a vertex of degree 3. We cut the graph by v . We obtain three connected components. One of them is the edge e connecting v to the root. The other two components (A and B) can be viewed as rooted Reeb graphs, where the root is v . Let us assume that the number of vertices $|V_A|$ is no fewer than the number of vertices $|V_B|$ of B ($|V_A| \geq |V_B|$).

Then there are three possibilities:

- 1) the marked edge to a vertex v is e ;
- 2) the marked edge belongs to A ;
- 3) the marked edge belongs to B .

Multiplying by 3 in formulas (2) and (3) is to indicate which subgraph with the marked edge is connected to.

If $k = 2n$, then the number of graph vertices is different ($|V_A| > |V_B|$), and A cannot be isomorphic to B . The number of non-isomorphic trees with i saddles is equal to K_i , and with $2n - 1 - i$ saddles is equal to K_{2n-1-i} , so the total number of trees G of this type is equal to $K_i K_{2n-1-i}$. By considering all possible values of i , we obtain formula (2).

If $k = 2n + 1$ and $|V_A| > |V_B|$, then we have similar products as in the case of $k = 2n$. Let's consider the case of $|V_A| = |V_B| = n$ in more detail. If edge e is not marked, then we relabel the graphs so that the marked edge belongs to graph A . Since each of the graphs can be chosen in K_n ways, there are K_n^2 graphs G of this type. If edge e is marked, then two cases are possible: 1) graph A is not isomorphic to graph B , 2) graphs A and B are isomorphic. In the first case, we have $K_n(K_n - 1)/2$ non-isomorphic graphs, and in the second case, K_n graphs. Summing everything together, we obtain formula (3). $\square$

The results of the calculations of the number of Reeb rooted trees on $\mathbb{R}P^2$ are presented in Tab. 1.

From theorem 1 it follows:

Corollary 1. *The Reeb graph on the projective plane has the following properties:*

- 1) the graph is a tree;
- 2) the graph has no vertices of type A_2 and only one vertex of type A_1 or B .

Proof. In formula (1), all four summands are non-negative, therefore, since the genus of the projective plane is $g = 1$, it follows that $E - V + 1 = 0$ (which is a criterion for a connected graph to be a tree), $N(A_2) = 0$, and one of the following holds: 1) $N(A_1) = 1$, $N(B) = 0$ or 2) $N(A_1) = 0$, $N(B) = 1$. $\square$

Theorem 3. *Let a graph G be given, which is a connected tree with vertices of degree 1 colored in two colors (B , C), and has one of two types:*

- 1) it has a single vertex of degree 2 (type A_1), while the remaining vertices have degrees 1 and 3, all vertices of degree 1 are of type C , and for the vertex of degree 2 and the vertices of degree 3, one incident edge is marked to each;
- 2) all vertices have degrees 1 or 3, one of the vertices of degree 1 is of type B , while the remaining vertices of degree 1 are of type C , and for the vertex of degree 3, one incident edge is marked to it.

Then there exists a simple Hamiltonian flow on the projective plane, whose Reeb graph is equivalent to G .

Proof. In the first case, we construct the atom A_1 and the flow on it as in Example 3. The vertex of degree 2 splits the graph into two rooted trees – the Reeb graphs of simple Morse functions on a two-dimensional disks, constant on the boundary. We consider the Hamiltonian flows generated by these functions. By attaching disks to the boundary of A_1 , we obtain an undirected surface with the desired flow on it.

In the second case, we start with the atom B in Example 4, and then, as in the first case, we construct the flow on the two-dimensional disk and attach this disk to the boundary of B . We smooth the obtained flows as it is done in (Gutierrez, 1986; Neumann, 1978). Since the field is smoothed at the gluing points and, according to these works, the topological structure of the flow is preserved during smoothing, the smoothing process actually consists of smoothing the Hamiltonian function. Consider a neighborhood of the edge as a collar $S^1 \times [0, 1)$ with a Hamiltonian on it $H_1(s, t) = H_1(t)$ for the first of the glued surfaces and $S^1 \times (-1, 0]$ with a Hamiltonian $H_2(s, t) = H_2(t)$. Here, $S^1 \times \{0\}$ is a common boundary for both surfaces. Let's define the Hamiltonian function on the glued surface as $H(s, t) = H_1(s, t)$ for the points of the first surface. And for the points of the second surface, $H(s, t) = H_2(s, t) + H_1(0) - H_2(0)$ if $H'_1(0)H'_2(0) > 0$ and $(s, t) = -H_2(s, t) + H_1(0) + H_2(0)$ if $H'_1(0)H'_2(0) < 0$. On the set $S^1 \times (-1, 1)$ we have $H(s, t) = H(t)$ and this function is monotonic and smooth except at point $t = 0$. Smoothing it at 0, we obtain the desired Hamiltonian function. $\square$

Remark. Theorem 3 also holds in the other direction: if G is the Reeb graph of a Hamiltonian flow on the projective plane, then it is of type 1) or 2). This follows directly from Theorem 1.

Theorem 4. *The number of topologically non-equivalent simple Hamiltonian flows with k saddles on the projective plane $\mathbb{R}P^2$ can be calculated using the formula*

$$N(\mathbb{R}P^2)_k = K_k + \sum_{i=0}^{k-1} K_i K_{k-i-1}.$$

Proof. Since the projective plane is a non-orientable surface of genus 1, the flow must contain either 1) one singular closed trajectory, or 2) one atom A_1 . In the case of a singular closed trajectory, it corresponds to a vertex B on the Reeb graph, which we consider as the root. Therefore, the number of such flows is equal to K_k .

In the second case, the complement to the atom A_1 is a disconnected union of two two-dimensional disks. On each of them, we have a Hamiltonian flow. The complete topological invariant of such flows is the rooted Reeb graphs (the root corresponds to the boundary of the 2-disk). We order these two 2-disks: the first disk lies in one sector of the saddle point of the atom A_1 , while the second disk lies in three sector of this point. Since the flow on the second disk does not depend on the flow on the first disk, and the total number of saddle points on these two 2-disks is equal to $k - 1$, the total number of flows satisfies the formula from the theorem. $\square$

The results of the calculations of the number of simple Hamiltonian flows on $\mathbb{R}P^2$ are presented in Tab. 1.

Table 1

Numbers of rooted trees and simple Hamiltonian flows on $\mathbb{R}P^2$									
k	0	1	2	3	4	5	6	7	8
K_k	1	2	6	25	111	540	2736	14396	77649
$N(\mathbb{R}P^2)_k$	1	3	10	41	185	898	4560	23985	129415

5. Reeb Graphs with Two Fixed Leaves

For a tree with two fixed leaves and with one and three degree vertices, we orient all edges according to the following rules: 1) if the edge is incident to the first leaf, then it is directed away from it, 2) if the marked edge for a saddle vertex is entering it, then two other edges are outgoing from it, 3) if the marked edge is exiting from a saddle, then two other edges are incoming to the saddle. We say that the isomorphism of the graph preserves the order of leaves if it maps the first fixed leaf to first fixed leaf and second fixed leaf to the second one.

Lemma 2. *Up to isomorphism that preserves the order of leaves, the number L_k of Reeb graphs that are trees with two ordered fixed leaves and k saddle points (saddles have degree three, and all other vertices have degree one) can be computed using the following recursive formula: $L_0 = 1$,*

$$L_k = 3(L_0 K_{k-1} + L_1 K_{k-2} + \dots + L_{k-1} K_0).$$

Proof. If $k = 0$, there is one graph with one edge: $L_0 = 1$. Suppose $k > 0$. Just as in the previous theorem, consider the saddle point v closest to the first fixed leaf. It is a vertex of degree 3, since $k > 0$. We cut the graph at v . This results in three connected components. One of them is the edge connecting v to the first leaf. The other two components can be viewed as the rooted Reeb graph G_1 and a graph G_2 with two ordered fixed leaves, where the root (the first leaf) is v .

Again, we have three options for selecting the marked edge in v , which explains the presence of the multiplier 3 in the formula:

- 1) in both graphs, the edges incident to v are outgoing (the marked edge does not belong to either G_1 or G_2);
- 2) this edge is outgoing in G_1 , while it is incoming in G_2 (the marked edge belongs to G_1);
- 3) it is incoming in G_2 , while it is outgoing in G_1 (the marked edge belongs to G_2).

The first and second graphs are not isomorphic, as the first has two fixed leaves, while the second has only one (the root). Therefore, the number of non-isomorphic original graphs with such subgraphs equal the number of non-isomorphic subgraphs multiplied together. This gives the desired formula. $\square$

Example. In the case $k = 1$ the graph has the shape of the letter Y. At the vertex of degree 3 there are three options for choosing the incident edge corresponding to the vertices of degree 1: 1) the first fixed vertex, 2) the second fixed vertex, 3) a non-fixed vertex. Therefore $L_1 = 3$.

Furthermore, we need the concept of a monotonic graph. A *monotonic graph* is a Reeb graph of a simple Morse function with two fixed leaves v_1, v_2 and each saddle vertex of the simple path from v_1 to v_2 has the marked edges belonging to the path. This means that an orientation can be assigned to the graph that is consistent with the fixed edges, and the path from v_1 to v_2 is directed (the restriction of the Hamiltonian to them is monotonic). Let M_k^S denote the number of non-isomorphic monotonic graphs with k saddles (up to isomorphism that can preserve or change the order of the fixed pair of vertices) and let M_k denote the number of non-isomorphic monotonic graphs with k saddles up to isomorphism, which preserves the order of the fixed pair of vertices (the orientation of the path from the first to the second vertex).

Let's find a recursive formula for M_k similar to Lemma 2. There exists a unique graph without saddles – a segment. Therefore, $M_0 = M_0^S = 1$. Let $k > 0$ and $v_1 v$ be an edge incident to v_1 . Then v is a vertex of degree 3 belonging to the path from v_1 to v_2 . It divides the graph into three parts: 1) the edge $v_1 v$, 2) a monotone graph with two fixed leaves v and v_2 , 3) a tree with root v . The monotone graph and the rooted tree have a total of $k - 1$ saddle vertices. Therefore, the following formula holds:

$$M_k = 2(M_0 K_{k-1} + M_1 K_{k-2} + \dots + M_{k-1} K_0).$$

Let's find the number of monotone graphs with k saddles for which there exists a graph isomorphism onto itself that swaps fixed leaves and preserves the distinguished edges. This isomorphism has an invariant vertex or edge. If it is a vertex, then it has degree 3 and one of the edges is distinguished, and therefore its image under the isomorphism is also a distinguished edge, which is impossible, because each saddle vertex has only one distinguished edge. Therefore, the

invariant is an edge. It divides the graph into two isomorphic subgraphs, which are monotone. Therefore, the number of such graphs is equal to $M_{k/2}$. Taking this into account, we have $M_k^S = \frac{1}{2}M_k$ for odd k , and $M_k^S = \frac{M_k + M_{k/2}}{2}$ for even k . The results of calculations using these formulas are shown in Tab. 2.

Example. If $k=1$, then the monotone graph with two fixed leaves has the form of the letter Y. Then, the marked edge for vertex 3 can be incident to either the first leaf or the second. If the isomorphism preserves the order of the leaves, then these two options are different, and if it can change the order of the leaves (the orientation of the path from one leaf to the other), then they are isomorphic. Therefore, $M_1 = 2$ and $M_1^S = 1$.

Table 2

Numbers of trees with two fixed leaves and monotonic graphs

k	0	1	2	3	4	5	6	7
L_k	1	3	15	81	462	2700	16074	96876
M_k	1	2	8	36	178	918	4904	26824
M_k^S	1	1	5	18	93	459	2470	13412

6. Simple Hamiltonian Flows on the Klein Bottle

A vertex of a graph is called an *R-vertex* for a cycle in the graph if the marked edge for this vertex does not belong to the cycle.

Since the non-orientable genus of the Klein bottle is 2, it follows from Theorem 1 that:

Corollary 2. *The Reeb graph on the Klein bottle has one of the following types:*

- the graph is a tree containing two vertices of degree 2, and the vertices of degree 1 are of type C;*
- the graph is a tree containing one vertex of degree 2, one vertex of type B, and the remaining vertices of degree 1 are of type C;*
- the graph is a tree that does not contain vertices of degree 2, has two vertices of type B, and the remaining vertices of degree 1 are of type C;*
- the graph is a tree that does not contain vertices of degree 2, has one vertex of type A_2 , and the remaining vertices of degree 1 are of type C;*
- the graph has one simple cycle, this cycle contains an odd number of R-vertices, and all vertices of the graph are either of type C or of degree three.*

Proof. As noted earlier, the condition $E - V + 1 = 0$ for a connected graph is equivalent to it being a tree. Therefore, we have the first four options. In the fifth option, $E = V$, which is equivalent to the existence of a cycle in the graph. In this case, the first three summands in the formula (1) are zero. If the cycle is cut at an interior point of an edge, we obtain a Reeb graph that is a tree and defines a Hamiltonian flow on the cylinder. We fix one of the boundary components of the cylinder. Projecting one base of the cylinder onto the other allows us to compare the orientations on them. If the orientations are the same, then after gluing the bases, we obtain a torus; if they are different, we obtain the Klein bottle. Moving a point along the cycle on the Reeb graph, we compare the orientations of the corresponding closed trajectories with the orientation of the initial one. Passing through R-vertices changes the orientation to the opposite, while other vertices do not change it. Therefore, to obtain a Klein bottle after gluing, there must be an odd number of R-vertices on the cycle. $\square$

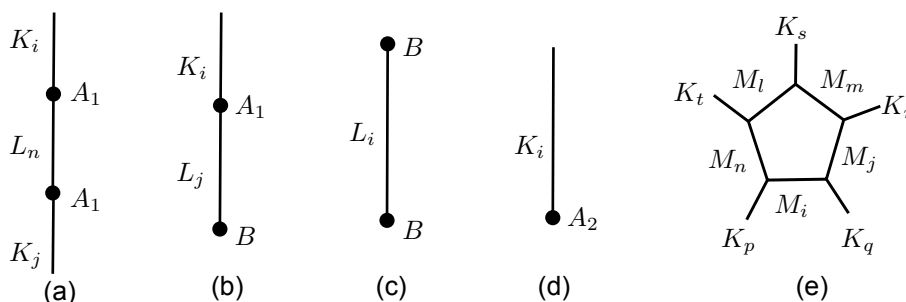


Fig. 5. Five types of Reeb graphs on Kl (each edge denotes a tree without A_1, A_2, B -vertices)

Theorem 5. *Let a graph G be given, which has one of the five types from the previous theorem. Then there exists a simple Hamiltonian flow on the Klein bottle, the Reeb graph of which is equivalent to the graph G .*

Proof. For each vertex of the graph, we construct an atom corresponding to the neighborhood of this vertex and a flow on it as was done in the examples of atoms. Each edge of the Reeb graph corresponds to a cylinder $S^1 \times [0, 1]$, on which a Hamiltonian flow is defined, generated by the projection onto the second factor. We attach the cylinders to the corresponding boundary components of the atoms, ensuring that the directions of movement along the trajectories on the boundary component of the cylinder and the boundary component of the atom coincide. We smooth the obtained flow as

it is done in (Gutierrez, 1986; Neumann, 1978). The Hamiltonian is smoothed like proof of Theorem 3. As a result, we obtain a closed surface that is non-orientable for the first four types of graphs due to the presence of a non-orientable atom. In the last type, the non-orientability of the surface is justified in the proof of the corollary. According to formula (1), the genus of the surface is equal to 2. Thus, we have a flow on the Klein bottle. On the non-orientable atoms, this flow lifts to a Hamiltonian flow on the double cover as in the examples, while on the orientable atoms and disks, it is duplicated twice for each of the lift copies. Therefore, the constructed flow is Hamiltonian with the given Reeb graph. $\square$

Theorem 6. *The number $N(Kl)_k$ of topologically non-equivalent simple Hamiltonian flows on the Klein bottle can be computed using the following formula for even $k = 2m$ or odd $k = 2m + 1$ numbers of saddles:*

$$N(Kl)_k = N(Kl)_k^a + N(Kl)_k^b + N(Kl)_k^c + N(Kl)_k^d + N(Kl)_k^e,$$

where the number $N(Kl)_k^a$ of Reeb graphs containing two A_1 -vertices (degree 2) is computed using the formula

$$N(Kl)_k^a = 2 \sum_{i+j+n=k-2} K_i K_j L_n + \sum_{2(i+j+1)=k} K_i L_j + \sum_{2i+2j+n=k-3} K_i L_j K_n;$$

the number $N(Kl)_k^b$ of Reeb graphs having one A_1 -vertex and one B -vertex is computed using the formula

$$N(Kl)_k^b = 2 \sum_{i+j=k-1} K_i L_j;$$

the number $N(Kl)_k^c$ of Reeb graphs having two vertices of type B is computed using the formula

$$N(Kl)_k^c = \frac{1}{2}(L_k + L_m + \sum_{2i+j+1=k} L_i K_j), \quad k = 2m,$$

$$N(Kl)_k^c = \frac{1}{2}(L_k + \sum_{2i+j+1=k} L_i K_j), \quad k = 2m + 1;$$

the number $N(Kl)_k^d$ of Reeb graphs that have one vertex of type A_2 is calculated using the formula

$$N(Kl)_k^d = K_{k-1};$$

and $N(Kl)_k^e$ is the number of Reeb graphs that contain a cycle.

Here, in all sums, $i, j, n \geq 0$ and the sums are taken over these indices.

Note that the cases (a)–(e) of the flows from the theorem correspond to Fig. 5 (a)–(e).

Proof. Case a. (Fig. 5 (a)). Two atoms A_1 divide the surface into three parts: two 2-disks and one cylinder. If the cylinder lies in one hyperbolic sector of one of the atoms A_1 (which we call the first), and in three hyperbolic sectors of the second atom A_1 , then we have a natural ordering of the atoms. Therefore, the number of possible flows in this case is equal to

$$S_{KKL}^{k-2} = \sum_{i+j+n=k-2} K_i K_j L_n.$$

The cases where the cylinder lies in one hyperbolic sector of each atom and where it lies in three hyperbolic sectors of each of the A_1 are identical, so we only consider the first case and multiply the resulting number of flows by 2. We fix the order of the A_1 atoms. The number N of Reeb graphs up to isomorphism that preserve an unordered pair of vertices is equal to the sum of the number N_1 of such graphs for which there is no isomorphism that swaps the vertices, and the number N_2 of such graphs for which the specified isomorphism exists. If we fix the order of the pair of vertices and consider graph isomorphisms that preserve this order, we obtain $K = 2N_1 + N_2$ graphs. Thus, $N = (K + N_2)/2$. Considering the doubling of the total number, we have that the number of non-isomorphic graphs equals the sum of the number of graphs with an ordered pair of fixed vertices (A_1) up to an isomorphism that preserve order of the vertexes and the number of self-symmetric graphs up to an isomorphism. As before, up to an isomorphism that preserves the given order of A_1 -vertices, the total number of Reeb graphs is equal to S_{KKL}^{k-2} . Changing the order of the A_1 atoms leads to a rotation of the orientation on the cylinder and the symmetry of the Reeb graph.

With an odd number of saddles, the symmetric graph can have one fixed edge (Fig. 6 (a1)). This edge divides the graph into two identical parts, each consisting of one root tree A and one tree with two fixed leaves B . Their total possible number is

$$\sum_{2(i+j+1)=k} K_i L_j.$$

If instead of the edge, C is invariant under symmetry subgraph (Fig. 6 (a2)), then there is a vertex v , which divide the Reeb graph onto 3 part, one of which is C . Other parts are isomorphic. Each of them consists of a root tree A and tree with two fixed leaves B . The number of Reeb graphs of such a type is equal $\sum_{2i+2j+n=k-3} K_i L_j K_n$. We add this to the total number of flows. Thus, we obtain the first formula.

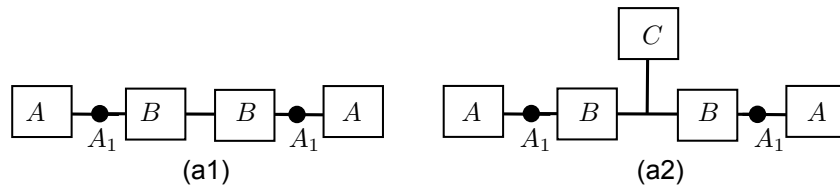


Fig. 6. Symmetric graphs with two A_1 vertices ($V(A) = i$, $V(B) = j$, $V(C) = n$)

Case b. The Reeb graph in this case does not allow symmetries that change two fixed points: the vertices A_1 and B . The vertex A_1 divides the Reeb graph into two parts: a root tree and a tree with two fixed leaves. The tree with two fixed leaves has a canonical orientation, generated at vertex B by the movement from vertex B to A_1 . Therefore, the total number of Reeb graphs is the sum of the products of K_i and L_j . The total number of saddles in these trees is $k - 1$. Since at vertex A_1 there is the possibility to marking each of the two incident edges, we double the total sum.

Case c. We consider two vertices B as marked leaves. The symmetry of the graph that swaps them can occur in two cases: 1) when there is no fixed vertex under the symmetry (possible only with an even number of saddles), 2) when there is a fixed tree. In the first case, graphs with such symmetries are determined by their halves, so their number equals L_m . In the second case, the complement to the tree consists of two isomorphic graphs with two leaves. Their common vertex is also a saddle (and is an R-vertex). Therefore, the total number of such graphs equals

$$\sum_{2i+j+1=k} L_i K_j.$$

All together gives the desired formulas.

Case d. We consider vertex A_2 as the root, thus we have a rooted tree with $k - 1$ saddles, and the number of Reeb graphs equals K_{k-1} . $\square$

7. Reeb graphs with a cycle

For Reeb graphs that have a cycle, the number of vertices in the cycle, with the marked edge not belonging to the cycle, is odd. If the cycle is a loop (Fig.17.1), then the vertex of the loop is the root of the complement to the loop. In the Reeb graph, the loop corresponds to a monotonic graph, and the complement to the loop are rooted trees. Since the graph can have symmetry that change orientation of the loop, the number of monotonic trees is calculated as M_i^S . Thus, the number of such graphs is

$$\sum_{i=0}^{k-1} M_i^S K_{k-1-i}.$$

In other cases, when calculating the number of graphs, one should consider the possible symmetries of the graph relative to the cycle (Fig. 5e). An algorithm for finding all cyclic Reeb graphs with n saddles can be constructed according to the following scheme:

1) We find this number separately for each odd m ($m \leq n$) of R-vertices on the cycle. We represent the cycle as a regular m -gon. The symmetries of the graph correspond to the symmetries of the m -gon. They are divided into two types: a) axial symmetries – the axis passes through a vertex and the center of the opposite side (Fig. 7.11 and 14), b) rotations – cyclic permutations of the vertices (Fig. 7.10 and 15).

2) R-vertices (vertices of the m -gon) partition the Reeb graph into rooted trees and monotonic graphs (Fig. 5e). Thus, each vertex of the m -gon corresponds to a rooted tree, and each edge corresponds to a monotonic graph. The total number of saddle vertices in the rooted trees and monotonic graphs equals $n - m$. We count based on the maximum number k of saddle vertices in each rooted tree and monotonic graph. We first consider the case when one vertex of the m -gon has k saddles. Then only the symmetry relative to the axis passing through this vertex is possible. If $n - m - k$ is odd, then there are no self-symmetric graphs, so the total number is half the sum of the products K_i and M_j , corresponding to the sides and vertices of the m -gon:

$$S(n - m - k) = \frac{1}{2} \sum_{i_2 + \dots + i_{m-1} + j_1 + \dots + j_{m-1} = n - m - k} K_{i_1} \dots K_{i_m} M_{j_1} \dots M_{j_m}.$$

Here $i_1 = k$. If $n - m - k$ is even, then there are graphs with an axis of symmetry, which are determined by their half, that is, a graph with $\frac{n-m-k}{2}$ vertices. In this case, the number of non-isomorphic Reeb graphs equals

$$S(n - m - k) + S\left(\frac{n - m - k}{2}\right).$$

If we have two vertices with the same number k , then there is a unique axial symmetry that maps one vertex to the other. Therefore, the calculations are analogous to the case of one vertex with k . In the case of three vertices with the same k and m being a multiple of three, in addition to axial symmetries, cyclic permutations are also possible. In this case, the graphs that are invariant under such permutations are defined by graphs with $\frac{n-m-k}{3}$ saddles. Here we assume that each $0 \leq j_i \leq k$.

Separately, we consider the case when the maximum number of vertices on an edge (in a monotone graph) equals k , while on all root trees it is less than k . If there is only one such edge (monotone tree), then the number of Reeb graphs of this type equals $M_k S(n-m-k)$ for odd k , and taking into account the symmetries relative to this edge – $(M_k + M_{k/2})S(n-m-k)$ for even k . If we have two or more edges with a maximum number of k , then the possible symmetries are accounted for similarly to those in the case of vertices.

8. Examples

Let us illustrate how the number $N(Kl)_6^C$ is obtained using the example of graphs with 6 saddles.

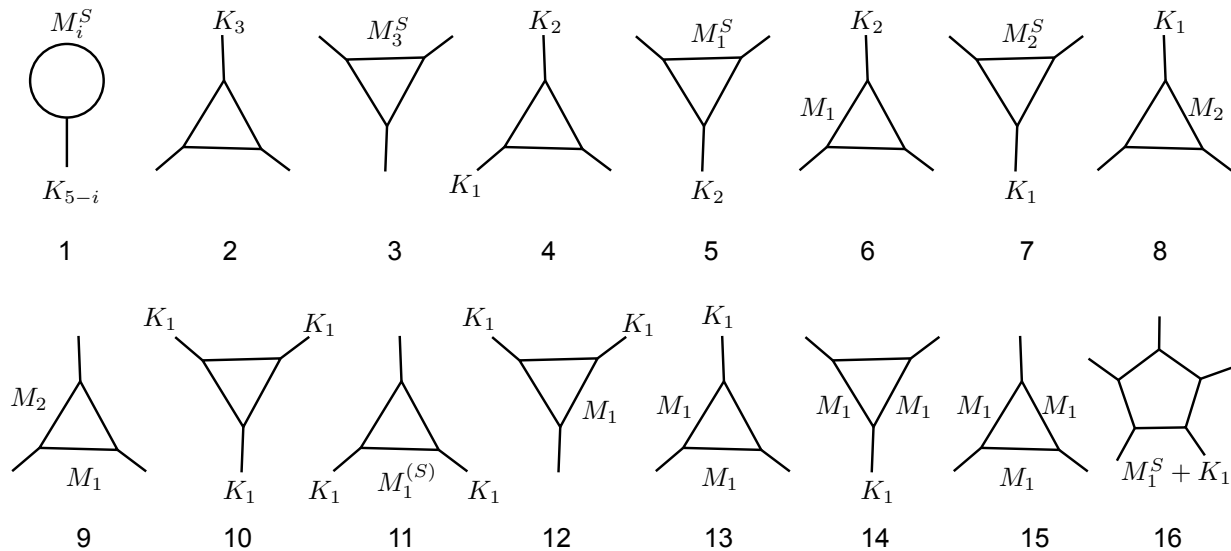


Fig. 7. Cyclic Reeb graphs with 6 saddles (the marked edges represent trees)

All possible types of cyclic graphs with no more than 6 saddles are shown in Fig. 7. Since R-vertices correspond to saddles, the following cases are possible: 1) one R-vertex, 2–15) three R-vertices, 16) five R-vertices. In the first case, the R-vertex divides the Reeb graph into a monotonic graph (loop) and a tree. Since the graph may have symmetry, we have M_i^S possible graphs for the loop, taking symmetry into account. For the tree, there are K_{5-i} graphs. In total, in the first variant, we obtain $\sum_{i=0}^5 M_i^S K_{5-i} = 1529$ cyclic graphs. In the case of three R-vertices, graphs 4, 6, 8, 9, 12, and 13 have no symmetries. Therefore, the number of cyclic graphs is determined as the product of the corresponding numbers. Additionally, symmetries are considered for graphs 2, 3, 5, 7, and 11. In the case of 14, with a fixed K_1 , we have two options when the graphs are symmetric to themselves, and two when they are symmetric to each other. Thus, taking K_1 into account, there are a total of 6 cyclic graphs of this type. Graph 10 can have both axial symmetries and symmetries with cyclic permutations of vertices (rotation by $2\pi/3$). To find the number of options, we denote the graphs with K_1 as G_1 and G_2 . In a cyclic graph, there can be from 0 to 3 graphs G_1 . Therefore, there are 4 cyclic graphs of type 10. For type 15, symmetries and cyclic permutations are also possible. We fix the direction of traversal of the cycle. In relation to it, there are two possible situations: 1) all monotonic graphs are the same, 2) two are the same, and the third is of a different type. Thus, there are only two cyclic graphs of type 15. In summary, we find 147 cyclic graphs with three R-vertices. In the case of type 16, we have 5 R-vertices. If the sixth (non-R) vertex lies on the monotonic graph, then there is exactly one variant up to symmetry. If it belongs to the tree, then there are 2 variants. In total, there are three cyclic graphs with 5 R-vertices. Overall, we have 1679 cyclic graphs with 6 saddles.

Results of Calculations

Using the obtained formulas and the described algorithm, the number of Reeb graphs of each type was found, as well as the total number of flow structures with no more than 8 saddles. The results of the calculations are presented in the Tab. 3.

Table 3

Numbers of simple Hamiltonian flows on the Klein bottle								
k	1	2	3	4	5	6	7	8
a	0	3	16	89	530	3214	19575	120407
b	2	10	54	308	1800	13016	64584	392938
c	2	9	45	254	1422	8400	50058	303211
d	1	2	6	25	111	540	2736	14396
e	1	3	14	62	316	1679	8987	49348
$N(Kl)_k$	6	27	135	738	4179	26849	145940	880300

9. Discussion and conclusions

A topological classification of simple Hamiltonian flows on the projective plane and the Klein bottle has been obtained. All possible topological structures of flows with no more than 8 saddles have been described. A recursive formula has

been derived for determining the number of topologically non-equivalent flows with a given number of critical points on the projective plane. Flows on the Klein bottle have been divided into five types; for the first four types, recursive formulas have been obtained to find the number of their possible structures, while for the last type, an algorithm for finding such structures has been described.

The results obtained may be useful for further research on the topological properties of Hamiltonian flows on various surfaces.

Authors' contribution: Alexandr Prishlyak — conceptualization; methodology; empirical research; Kateryna Butok — software; empirical data collection and validation; analysis of sources, preparation of literature review and theoretical foundations of research.

Acknowledgments. The authors would like to thank the anonymous referees who provided useful and detailed comments on a previous/earlier version of the manuscript.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Bolsinov, A., & Fomenko, A. (2004). *Integrable hamiltonian systems. geometry, topology, classification*. CRC Press. <https://doi.org/10.1201/9780203643426>
- Cobo, M., Gutierrez, C., & Llibre, J. (2010). Flows without wandering points on compact connected surfaces. *Transactions of the American Mathematical Society*, 362(9), 4569–4580. <https://doi.org/10.1090/S0002-9947-10-05113-5>
- de Rezende, K., & Franzosa, R. (1993). Lyapunov graphs and flows on surfaces. *Transactions of the American Mathematical Society*, 340(2), 767–784. <https://doi.org/10.1090/S0002-9947-1993-1127155-9>
- Gutierrez, C. (1986). Smoothing continuous flows on two-manifolds and recurrences. *Ergodic Theory and Dynamical Systems*, 6(1), 17–44. <https://doi.org/10.1017/S0143385700003278>
- Hladysh, B., & Prishlyak, A. (2019). Simple Morse functions on an oriented surface with boundary. *Journal of Mathematical Physics, Analysis, Geometry*, 15(3), 354–368. <https://doi.org/10.15407/mag15.03.354>
- Khohliyk, O., & Maksymenko, S. (2020). Diffeomorphisms preserving Morse–Bott functions. *Indagationes Mathematicae*, 31(2), 185–203. <https://doi.org/10.1016/j.indag.2019.12.004>
- Kravchenko, A., & Maksymenko, S. (2019). Automorphisms of Kronrod-Reeb graphs of Morse functions on compact surfaces. *European Journal of Mathematics*, 6(1), 114–131. <https://doi.org/10.1007/s40879-019-00379-8>
- Maksymenko, S. (2006). Stabilizers and orbits of smooth functions. *Bulletin des Sciences Mathématiques*, 130(4), 279–311. <https://doi.org/10.1016/j.bulsci.2005.11.001>
- Maksymenko, S. (2009). ∞ -jets of diffeomorphisms preserving orbits of vector fields. *Open Mathematics*, 7(2), 272–298. <https://doi.org/10.2478/s11533-009-0010-y>
- Maksymenko, S. (2020). Deformations of functions on surfaces by isotopic to the identity diffeomorphisms. *Topology and its Applications*, 282, 107312. <https://doi.org/10.1016/j.topol.2020.107312>
- Neumann, D. (1978). Smoothing continuous flows on 2-manifolds. *Journal of Differential Equations*, 28(3), 327–344. [https://doi.org/10.1016/0022-0396\(78\)90131-6](https://doi.org/10.1016/0022-0396(78)90131-6)
- Neumann, D., & O'Brien, T. (1976). Global structure of continuous flows on 2-manifolds. *Journal of Differential Equations*, 22(1), 89–110. [https://doi.org/10.1016/0022-0396\(76\)90006-1](https://doi.org/10.1016/0022-0396(76)90006-1)
- Nikolaev, I., & Zhuzhoma, E. (1999). *Flows on 2-dimensional manifolds*. Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0093599>
- Oshemkov, A., & Sharko, V. (1998). Classification of morse-smale flows on two-dimensional manifolds. *Matem. Sbornik*, 189(8), 93–140.
- Prishlyak, A. (2000). Conjugacy of Morse functions on surfaces with values on a straight line and circle. *Ukrainian Mathematical Journal*, 52(10), 1623–1627. <https://doi.org/10.1023/A:1010461319703>
- Prishlyak, A., & Loseva, M. (2020). Topology of optimal flows with collective dynamics on closed orientable surfaces. *Proceedings of the International Geometry Center*, 13(2), 50–67. <https://doi.org/10.15673/tmge.v13i2.1731>
- Reeb, G. (1946). Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Comptes Rendus de l'Académie des Sciences*, 222, 847–849.
- Sharko, V. (2003). Smooth and topological equivalence of functions on surfaces. *Ukrainian Mathematical Journal*, 55(5), 832–846. <https://doi.org/10.1023/B:UKMA.0000010259.21815.d7>
- Wang, X. (1990). The C^* -algebras of Morse–Smale flows on two-manifolds. *Ergodic Theory and Dynamical Systems*, 10(3), 565–597. <https://doi.org/10.1017/S0143385700005757>
- Yokoyama, T. (2022). Refinements of topological invariants of flows. *Discrete and Continuous Dynamical Systems*, 42(5), 2295. <https://doi.org/10.3934/dcds.2021191>

Отримано редакцією журналу / Received: 05.01.25
Прорецензовано / Revised: 23.09.25
Схвалено до друку / Accepted: 10.10.25

Олександр ПРИШЛЯК, д-р фіз.-мат. наук, проф.

ORCID ID: 0000-0002-7164-807X

e-mail: prishlyak@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Катерина БУТОК, д-р філософії, наук. співроб.

ORCID ID: 0009-0003-3000-4416

e-mail: katro.butok@gmail.com

Університет Отто фон Герике, Магдебург, Німеччина

ТОПОЛОГІЧНА СТРУКТУРА ПРОСТИХ ГАМІЛЬТОНОВИХ ПОТОКІВ НА ПРОЄКТИВНІЙ ПЛОЩИНІ ТА ПЛЯЩІ КЛЕЙНА

Для кожної неорієнтованої поверхні існує орієнтоване дволисте накриття. Ми розглядаємо потоки, підняття яких на дволисте накриття є гамільтоновими потоками з гамільтоніаном – функцією Морса. За аналогією з орієнтованими многовидами, потік простий, якщо в ньому не існує сідлових зв'язок між різними сідлами. Для опису топологічних властивостей функцій на двовимірних многовидах використовуємо інваріант, що називається графом Ріба. У випадку простих функцій Морса на замкнутих орієнтованих двовимірних многовидах – це повний топологічний інваріант для пошарової еквівалентності і топологічної еквівалентності гамільтонових потоків.

Проте в нашому випадку, на неорієнтовному многовиді не існує функції, що породжує гамільтоновий потік. З огляду на це поняття графа Ріба потребує перевизначення.

Для дослідження топологічних властивостей простих гамільтонових потоків на неорієнтовних поверхнях описано структуру всіх можливих атомів, що є аналогом атомів Болсінова-Фоменка для функцій Морса та гамільтонових потоків на орієнтованих поверхнях. Зокрема, на неорієнтованих поверхнях є п'ять типів таких атомів. Побудовано повний топологічний інваріант, що є графом Ріба з маркуванням його вершин та інцидентних до вершин піребер, описано його властивості та доведено критерій топологічної еквівалентності потоків. Для проєктивної площини та пляшки Клейна ми доводимо теореми про існування потоків із заданим графом Ріба. Для циклічних графів Ріба на пляшці Клейна описано алгоритм для їхнього перерахування. Для всіх інших типів графів Ріба на проєктивній площині та пляшці Клейна отримано рекурсивні формули для знаходження їхнього числа. Наведено результати обчислень числа топологічних структур потоків з не більш ніж восьми сідлами. Отримані результати відкривають можливість дослідження структур гамільтонових потоків більшої складності на компактних неорієнтованих поверхнях.

Ключові слова: топологічна еквівалентність, граф Ріба, класифікація, неорієнтовна поверхня.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 519.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.9>

Mykyta SADOK, PhD Student
ORCID ID: 0009-0000-0215-4204
e-mail: sadokmykyta@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

LIMIT THEOREMS FOR GENERALISED CONVEX HULLS

The $(K, \mathbb{H})$ -hull of a set $A \subseteq \mathbb{R}^d$ is a generalization of the standard closed convex hull of A which can be defined as the intersection of all images that contain A of a closed convex subset $K \subset \mathbb{R}^d$ under the action of a subset $\mathbb{H}$ of the group of invertible affine transformations of $\mathbb{R}^d$. We demonstrate that the scaled normalization of the $(K, \mathbb{H})$ -hull of a random sample, distributed according to a probability measure μ on K with a power-like behavior near the boundary ∂K of K , converges in distribution to a random closed set which can be viewed as the zero cell of a certain Poisson hyperplane tessellation with explicit intensity measure.

Key words: Generalized convexity, Poisson hyperplane tessellation, zero cell, curvature measure.

AMS 2020 classification: 60D05.

Introduction

A random closed set in a finite-dimensional Euclidean space is a random element taking values in the family of closed sets in this space equipped with the Fell topology. More information on the Fell topology can be found in Appendix B in (Molchanov, 2005). Over the past decades, considerable attention has been paid to the problem of generalized convexity and the related constructions emerging from random point samples. For instance, recall that for a subset A of the unit Euclidean ball, the ball hull of A is defined as the intersection of all translated unit balls that contain A . The paper (Fodor, Kevei, & Vígh, 2014) investigates ball hulls of random samples in $\mathbb{R}^2$, that is, the case when $A = \Xi_n = \{\xi_1, \dots, \xi_n\}$ consists of independent points uniformly distributed in the unit Euclidean disk. This study has been subsequently extended to ball hulls in $\mathbb{R}^d$ in (Fodor, 2020). This concept was further generalized by replacing the unit Euclidean ball with an arbitrary closed convex body K in (Fodor, Papvári, & Vígh, 2020) and (Marynych, & Molchanov, 2022). The approach evolved further in a recent study (Kabluchko, Marynych, & Molchanov, 2025), introducing the concept of a $(K, \mathbb{H})$ -hull in the following manner. Let K be a closed convex subset of $\mathbb{R}^d$, distinct from the entire space. Let $\mathbb{H}$ be a nonempty subset of $\mathbb{R}^d \times \mathbb{GL}_d$, where $\mathbb{GL}_d$ represents the group of all invertible linear transformations in $\mathbb{R}^d$. The $(K, \mathbb{H})$ hull of a set $A \subseteq \mathbb{R}^d$ is defined in (Kabluchko, Marynych, & Molchanov, 2025) as follows:

$$\text{conv}_{(K, \mathbb{H})}(A) = \bigcap_{(g, x) \in \mathbb{H}: A \subseteq g(K+x)} g(K+x),$$

where $g(B) = \{gz : z \in B\}$ and $B+x = \{z+x : z \in B\}$, for $(g, x) \in \mathbb{H}$ and $B \subseteq \mathbb{R}^d$.

We assume that $K \in \mathcal{K}_{(0)}^d$, where $\mathcal{K}_{(0)}^d$ is the family of compact convex sets that contain the origin in their interior. Let μ be a probability measure supported by K and $\Xi_n = \{\xi_1, \dots, \xi_n\}$ a set of n independent random points with the common distribution μ . Let M_d denote the space of all real-valued $d \times d$ matrices, and let $\exp : M_d \mapsto \mathbb{GL}_d$ denote the standard matrix exponent. The following set

$$\mathfrak{X}_n := \{(x, C) \in \mathbb{R}^d \times M_d : \Xi_n \subseteq \exp(C)(K+x)\},$$

which is the main object of investigation in the present paper, has been introduced and studied in (Kabluchko, Marynych, & Molchanov, 2025) in the case when μ is the uniform distribution on K . The importance of this set and its connections to $\text{conv}_{(K, \mathbb{H})}(\Xi_n)$ and also to the Lie algebra of $\mathbb{R}^d \times \mathbb{GL}_d$ are explained in (Kabluchko, Marynych, & Molchanov, 2025).

Assuming that μ is the uniform distribution on K , it was shown in (Kabluchko, Marynych, & Molchanov, 2025) that for any closed set $\mathfrak{D} \subset \mathbb{R}^d \times M_d$, the random closed set $n\mathfrak{X}_n \cap \mathfrak{D} \subset \mathbb{R}^d \times M_d$ converges in distribution, in the Fell topology, to a random closed set defined as a certain intersection of half-spaces in $\mathbb{R}^d \times M_d$. These half-spaces are derived from a Poisson point process whose intensity measure is determined by the curvature measure of K . The present study aims to determine the limiting behavior of an appropriately rescaled $\mathfrak{X}_n$ when the measure μ is not uniform. Our main result (Theorem 2 below) states that, under suitable normalization, the intersection $\mathfrak{X}_n \cap \mathfrak{D}$ converges in distribution to a random closed set, which is also defined as an intersection over the points of a Poisson point process. However, in this case, the intensity measure of the Poisson process depends not only on the curvature measure of K but also on the behavior of the measure μ near the boundary of K . The limiting random closed set can be interpreted as the zero cell of a Poisson hyperplane tessellation in this space. When $\mathbb{R}^d \times M_d$ is replaced by $\mathbb{R}^d \times \{0\}$, a similar result has been established in (Visona, 2024).

The remainder of the paper is structured as follows. In Section 1, we review relevant definitions from convex geometry and introduce our assumptions on the distribution μ . Section 2 is dedicated to the proof of Lemma 2, which is a variant of Proposition 2.2 from (Visona, 2024), as well as results from (Kiderlen, & Rataj, 2006). This lemma plays a central role in the proof of our main theorem. In Section 3, we state and prove our main result concerning the convergence of a suitably rescaled $\mathfrak{X}_n$. Several examples are presented at the end of Section 3.

1. Preliminaries and necessary definitions

Let us introduce some necessary definitions. $\text{Int } A$ denotes topological interior of a set $A \subset \mathbb{R}^d$. The Minkowski sum of two sets K and L is the following set, see Section 3.1 in (Schneider, 2013):

$$K \oplus L = \{k + l : k \in K, l \in L\}.$$

Minkowski difference of two sets K and L is defined as follows:

$$K \ominus L = \{x : x + L \subset K\}.$$

Let S^{d-1} denote the unit sphere in $\mathbb{R}^d$. By $B_r(a)$ we denote the closed ball in $\mathbb{R}^d$ of radius r and center $a \in \mathbb{R}^d$.

Even though in the present work we are interested in convex closed sets K , we find it useful to formulate some auxiliary results for a more general class of closed sets called gentle sets. To this end, we need to introduce some additional notation. For an arbitrary closed (and not necessarily convex) set K , the exoskeleton $\text{exo}(K)$ is the set of points that do not admit a unique nearest point in K . It is known that the set $\text{exo}(K)$ is Lebesgue measurable and, with V_d denoting the d -dimensional Lebesgue measure, $V_d(\text{exo}(K)) = 0$, see Section 2 in (Kiderlen, & Rataj, 2006). The metric projection $\xi_K : \mathbb{R}^d \setminus \text{exo}(K) \rightarrow K$ is defined by letting $\xi_K(a)$ be a unique point in K closest to a .

Let $K \subseteq \mathbb{R}^d$ be a closed set. The reduced normal bundle of K is

$$N(K) := \left\{ \left(\xi_K(z), \frac{z - \xi_K(z)}{\|z - \xi_K(z)\|} \right) : z \notin K \cup \text{exo}(K) \right\}.$$

The set

$$\hat{N}(K) := N(\partial K) \setminus N(K)$$

is called the inner reduced normal bundle of K . The reach function of K is a mapping $\delta(K, \cdot, \cdot) : N(K) \rightarrow [0, +\infty]$ defined by

$$\delta(K, a, u) := \inf\{t \geq 0 : a + tu \in \text{exo}(K)\}, \quad (a, u) \in N(K),$$

assuming that $\delta(K, a, u) = +\infty$ if the set $\{t \geq 0 : a + tu \in \text{exo}(K)\}$ is empty.

Let $\mathcal{H}^{d-1}$ denote $(d-1)$ -dimensional Hausdorff measure. A closed set K is said to be gentle, see (Kiderlen, & Rataj, 2006), if the following holds:

1. $\mathcal{H}^{d-1}(\{a \in B : (a, u) \in N(\partial K), u \in S^{d-1}\}) < \infty$ for all bounded Borel sets $B \subset \mathbb{R}^d$;
2. For $\mathcal{H}^{d-1}$ -almost all $a \in \partial K$, there are non-degenerate balls B_i and B_o containing a such that $B_i \subset K$ and $\text{Int } B_o \subset K^c$, where K^c denotes the set complement.

It is known, see p. 107 in (Kiderlen, & Rataj, 2006), that for $\mathcal{H}^{d-1}$ -almost all $a \in \partial K$, there exists a unique outer normal vector $u = u(a)$ such that $(a, u) \in N(K)$. Let $C_{d-1}(K, \cdot)$ be a measure defined as the image of $\mathcal{H}^{d-1}$ on ∂K under the map

$$\partial K \ni a \mapsto (a, u) \in N(K).$$

As explained above, this map is defined $\mathcal{H}^{d-1}$ -almost everywhere on ∂K and is measurable. The measure $C_{d-1}(K, \cdot)$ is called the curvature measure of K . Let $C_{d-1}^*(K, \cdot)$ be the image measure of $C_{d-1}(K, \cdot)$ under the reflection $(a, u) \in N(K) \mapsto (a, -u) \in \hat{N}(K)$, this reflection is well-defined almost everywhere.

We will use the abbreviation for almost all $(a, u) \in N(K)$ instead of for $C_{d-1}(K, \cdot)$ -almost all $(a, u) \in N(K)$. The same convention is used for $\hat{N}(K)$ equipped with the measure $C_{d-1}^*(K, \cdot)$.

The following assumptions on the measure μ on K are borrowed from (Visona, 2024). In what follows we shall work under these assumptions.

Let μ be a finite measure, supported by K , and satisfying:

- (M1) The measure μ has a compact support and is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ and with density f .
- (M2) There exists an $\alpha > -1$ such that, for almost all $(a, u) \in \hat{N}(K)$,

$$\lim_{t \downarrow 0} \frac{f(a + tu)}{t^\alpha} = \hat{g}(a, u) = g(a) \in [0, +\infty), \quad (1)$$

where the function $\hat{g}$ is strictly positive on a subset of ∂K of positive measure and is bounded almost everywhere.

Note that if K is a compact set, then μ always has a compact support.

We will also need some definitions for a convex compact set S . The support plane of S with outer normal vector $u \neq 0$ is the following set:

$$H(S, u) := \{x \in \mathbb{R}^d : \langle x, u \rangle = h(S, u)\},$$

where $h(S, \cdot)$ is the support function of S in the direction u defined by¹

$$h(S, u) := \sup\{\langle x, u \rangle : x \in S\}, \quad u \in S^{d-1}.$$

¹Observe that $h(S, \cdot)$ is well-defined for any compact set S (not necessarily convex) and $h(S, \cdot) \equiv h(\text{conv}(S), \cdot)$.

Let $H_u^-(t)$ denote the closed half-space defined by:

$$H_u^-(t) := \{x \in \mathbb{R}^d : \langle x, u \rangle \leq t\}, \quad t \in \mathbb{R}, \quad u \in S^{d-1}.$$

For v , which belong to the boundary ∂S of a convex compact set S , the normal cone $N(S, v)$ to S at v is defined by

$$N(S, v) := \{u \in \mathbb{R}^d \setminus \{0\} : v \in H(S, u)\} \cup \{0\}.$$

Let $Nor(S)$ denote the normal bundle of the closed convex S , that is, a subset of $\partial S \times S^{d-1}$, which is the collection $(x, N(S, x) \cap S^{d-1})$ for $x \in \partial S$. For the closed convex set S , the reduced normal bundle $N(S)$ coincides with $Nor(S)$, the normal bundle.

2. Measure change after taking Minkowski difference

Our proof uses a version of the Proposition 2.2 in (Visona, 2024). This theorem describes the asymptotic properties of a measure satisfying (M1) and (M2) with respect to the Minkowski difference between the gentle set K and a scaled compact set L . While in (Visona, 2024) the theorem is proven for the measure of $K \setminus (K \ominus \varepsilon L)$, as $\varepsilon \downarrow 0$, we extend this result in Lemma 2 to the case when the measured set under the limit is $\xi_{\partial K}^{-1}(C) \cap (K \setminus K \ominus \varepsilon L)$, where C is a bounded Borel set.

The following lemma was formulated and proven in (Visona, 2024), see Lemma 2.1 therein. We provide only the formulation of this lemma.

Lemma 1. *Let K be a gentle set and let L be a compact set in $\mathbb{R}^d$. Let f be a nonnegative integrable function and let $h(L, -u)^+$ denote positive part of the support function. Then, for almost all $(a, u) \in \hat{N}(K)$, there exist functions t_+ and t_- , which satisfy*

$$\lim_{\varepsilon \downarrow 0} \frac{t_{\pm}(\varepsilon)}{\varepsilon} = h(L, -u)^+,$$

and such that

$$\int_0^{t_+(\varepsilon)} f(a + tu) dt \leq \int_0^{\delta(\partial K, a, u)} f(a + tu) \mathbf{1}_{K \setminus K \ominus \varepsilon L}(a + tu) dt \leq \int_0^{t_-(\varepsilon)} f(a + tu) dt$$

for all sufficiently small ε .

Proof. The proof of this result can be found in Lemma 2.1 in (Visona, 2024). □

Lemma 2. *Let K be a gentle set, let C be a bounded Borel set from $\mathbb{R}^d$, and let L be a compact set in $\mathbb{R}^d$. Suppose that μ is a finite measure on K which has compact support, is absolutely continuous in a neighborhood of ∂K with density f that satisfies (M2). Put $\gamma := (\alpha + 1)^{-1}$. Then*

$$\lim_{\varepsilon \downarrow 0} \frac{\mu(\xi_{\partial K}^{-1}(C) \cap (K \setminus K \ominus \varepsilon^\gamma L))}{\varepsilon} = \int_{N(K)} \mathbf{1}_C(a) g(a) \frac{(h(L, u)^+)^{\alpha+1}}{\alpha + 1} C_{d-1}(K, d(a, u)).$$

Proof. Our demonstration closely mirrors the proof of Proposition 2.2 in (Visona, 2024). There exists a certain constant $s > 0$ such that the finite measure μ is absolutely continuous on the boundary $\partial K \oplus B_s(0)$. Additionally, there exists a constant $r > 0$ with $L \subseteq B_r(0)$.

For $\varepsilon < (s/r)^{1/\gamma}$, Proposition 4 from (Kiderlen, & Rataj, 2006) implies that

$$\begin{aligned} \mu(\xi_{\partial K}^{-1}(C) \cap (K \setminus K \ominus \varepsilon^\gamma L)) &= \int_{\mathbb{R}^d} f(x) \mathbf{1}_C(\xi_{\partial K}(x)) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(x) dx = \\ &= \sum_{i=1}^d \kappa_i \int_{N(\partial K)} \mathbf{1}_C(a) \int_0^{\delta(\partial K, a, u)} t^{i-1} f(a + tu) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a + tu) dt \nu_{d-i}(\partial K, d(a, u)), \end{aligned} \quad (2)$$

where the measures on ∂K , denoted as $\nu_0(\partial K, \cdot), \dots, \nu_{d-1}(\partial K, \cdot)$ are called support measures, with κ_i representing the volume of the i -dimensional unit ball for $i = 1, \dots, d$. According to Corollary 2.5 and equation (2.13) in (Hug, Last, & Weil, 2004), these support measures possess locally finite total variation. It is also important to highlight that the measures are signed for convex bodies.

The conditions outlined in Proposition 4 of (Kiderlen, & Rataj, 2006) necessitate that f is bounded. However, the proposition can be applied to integrable functions that are not necessarily bounded. Consider the sequence of functions defined by $f_n := f \mathbf{1}_{\{f \leq n\}}$, which is both monotonic and composed of bounded elements. Due to the principle of monotone convergence, equation (2) holds true even for integrable functions that are unbounded.

We next verify that for $i \geq 2$, the terms in (2) are on the order of $o(\varepsilon)$ as $\varepsilon \rightarrow 0$. The function $\mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}$ is supported within $\partial K \oplus B_{2r\varepsilon^\gamma}(0)$, implying that

$$\left| \frac{1}{\varepsilon} \int_{N(\partial K)} \mathbf{1}_C(a) \int_0^{\delta(\partial K, a, u)} t^{i-1} f(a + tu) \times \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a + tu) dt \nu_{d-i}(\partial K, d(a, u)) \right| \leq$$

$$\begin{aligned} &\leq \frac{1}{\varepsilon} \int_0^{2r\varepsilon^\gamma} \int_{N(\partial K)} \mathbf{1}_C(a) t^{i-1+\alpha} \frac{f(a+tu)}{t^\alpha} dt |\nu_{d-i}|(\partial K, d(a, u)) \leq \\ &\leq \frac{(2r\varepsilon^\gamma)^{i+\alpha}}{(i+\alpha)\varepsilon} \sup_{((a,u),t) \in N(\partial K) \times [0, 2r\varepsilon^\gamma]} \frac{f(a+tu)}{t^\alpha} \times |\nu_{d-i}|(\partial K, \partial K(\varepsilon) \times S^{d-1}) \rightarrow 0, \end{aligned}$$

as $\varepsilon \rightarrow 0$, for all $i \neq 1$, with $\gamma = (\alpha + 1)^{-1}$ and $\alpha > -1$, where

$$\partial K(\varepsilon) := \{x \in \partial K : x + y \in \text{supp } f \text{ for } y \in B_{2r\varepsilon^\gamma}(0)\},$$

which remains compact for all sufficiently small ε because the support of f is compact on ∂K . Consequently, the measure $|\nu_{d-i}|(\partial K, \partial K(\varepsilon) \times S^{d-1})$ is bounded for each integer i in the range 1 to $d-1$, due to $\partial K(\varepsilon)$ being bounded and these measures having a locally finite total variation. Additionally, from assumption (M2), the expression $f(a+tu)/t^\alpha$ remains constrained for almost every $((a, u), t)$ within the set $N(\partial K) \times [0, 2r\varepsilon^\gamma)$ when ε is sufficiently minimal.

Propositions 4.1 and 5.1 in (Hug, Last, & Weil, 2004) imply the equation:

$$2\nu_{d-1}(\partial K, \cdot) = C_{d-1}(K, \cdot) + C_{d-1}^*(K, \cdot).$$

For almost every $(a, u) \in N(K)$, we defined $\delta_\pm := \delta(\partial K, a, \pm u)$. Using that $\kappa_1 = 2$ we conclude

$$\begin{aligned} &\lim_{\varepsilon \downarrow 0} \frac{\mu(\xi_{\partial K}^{-1}(C) \cap (K \setminus K \ominus \varepsilon^\gamma L))}{\varepsilon} = \\ &= \lim_{\varepsilon \downarrow 0} \frac{2}{\varepsilon} \int_{N(\partial K)} \mathbf{1}_C(a) \int_0^{\delta(\partial K, a, u)} f(a+tu) \times \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a+tu) dt \nu_{d-1}(\partial K, d(a, u)) = \\ &= \lim_{\varepsilon \downarrow 0} \frac{2}{2\varepsilon} \int_{N(K)} \mathbf{1}_C(a) \int_0^{\delta^+} f(a+tu) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a+tu) dt C_{d-1}(K, d(a, u)) + \\ &+ \frac{2}{2\varepsilon} \int_{\hat{N}(K)} \mathbf{1}_C(a) \int_0^{\delta^-} f(a+tu) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a+tu) dt C_{d-1}^*(K, d(a, u)) = \\ &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_{\hat{N}(K)} \mathbf{1}_C(a) \int_0^{\delta^-} f(a+tu) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a+tu) dt C_{d-1}^*(K, d(a, u)), \end{aligned}$$

since the density function f vanishes outside K . Applying Lemma 1, we obtain for almost all $(a, u) \in \hat{N}(K)$,

$$\begin{aligned} g(a) \frac{(h(L, -u)^+)^{\alpha+1}}{\alpha+1} &= \lim_{\varepsilon \downarrow 0} \inf_{s \in [0, t_+(\varepsilon^\gamma))} \frac{f(a+su)}{s^\alpha} \frac{1}{\varepsilon} \int_0^{t_+(\varepsilon^\gamma)} t^\alpha dt \leq \\ &\leq \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^{t_+(\varepsilon^\gamma)} f(a+tu) dt \leq \\ &\leq \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^{\delta^-} f(a+tu) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a+tu) dt \leq \\ &\leq \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^{t_-(\varepsilon^\gamma)} f(a+tu) dt \leq \\ &\leq \lim_{\varepsilon \downarrow 0} \sup_{s \in [0, t_-(\varepsilon^\gamma))} \frac{f(a+su)}{s^\alpha} \frac{1}{\varepsilon} \int_0^{t_-(\varepsilon^\gamma)} t^\alpha dt = \\ &= g(a) \frac{(h(L, -u)^+)^{\alpha+1}}{\alpha+1}. \end{aligned}$$

Thus,

$$F_\varepsilon(a, u) := \frac{1}{\varepsilon} \int_0^{\delta^-} f(a+tu) \mathbf{1}_{K \setminus K \ominus \varepsilon^\gamma L}(a+tu) dt \rightarrow g(a) \frac{(h(L, -u)^+)^{\alpha+1}}{\alpha+1} \quad \text{as } \varepsilon \downarrow 0$$

for almost all $(a, u) \in \hat{N}(K)$. The limiting function is bounded and has compact support, because g is a bounded function with compact support and L is compact.

The sequence F_ε is bounded for all $\varepsilon > 0$ and for almost all $(a, u) \in \hat{N}(K) \cap ((\text{supp } f \cap \partial K) \times S^{d-1})$, which is compact. We can then find an upper bound and apply the dominated convergence theorem.

We conclude by noticing that

$$\int_{N(K)} \mathbf{1}_C(a) g(a) \frac{(h(L, -u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}^*(K, d(a, u)) = \int_{N(K)} \mathbf{1}_C(a) g(a) \frac{(h(L, u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a, u)).$$

□

The following corollary follows from the fact that $K \subseteq \mathcal{K}_{(0)}^d$ is also a gentle set, see Section 2 in (Kiderlen, & Rataj, 2006).

Corollary 1. *The limit relation in Lemma 2 holds true if $K \in \mathcal{K}_{(0)}^d$.*

3. Limit theorem for $\mathfrak{X}_n$

From now on, we will assume that $K \in \mathcal{K}_{(0)}^d$. Each set from $\mathcal{K}_{(0)}^d$ is a convex body (a compact convex set with a nonempty interior).

Let $\mathcal{P}_K$ denote Poisson point process on $(0, \infty) \times N(K)$, with the intensity measure ν defined as a product of an absolutely continuous measure on $(0, \infty)$ with density t^α and a measure ν^* on $N(K)$, defined by

$$\nu^*(K, W \cap N(K)) = \int_{N(K)} \mathbf{1}_{(a,u) \in W} g(a) C_{d-1}(K, d(a, u)),$$

where W is a Borel subset of $N(K)$.

We start with the lemma which is the main technical result, needed for the proof of our main result, Theorem 2 below. The proof follows the same reasoning as in Lemma 4.4 in (Kabluchko, Marynych, & Molchanov, 2025), with only minor modifications for our specific setting.

Lemma 3. *Let K be an element of $\mathcal{K}_{(0)}^d$, and assume μ meets conditions (M1) and (M2). It follows that*

$$\sum_{k=1}^n \delta_{(n^{-\gamma}(K - \xi_k)^\circ, \xi_k)} \xrightarrow{d} \sum_{(t,a,u) \in \mathcal{P}_K} \delta_{([0, t^{-1}u], a)} \quad \text{as } n \rightarrow \infty, \quad (3)$$

with the convergence interpreted as convergence in distribution within the space of point measures on $(\mathcal{K}_0^d \setminus \{0\}) \times \mathbb{R}^d$ furnished with the vague topology. Here, $\mathcal{K}_0^d$ represents the set of compact convex sets that include the origin (which may not be interior). As previously, we define $\gamma = (\alpha + 1)^{-1}$.

Proof. For $L \in \mathcal{K}_0^d \setminus \{0\}$ let L° denote its polar set defined by $L^\circ := \{x \in \mathbb{R}^d : h(L, x) \leq 1\}$. To begin, we observe that the subsequent equality is valid for the limiting Poisson process. $L \in \mathcal{K}_0^d \setminus \{0\}$ and every Borel $R \subseteq \mathbb{R}^d$

$$\begin{aligned} & \mathbf{P}\{[0, t^{-1}u] \subseteq L \text{ or } a \notin \xi_{\partial K}(R) \text{ for all } (t, a, u) \in \mathcal{P}_K\} = \\ &= \exp(-\nu(\{(t, a, u) : [0, t^{-1}u] \not\subseteq L, a \in \xi_{\partial K}(R)\})) = \\ &= \exp(-\nu(\{(t, a, u) : L^\circ \not\subseteq H_u^-(t), a \in \xi_{\partial K}(R)\})) = \\ &= \exp(-\nu(\{(t, a, u) : h(L^\circ, u) > t, a \in \xi_{\partial K}(R)\})) = \\ &= \exp\left(-\int_{N(K)} \mathbf{1}_{\{a \in \xi_{\partial K}(R)\}} g(a) \frac{(h(L^\circ, u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a, u))\right). \end{aligned}$$

According to Proposition 7.9 in the Appendix of (Kabluchko, Marynych, & Molchanov, 2025), we need to show that, for every $L \in \mathcal{K}_0^d \setminus \{0\}$ and every Borel $R \subseteq \mathbb{R}^d$, as $n \rightarrow \infty$,

$$n\mathbf{P}\{n^{-\gamma}(K - \xi)^\circ \not\subseteq L, \xi \in R\} \longrightarrow \int_{N(K)} \mathbf{1}_{\{a \in \xi_{\partial K}(R)\}} g(a) \frac{(h(L^\circ, u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a, u)). \quad (4)$$

Note that

$$\begin{aligned} & \lim_{n \rightarrow \infty} n\mathbf{P}\{n^{-\gamma}(K - \xi)^\circ \not\subseteq L, \xi \in R\} = \lim_{n \rightarrow \infty} n\mathbf{P}\{n^{-\gamma}L^\circ \not\subseteq (K - \xi), \xi \in R\} = \\ &= \lim_{n \rightarrow \infty} n\mathbf{P}\{\xi \notin K \ominus n^{-\gamma}L^\circ, \xi \in R\} = \lim_{n \rightarrow \infty} n\mu(R \cap (K \setminus (K \ominus n^{-\gamma}L^\circ))) = \\ &= \int_{N(K)} \mathbf{1}_{\{a \in R\}} g(a) \frac{(h(L^\circ, u)^+)^{\alpha+1}}{\alpha+1} C_{d-1}(K, d(a, u)), \end{aligned}$$

where the last equality follows from Corollary 1, where $\varepsilon = n^{-1}$, L is replaced by L° and $C = \xi_{\partial K}(R)$. Thus, (4) is verified and the proof is complete. $\square$

Let us formulate now the main theorem of this work.

Theorem 2. *Assume that $K \in \mathcal{K}_{(0)}^d$ and the measure μ satisfies (M1) and (M2). Suppose that $\mathfrak{D}$ is a closed convex set in $\mathbb{R}^d \times \mathbb{M}_d$ which contains the origin and let $\gamma = (\alpha + 1)^{-1}$ be a real parameter, with $\alpha > -1$. The sequence of random closed sets $((n^\gamma \mathfrak{X}_n) \cap \mathfrak{D})_{n \in \mathbb{N}}$ converges in distribution in the space of closed subsets of $\mathbb{R}^d \times \mathbb{M}_d$ endowed with the Fell topology to a random closed convex set $\check{\mathfrak{Z}}_K \cap \mathfrak{D}$, where*

$$\mathfrak{Z}_K := \bigcap_{(t,\eta,u) \in \mathcal{P}_K} \left\{ (x, C) \in \mathbb{R}^d \times \mathbb{M}_d : \langle C\eta + x, u \rangle \leq t \right\} \quad (5)$$

and $\check{\mathfrak{Z}}$ is the reflection of $\mathfrak{Z}_K$ with respect to the origin in $\mathbb{R}^d \times \mathbb{M}_d$.

Proof. The structure of this proof will closely resemble that of Theorem 4.1 in (Kabluchko, Marynych, & Molchanov, 2025). Referring to Lemma 7.3 in the same source, it is necessary to demonstrate that $(n^\gamma \mathfrak{X}_n) \cap \mathfrak{D} \cap \mathfrak{K}$ converges to $\check{\mathfrak{Z}}_K \cap \mathfrak{D} \cap \mathfrak{K}$.

This should be shown for an arbitrary compact convex subset $\mathfrak{K}$ of $\mathbb{R}^d \times M_d$ that contains the origin in its interior, and subsequently, one must take the limit as $\mathfrak{K}$ approaches $(\mathbb{R}^d \times M_d)$. Consequently, it holds

$$\begin{aligned} (n^\gamma \mathfrak{X}_n) \cap \mathfrak{K} &= \left\{ (x, C) \in \mathfrak{K} : \Xi_n \subseteq \exp(C/n^\gamma)(K + x/n^\gamma) \right\} = \\ &= \bigcap_{k=1}^n \left\{ (x, C) \in \mathfrak{K} : \xi_k \in \exp(C/n^\gamma)(K + x/n^\gamma) \right\} = \\ &= \bigcap_{k=1}^n \left\{ (x, C) \in \mathfrak{K} : \exp(-C/n^\gamma)\xi_k \in K + x/n^\gamma \right\} = \\ &= \bigcap_{k=1}^n \left\{ (x, C) \in \mathfrak{K} : (n^\gamma(\exp(-C/n^\gamma) - I))\xi_k - x \in n^\gamma(K - \xi_k) \right\}, \end{aligned}$$

where I is the identity matrix in M_d . Let

$$a_m := \sup_{n \geq m} \sup_{(x, C) \in \mathfrak{K}, y \in K} \left\| (n^\gamma(\exp(-C/n^\gamma) - I)y + Cy) \right\|, \quad m \in \mathbb{N}.$$

Note that $a_m \rightarrow 0$ as $m \rightarrow \infty$, because $n^\gamma(\exp(-C/n^\gamma) - I) \rightarrow -C$ locally uniformly in C as $n \rightarrow \infty$, the set $\mathfrak{K}$ is compact in $\mathbb{R}^d \times M_d$, and K is compact in $\mathbb{R}^d$.

Recall that B_{a_m} denotes the closed ball of radius a_m in $\mathbb{R}^d$ centred at the origin. For each $m \in \mathbb{N}$ and $n^\gamma \geq m$, we have

$$\mathfrak{N}_{m,n}^- \subseteq ((n^\gamma \mathfrak{X}_n) \cap \mathfrak{K}) \subseteq \mathfrak{N}_{m,n}^+, \quad (6)$$

where

$$\mathfrak{N}_{m,n}^+ := \bigcap_{k=1}^n \left\{ (x, C) \in \mathfrak{K} : -C\xi_k - x \in n^\gamma(K - \xi_k) + B_{a_m} \right\}$$

and

$$\mathfrak{N}_{m,n}^- := \bigcap_{k=1}^n \left\{ (x, C) \in \mathfrak{K} : -C\xi_k - x + B_{a_m} \subseteq n^\gamma(K - \xi_k) \right\}. \quad (7)$$

The benefit of the lower and upper bounds in (6) is the convex nature of $\mathfrak{N}_{m,n}^\pm$, simplifying their analysis. We plan to utilize Lemma 7.8 from (Kabluchko, Marynych, & Molchanov, 2025) by setting $Y_{m,n}^\pm = \mathfrak{N}_{m,n}^\pm \cap \mathfrak{D}$ and $X_n = (n^\gamma \mathfrak{X}_n) \cap \mathfrak{K} \cap \mathfrak{D}$.

Let $\mathfrak{L}$ be a compact subset of $\mathfrak{K}$. Denote

$$\begin{aligned} M_m^+(\mathfrak{L}) &:= \{(L, y) \in \mathcal{K}_0^d \times \mathbb{R}^d : -Cy - x \in L^\circ + B_{a_m} \text{ for all } (x, C) \in \mathfrak{L}\}, \\ M_m^-(\mathfrak{L}) &:= \{(L, y) \in \mathcal{K}_0^d \times \mathbb{R}^d : -Cy - x + B_{a_m} \subseteq L^\circ \text{ for all } (x, C) \in \mathfrak{L}\}, \\ M(\mathfrak{L}) &:= \{(L, y) \in \mathcal{K}_0^d \times \mathbb{R}^d : -Cy - x \in L^\circ \text{ for all } (x, C) \in \mathfrak{L}\}. \end{aligned}$$

Then

$$\mathbf{P}\{\mathfrak{L} \subseteq \mathfrak{N}_{m,n}^\pm\} = \mathbf{P}\{(n^{-\gamma}(K - \xi_i)^\circ, \xi_i) \in M_m^\pm(\mathfrak{L}), i = 1, \dots, n\}.$$

By Lemma 3, the point process $\{(n^{-\gamma}(K - \xi_i)^\circ, \xi_i), i = 1, \dots, n\}$ converges in distribution to the Poisson process $\{([0, t^{-1}u], \eta) : (t, \eta, u) \in \mathcal{P}_K\}$. The sets $M(\mathfrak{L})$, $M_m^\pm(\mathfrak{L})$ are continuity sets for the distribution of the limiting Poisson process. Indeed, for each $(t, \eta, u) \in (0, \infty) \times N(K)$,

$$\begin{aligned} \{([0, t^{-1}u], \eta) \in \partial M_m^+(\mathfrak{L})\} &= \\ &= \left\{ -C\eta - x \in H_u^-(t + a_m) \text{ for all } (x, C) \in \mathfrak{L} \right\} \setminus \left\{ -C\eta - x \in \text{Int } H_u^-(t + a_m) \text{ for all } (x, C) \in \mathfrak{L} \right\} = \\ &= \left\{ \langle -C\eta - x, u \rangle \leq t + a_m \text{ for all } (x, C) \in \mathfrak{L} \right\} \setminus \left\{ \langle -C\eta - x, u \rangle < t + a_m \text{ for all } (x, C) \in \mathfrak{L} \right\} = \\ &= \left\{ \langle -C\eta - x, u \rangle \leq t + a_m \text{ for all } (x, C) \in \mathfrak{L} \right\} \cap \left\{ \langle -C\eta - x, u \rangle = t + a_m \text{ for some } (x, C) \in \mathfrak{L} \right\}, \end{aligned}$$

where Int denotes the topological interior. Since the probability of the latter event for some $(t, \eta, u) \in \mathcal{P}_K$ vanishes, it follows that $M_m^+(\mathfrak{L})$ is a continuity set for the Poisson point process $\{([0, t^{-1}u], \eta) : (t, \eta, u) \in \mathcal{P}_K\}$. Letting $a_m = 0$, we obtain that $M(\mathfrak{L})$ is also a continuity set. The argument for $M_m^-(\mathfrak{L})$ is similar by replacing a_m with $(-a_m)$.

Thus, for all $m \in \mathbb{N}$,

$$\mathbf{P}\{\mathfrak{L} \subseteq \mathfrak{N}_{m,n}^+\} \rightarrow \mathbf{P}\{\{([0, t^{-1}u], \eta) : (t, \eta, u) \in \mathcal{P}_K\} \subseteq M_m^+(\mathfrak{L})\} = \mathbf{P}\{\mathfrak{L} \subseteq \mathfrak{N}_m^+\}$$

as $n \rightarrow \infty$, where

$$\mathfrak{N}_m^+ := \bigcap_{(t, \eta, u) \in \mathcal{P}_K} \left\{ (x, C) \in \mathfrak{K} : -C\eta - x \in H_u^-(t + a_m) \right\}.$$

The random closed sets $\mathfrak{N}_{m,n}^+$ and $\mathfrak{N}_m^+$ are convex and almost surely contain a neighbourhood of the origin in $\mathbb{R}^d \times M_d$, hence, are regular closed, meaning that they coincide with the closure of their respective interiors. By Theorem 7.5 from (Kabluchko, Marynych, & Molchanov, 2025) applied to the space $\mathbb{R}^d \times M_d$, the random convex set $\mathfrak{N}_{m,n}^+$ converges in distribution to $\mathfrak{N}_m^+$. Since $\mathfrak{N}_{m,n}^+ \xrightarrow{d} \mathfrak{N}_m^+$ as $n \rightarrow \infty$, and the involved sets almost surely contain the origin in their interiors, Corollary 7.7 from (Kabluchko, Marynych, & Molchanov, 2025) yields that $(\mathfrak{N}_{m,n}^+ \cap \mathfrak{D}) \xrightarrow{d} (\mathfrak{N}_m^+ \cap \mathfrak{D})$ as $n \rightarrow \infty$, for each closed convex set $\mathfrak{D}$ which contains the origin in $\mathbb{R}^d \times M_d$ (not necessarily as an interior point). Thus, we have checked part (i) of Lemma 7.8 from (Kabluchko, Marynych, & Molchanov, 2025).

We proceed with checking part (ii) of Lemma 7.8 from (Kabluchko, Marynych, & Molchanov, 2025) with $Y_m^- := \mathfrak{N}_m^-$, where

$$\mathfrak{N}_m^- := \bigcap_{(t,\eta,u) \in \mathcal{P}_K} \left\{ (x, C) \in \mathfrak{K} : -C\eta - x \in H_u^-(t - a_m) \right\}.$$

Note that the random sets $\mathfrak{N}_{m,n}^-$ and $\mathfrak{N}_m^-$ may be empty and otherwise not necessarily contain the origin. We need to check 7.14 from (Kabluchko, Marynych, & Molchanov, 2025), which in our case reads as follows

$$\mathbf{P} \left\{ \mathfrak{N}_{m,n}^- \cap \mathfrak{D} \cap \mathfrak{L} \neq \emptyset, 0 \in \mathfrak{N}_{m,n}^- \right\} \rightarrow \mathbf{P} \left\{ \mathfrak{N}_m^- \cap \mathfrak{D} \cap \mathfrak{L} \neq \emptyset, 0 \in \mathfrak{N}_m^- \right\} \quad \text{as } n \rightarrow \infty, \quad (8)$$

for all compact sets $\mathfrak{L}$ which are continuity sets of $\mathfrak{N}_m^- \cap \mathfrak{D}$. We shall prove (8) for all compact sets $\mathfrak{L}$ in $\mathbb{R}^d \times M_d$. To this end, we shall employ Lemma 7.6 (Kabluchko, Marynych, & Molchanov, 2025) and divide the derivation (8) into several steps, each devoted to checking one condition of Lemma 7.6 from (Kabluchko, Marynych, & Molchanov, 2025).

Step 1. Let us check that, for sufficiently large $n \in \mathbb{N}$,

$$\mathbf{P} \left\{ (0, 0) \in \mathfrak{N}_{m,n}^- \right\} = \mathbf{P} \left\{ (0, 0) \in \text{Int} \mathfrak{N}_{m,n}^- \right\} > 0, \quad m \in \mathbb{N}.$$

Since the interior of a finite intersection is the intersection of the interiors, and using independence, it suffices to check this for each of the sets which appear in the intersection in (7). If $(0, 0)$ belongs to $Y_k := \{(x, C) \in \mathfrak{K} : -C\xi_k - x + B_{a_m} \subseteq n^\gamma(K - \xi_k)\}$, then $B_{a_m} \subseteq n^\gamma(K - \xi_k)$. Since ξ_k is not concentrated on ∂K , we have

$$\mathbf{P} \{ B_{a_m} \subseteq n^\gamma(K - \xi_k) \} = \mathbf{P} \{ B_{a_m} \subseteq n^\gamma \text{Int}(K - \xi_k) \}$$

for all n . If $B_{a_m} \subseteq n^\gamma \text{Int}(K - \xi_k)$, then $-C\xi_k - x + B_{a_m} \subseteq n^\gamma(K - \xi_k)$ for all x and C from a sufficiently small neighbourhood of the origin in $\mathbb{R}^d \times M_d$. Furthermore,

$$\mathbf{P} \left\{ (0, 0) \in \mathfrak{N}_{m,n}^- \right\} = \mathbf{P} \{ B_{a_m} \subseteq n^\gamma(K - \xi_k), k = 1, \dots, n \} = \mathbf{P} \{ B_{a_m/n^\gamma} \subseteq K \ominus \Xi_n \} > 0$$

for all sufficiently large n .

Step 2. Let us check that, for each $m \in \mathbb{N}$,

$$\mathbf{P} \left\{ (0, 0) \in \mathfrak{N}_m^- \right\} = \mathbf{P} \left\{ (0, 0) \in \text{Int} \mathfrak{N}_m^- \right\} > 0.$$

The equality above follows from the observation that the origin lies on the boundary of $\{(x, C) \in \mathfrak{K} : -C\eta - x \in H_u^-(t - a_m)\}$ only if $t = a_m$, which happens with probability zero. Furthermore, $(0, 0) \in \mathfrak{N}_m^-$ if $t \geq a_m$ for all $(t, \eta, u) \in \mathcal{P}_K$, which has positive probability.

Step 3. By a similar argument as we have used for $\mathfrak{N}_{m,n}^+$, for every compact subset $\mathfrak{L}$ of $\mathfrak{K}$ and $m \in \mathbb{N}$, we have

$$\mathbf{P} \left\{ \mathfrak{L} \subseteq \mathfrak{N}_{m,n}^- \right\} \rightarrow \mathbf{P} \left\{ \left\{ ([0, t^{-1}u], \eta) : (t, \eta, u) \in \mathcal{P}_K \right\} \subseteq M_m^-(\mathfrak{L}) \right\} = \mathbf{P} \left\{ \mathfrak{L} \subseteq \mathfrak{N}_m^- \right\} \quad \text{as } n \rightarrow \infty.$$

In summary, we have checked all conditions of Lemma 7.6 (Kabluchko, Marynych, & Molchanov, 2025). This finishes the proof of (8) and shows that all the conditions of part (ii) of Lemma 7.8 from (Kabluchko, Marynych, & Molchanov, 2025) hold. It remains to note that

$$(\mathfrak{N}_m^+ \cap \mathfrak{D}) \downarrow (\check{\mathfrak{Z}}_K \cap \mathfrak{K} \cap \mathfrak{D}), \quad (\mathfrak{N}_m^- \cap \mathfrak{D}) \uparrow (\check{\mathfrak{Z}}_K \cap \mathfrak{K} \cap \mathfrak{D}) \quad \text{a.s. as } m \rightarrow \infty$$

in the Fell topology, and

$$\lim_{m \rightarrow \infty} \mathbf{P} \{ 0 \in \mathfrak{N}_m^- \} = 1.$$

Thus, by Lemma 7.8 from (Kabluchko, Marynych, & Molchanov, 2025) $(n^\gamma \mathfrak{X}_n) \cap \mathfrak{K} \cap \mathfrak{D}$ converges in distribution to $\check{\mathfrak{Z}}_K \cap \mathfrak{K} \cap \mathfrak{D}$ as $n \rightarrow \infty$. By Lemma 7.3 from (Kabluchko, Marynych, & Molchanov, 2025), $(n^\gamma \mathfrak{X}_n) \cap \mathfrak{D}$ converges in distribution to $\check{\mathfrak{Z}}_K \cap \mathfrak{D}$. $\square$

Note that $\check{\mathfrak{Z}}_K$ is almost surely a closed set as the intersection of half-spaces. The question of boundedness depends on the properties of measure ν^* . If it is supported by sets contained in $\partial K \times H \subset \partial K \times S^{d-1}$, where H denotes a closed hemisphere of S^{d-1} , then this set is almost surely unbounded and is almost surely bounded otherwise.

As set $\check{\mathfrak{Z}}_K$ is an element of space $\mathbb{R}^d \times M_d$, the latter space can be turned into a Euclidian one, given the proper scalar product, see Section 5.1 in (Kabluchko, Marynych, & Molchanov, 2025). Then, the set $\check{\mathfrak{Z}}_K$ can be viewed as the zero cell of the Poisson hyperplane tessellation in space $\mathbb{R}^d \times M_d$.

Corollary 3. *Convergence in distribution of random closed sets in Hausdorff metric also implies convergence in distribution*

of intrinsic volumes, so

$$V_j(n^\gamma \mathfrak{X}_n \cap \mathfrak{D}) \xrightarrow{d} V_j(\mathfrak{Z}_K \cap \mathfrak{D}) \quad \text{as } n \rightarrow \infty, j = 0, 1, \dots, d,$$

where V_j denotes the j -th intrinsic volume.

Before we proceed with example, it is necessary to introduce some new definitions as well as some assumptions, borrowed from (Kabluchko, Marynych, & Molchanov, 2025). Let $\mathbb{V}$ be a sufficiently small neighbourhood of I in $\mathbb{GL}_d$, where the matrix exponent is bijective. Let $\log : \mathbb{V} \mapsto M_d$ be the inverse of the matrix exponent. Define mapping $\widetilde{\exp} : \mathbb{R}^d \times \log \mathbb{V} \mapsto \mathbb{R}^d \times \mathbb{V}$ by $\widetilde{\exp}(x, h) = (x, \exp h)$, $x \in \mathbb{R}^d$, $h \in \log \mathbb{V}$.

Assume that the following objects exist:

- a neighbourhood $\mathfrak{U} \subseteq \mathbb{R}^d \times \log \mathbb{V}$ of $(0, 0)$ in $\mathbb{R}^d \times M_d$;
- a neighbourhood $\mathfrak{U} \subseteq \mathbb{R}^d \times \mathbb{V}$ of $(0, I)$ in $\mathbb{R}^d \times \mathbb{GL}_d$;
- a closed convex cone $\mathfrak{C}_{\mathbb{H}}$ in $\mathbb{R}^d \times M_d$ with apex at $(0, 0)$ such that

$$\mathbb{H} \cap \mathfrak{U} = \widetilde{\exp}(\mathfrak{C}_{\mathbb{H}} \cap \mathfrak{U}).$$

Similarly to (Kabluchko, Marynych, & Molchanov, 2025), in our examples we will look at intersections of $\mathfrak{Z}_K$ with the cone $\mathfrak{C}_{\mathbb{H}}$. More information on properties of this cone can be found on pages 12-13 of (Kabluchko, Marynych, & Molchanov, 2025). In the following example, we examine properties of such $\mathfrak{Z}_K$ that arise as part of the limit in Theorem 2, where $\mathfrak{X}_n$ is defined using Ξ_n sampled according to a probability measure μ satisfying assumptions (M1) and (M2).

Example. Let K be a unit ball B_1 and $\mathbb{H} = \mathbb{T} \times \mathbb{G} = \{0\} \times \mathbb{GL}_d$. In this case, $\mathfrak{C}_{\mathbb{H}} = \{0\} \times M_d$ and

$$\mathfrak{Z}_{B_1} \cap (\{0\} \times M_d) = \{0\} \times \bigcap_{(t,u) \in \mathcal{P}} \{C \in M_d : \langle Cu, u \rangle \leq t\},$$

where $\mathcal{P}$ is a Poisson point process on $\mathbb{R}_+ \times S^{d-1}$ with intensity measure being the product of an absolutely continuous measure with density t^α and measure ν^* on S^{d-1} , defined as

$$\nu^*(K, W \cap S^{d-1}) = \int_{N(B_1)} \mathbf{1}_{u \in Wg(a)} C_{d-1}(B_1, d(a, u)),$$

where $g(a)$ is defined in the same way as in (1).

Discussion and conclusions

In this work, we studied the asymptotic behavior of generalized convex hulls defined via group actions when the underlying point sample is drawn from a non-uniform distribution on a convex body K . Although the construction as well as limit theorems for the $(K, \mathbb{H})$ -hulls were introduced in (Kabluchko, Marynych, & Molchanov, 2025) under the assumption of uniform sampling, our main contribution lies in identifying how the local behavior of the sampling density near the boundary of K affects the limit. We proved that under power-like decay conditions, the scaled hull converges in distribution to a random closed set defined as the zero cell of a Poisson hyperplane tessellation, where the intensity measure now depends not only on the curvature measure of K , but also explicitly on the boundary behavior of the density. This result confirms and extends previous findings from (Visona, 2024) on randomly translated sets and shows that the limit structure is sensitive to the anisotropy introduced by the sampling scheme. Our approach demonstrates that non-uniformity introduces additional geometric information into the scaling limit, opening new perspectives for modeling and analysis. These insights are likely relevant in applications where data is naturally concentrated near certain regions of a domain such as in image processing, statistical physics, or non-homogeneous spatial processes.

Acknowledgments. I express my deep gratitude to Professor Alexander Marynych for his invaluable help, insightful advice, and careful reading of the drafts of this paper. His suggestions and corrections greatly improved both the clarity and the quality of the final work.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Fodor, F. (2020). *Random ball-polytopes in smooth convex bodies*. <https://arxiv.org/abs/1906.11480>
- Fodor, F., Kevei, P., & Vigh, V. (2014). On random disc polygons in smooth convex discs. *Advances in Applied Probability*, 46(4), 899–918. <https://doi.org/10.1239/aap/1418396236>
- Fodor, F., Papvári, D. I., & Vigh, V. (2020). On random approximations by generalized disc-polygons. *Mathematika*, 66(2), 498–513. <https://doi.org/10.1112/mtk.12027>
- Hug, D., Last, G., & Weil, W. (2004). A local steiner-type formula for general closed sets and applications. *Mathematische Zeitschrift*, 246, 237–272. <https://doi.org/10.1007/s00209-003-0597-9>
- Kabluchko, Z., Marynych, A., & Molchanov, I. (2025). Generalised convexity with respect to families of affine maps. *Israel Journal of Mathematics*, 266, 131–175. <https://doi.org/10.1007/s11856-025-2751-0>
- Kiderlen, M., & Rataj, J. (2006). On infinitesimal increase of volumes of morphological transforms. *Mathematika*, 53(1), 103–127. <https://doi.org/10.1112/S002557930000005X>
- Marynych, A., & Molchanov, I. (2022). Facial structure of strongly convex sets generated by random samples. *Advances in Mathematics*, 395, 108086. <https://doi.org/10.1016/j.aim.2021.108086>
- Molchanov, I. (2005). *Theory of random sets*. Springer London. <https://doi.org/10.1007/1-84628-150-4>

- Schneider, R. (2013). *Convex bodies: The brunn-minkowski theory* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139003858>
- Visona, T. (2024). Intersections of randomly translated sets. *Journal Of Theoretical Probability*, 38, 3. <https://doi.org/10.1007/s10959-024-01371-z>

Отримано редакцією журналу / Received: 11.06.25
Прорецензовано / Revised: 22.09.25
Схвалено до друку / Accepted: 10.10.25

Микита САДОК, асп.
ORCID ID: 0009-0000-0215-4204
e-mail: sadokmykyta@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

ГРАНИЧНІ ТЕОРЕМИ ДЛЯ УЗАГАЛЬНЕНИХ ОПУКЛИХ ОБОЛОНОК

$(K, \mathbb{H})$ -оболонка множини $A \subseteq \mathbb{R}^d$ – це узагальнення стандартної замкненої опуклої оболонки A , що може бути визначена як перетин усіх образів опуклої замкненої підмножини $K \subset \mathbb{R}^d$ під дією підмножини $\mathbb{H}$ групи усіх оборотних афінних перетворень в $\mathbb{R}^d$, що містять A . Доведено, що масштабована нормалізація $(K, \mathbb{H})$ -оболонки випадкової вибірки, розподіленої згідно з імовірнісною мірою μ на K , яка веде себе як степінь біля межі ∂K , збігається за розподілом до випадкової замкненої множини, яку можна розглядати як нульову клітинку деякої пуассонівської гіперплощинної мозаїки з явною мірою інтенсивності.

Ключові слова: узагальнена опуклість, пуассонівська гіперплощинна мозаїка, нульова клітинка, міра кривизни.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження(у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 519.21

DOI: <https://doi.org/10.17721/1812-5409.2025/2.10>

Dmytro TYKHONENKO, PhD Student

ORCID ID: 0009-0002-1144-6852

e-mail: tykhonenko.dmytro@gmail.com

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Rostyslav YAMNENKO, DSc (Phys. & Math.)

ORCID ID: 0000-0002-9612-7959

e-mail: rostyslav.yamnenko@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Katalin KUCHINKA, PhD (Phys. & Math.)

ORCID ID: 0000-0001-8129-9937

e-mail: kucsinka.katalin@unithe.hu

Ferenc Rakoczi II Transcarpathian Hungarian University, Berehove, Ukraine; University Tokaj, Sárospatak, Hungary

IMPROVEMENT OF CHARACTERIZATION INEQUALITY FOR NORMS OF φ -SUB-GAUSSIAN RANDOM VARIABLE

The paper is focused on refining the characterization inequality for the norm of φ -sub-Gaussian random variables. The core of this improvement lies in a detailed analysis of the remainder terms in Stirling's formula, which provides an approximation for $n!$. This study conducts a comparative analysis of several pairs of known remainder terms, α_n and β_n , to identify the pair that minimizes the approximation error. By calculating the absolute difference for these pairs, we demonstrate that the tightest bounds are achieved with a specific selection of higher-order terms. This optimized approximation for $n!$ is then applied to sharpen the constant in a characterization inequality for comparing different norms of a φ -sub-Gaussian random variable and also in the theorem on φ -sub-Gaussianity of a random variable.

Key words: Stirling's formula, remainder, characterization inequality, norm, φ -sub-Gaussian random variable.

AMS 2020 classification: 60E15, 26D15.

Introduction

Stirling's formula is a cornerstone of mathematical analysis, offering a powerful approximation for the factorial function, $n!$. The classical representation of this formula, known since the 18th century, states that as n approaches infinity, the ratio of $n!$ to $\sqrt{2\pi n}(n/e)^n$ converges to 1. This approximation is indispensable in various fields, including probability theory, statistics, and physics, where direct computation of factorials for large n is impractical.

While the limit-based formula is powerful, its practical application often requires more precise error bounds. Consequently, a significant body of research has been dedicated to refining Stirling's original approximation. The primary objective was to conduct a comparative analysis of several of these well-established remainder terms. The goal is to identify the specific pair of α_n and β_n that yields the least approximation error among the considered options.

The refined result is then applied to sharpen the characterization inequality for the norm of φ -sub-Gaussian random variables. These variables are a generalization of Gaussian random variables and have been widely researched and described in monographs (Buldygin, & Kozachenko, 2000; Vasylyk, Kozachenko, & Yamnenko, 2008).

1. Stirling's Formula Remainders

Since Stirling's time (1764), it has been known that $n! \sim C\sqrt{n}(n/e)^n$, with $C = \sqrt{2\pi}$, that is

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n}(n/e)^n} = 1. \quad (1)$$

Formula (1) is known as Stirling's formula, and many elementary proofs of it have been given. Most such proofs use Wallis' formula

$$\lim_{n \rightarrow \infty} \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot \dots \cdot 2n \cdot 2n}{1 \cdot 3 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)(2n+1)} = \frac{\pi}{2}, \quad (2)$$

to show that $C = \sqrt{2\pi}$. However, one of the elementary proofs by W. Feller (Feller, 1967) avoids any appeal to Wallis' formula. Since completely elementary proofs of (2) are well known (see, for example, (Schwartz, 1967)), this does not seem to be essential. In addition, a number of upper and lower bounds for $n!$, most of which imply (1), have been obtained by various authors. One of the most elementary estimates $e^{11/12}\sqrt{n}(n/e)^n < n! < e\sqrt{n}(n/e)^n$ was obtained by Hummel (Hummel, 1940) in 1940. Most bounds are of the form

$$\sqrt{2\pi n}(n/e)^n e^{\alpha_n} < n! < \sqrt{2\pi n}(n/e)^n e^{\beta_n}, \quad (3)$$

where α_n and β_n tend to zero through positive values. For example, $\beta_n = (12n)^{-1}$ was proved in (Uspensky, 1937), while the successively better values

$$\alpha_n^{(1)} = \frac{1}{\left(12n + \frac{3}{2(2n+1)}\right)}, \quad \alpha_n^{(2)} = \frac{1}{12n} - \frac{1}{360n^3}, \quad \alpha_n^{(3)} = \frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{\left(1260\left(1 + \frac{2}{n(n+1)}\right)n^5\right)}, \quad (4)$$

$$\begin{aligned}\beta_n^{(1)} &= \frac{1}{12n} - \frac{1}{((360 + \gamma_n)n^3)}, \quad \text{where } \gamma_n = 30 \frac{7n(n+1)+1}{n^2(n+1)^2}, \\ \beta_n^{(2)} &= \frac{1}{12n} - \frac{1}{(360n(n+1)(n+2))}, \\ \beta_n^{(3)} &= \frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5}\end{aligned}\quad (5)$$

were obtained in (Maria, 1965; Nanjundiah, 1965; Beesack, 1969; Hu, Lui, & Shi, 2006; Long, 2014; Macyz, 2021). As proofs for the mentioned remainders are given – we focus on choosing the left and right remainders that will give us the least $n!$ approximation error among those listed.

To do this, we need to calculate the possible error for each pair of left and right remainders. In common sense, it looks like $\frac{e^{\alpha n} - e^{\beta n}}{e^{\beta n}} = e^{\alpha n - \beta n} - 1$. And for an equation like this – we can use the fact that for small x , $e^x - 1 \approx x$.

Let's calculate the first pair of α and β with γ_n defined in (5).

$$\beta_n^{(1)} = \frac{1}{12n} - \frac{1}{(360 + \gamma_n)n^3} = \frac{1}{12n} - \frac{1}{(360 + \frac{210n^2 + 210n + 30}{(n^2 + n)^2})n^3} = \frac{33n^3 + 91n^2 + 123n + 60}{60n(7n^3 + 19n^2 + 25n + 12)}.\quad (6)$$

$$\alpha_n^{(1)} = \frac{1}{\left(12n + \frac{3}{2(2n+1)}\right)} = \frac{4n + 2}{48n^2 + 24n + 3}.\quad (7)$$

With the equations calculated above (6) and (7), we can compute the absolute difference between α and β .

$$\begin{aligned}|\alpha_n^{(1)} - \beta_n^{(1)}| &= \left| \frac{4n + 2}{48n^2 + 24n + 3} - \left(\frac{1}{12n} - \frac{1}{(360 + \gamma_n)n^3} \right) \right| \\ &= \left| \frac{32n^5 + 80n^4 + 31n^3 - 75n^2 - 123n - 60}{60n(112n^5 + 360n^4 + 559n^3 + 411n^2 + 121n + 12)} \right|.\end{aligned}\quad (8)$$

Now let's do the same for the next two pairs of α and β from (4) and (5) respectively.

$$|\alpha_n^{(2)} - \beta_n^{(2)}| = \left| \frac{1}{12n} - \frac{1}{360n^3} - \left(\frac{1}{12n} - \frac{1}{360n(n+1)(n+2)} \right) \right| = \left| \frac{n^2 - (n+1)(n+2)}{360n^3(n+1)(n+2)} \right| = \frac{3n + 2}{360n^3(n+1)(n+2)},\quad (9)$$

$$|\alpha_n^{(3)} - \beta_n^{(3)}| = \left| \frac{1}{1260 \left(\frac{n^2 + n + 2}{n(n+1)} \right) n^5} - \frac{1}{1260n^5} \right| = \left| \frac{n^2 + n - (n^2 + n + 2)}{1260(n^2 + n + 2)n^5} \right| = \frac{2}{1260(n^2 + n + 2)n^5}.\quad (10)$$

As we consider $n \geq 2$, it's easy to see that α and β achieve their supremum with $n = 2$. So, to calculate the least possible error – let's substitute it into equations (8), (9), (10):

$$\begin{aligned}|\alpha_n^{(1)} - \beta_n^{(1)}| &= \frac{1946}{120 \cdot 15714} \approx 1.03e^{-3}, \\ |\alpha_n^{(2)} - \beta_n^{(2)}| &= \frac{3 \cdot 2 + 2}{360 \cdot 8 \cdot 3 \cdot 4} \approx 2.3e^{-4}, \\ |\alpha_n^{(3)} - \beta_n^{(3)}| &= \frac{2}{1260 \cdot (4 + 2 + 2) \cdot 32} \approx 6.2e^{-6}.\end{aligned}\quad (11)$$

As we convinced earlier, $\alpha_n^{(3)}$ and $\beta_n^{(3)}$ from (4) and (5) respectively give us the least error (11). That is why we will be using it further. Now let's assume that $n = 2$.

Then

$$\beta_n^{(3)} = \frac{1}{12 \cdot 2} - \frac{1}{360 \cdot 8} + \frac{1}{1260 \cdot 32} = \frac{1680 - 14 + 1}{40320} = \frac{1667}{40320}$$

and

$$\alpha_n^{(3)} = \frac{1}{12 \cdot 2} - \frac{1}{360 \cdot 8} + \frac{2 \cdot 3}{1260 \cdot (2 \cdot 3 + 2) \cdot 32} = \frac{13334}{322560} = \frac{6667}{161280}.\quad (12)$$

2. Improvements of inequality for norm of space of φ -sub-Gaussian random variables

φ -sub-Gaussian random variables, spaces, and their properties were introduced in (Giuliano-Antonini, Kozachenko, & Nikitina, 2003). We will try to improve the accuracy of the inequality for the norms of the space of φ -sub-Gaussian random variables, referring to the lemmas from there.

Lemma 1. Let $\xi \in \text{Sub}_\varphi(\Omega)$. Then for every $a > 0$ the next inequality holds true:

$$\mathbb{E} |\xi|^a \leq 2 \left(\frac{a}{e} \right)^a (\tau_\varphi(\xi))^a \inf_{t>0} \exp(\varphi(t - a \ln t)) \leq 2 (\tau_\varphi(\xi))^a \left(\frac{a}{\varphi^{-1}(a)} \right)^a.$$

Then, from 1, we can obtain the following result.

Corollary 1. Let $\xi \in Sub_{\varphi}(\Omega)$. Then the next inequality holds true

$$\tau_{\varphi}(\xi) \geq \theta_{\varphi}(\xi)/\sqrt{2},$$

where $\theta_{\varphi}(\xi) = \sup_{n \geq 2} E(|\xi|^n)^{\frac{1}{n}} \frac{\varphi^{-1}(n)}{n}$, and $\theta_{\varphi}(\xi)$ is a norm for $Sub_{\varphi}(\Omega)$.

Lemma 2. Let $\xi \in Sub_{\varphi}(\Omega)$. Then for $k = 1, 2, \dots$ the following inequalities hold true:

$$|E \xi^k| \leq E |\xi|^k \leq 2 (\tau_{\varphi}(\xi))^k \frac{e^k}{(\varphi^{-1}(k))^k} k!.$$

Corollary 2. If $\xi \in Sub_{\varphi}(\Omega)$, then the next inequality holds true

$$\tau_{\varphi}(\xi) \geq \frac{1}{e\sqrt{2}} \nu_{\varphi}(\xi),$$

where $\nu_{\varphi}(\xi) = \sup_{n \geq 2} |E \xi^n|^{\frac{1}{n}} \frac{\varphi^{-1}(n)}{(n!)^{\frac{1}{n}}}$.

Norms $\theta_{\varphi}(\cdot)$ and $\nu_{\varphi}(\cdot)$ are less popular than $\tau_{\varphi}(\cdot)$, yet they are used in applications of φ -sub-Gaussian processes, see, for example (Troshki, 2017).

Lemma 3. Let $\xi \in Sub_{\varphi}(\Omega)$ and $\nu_{\varphi}(\xi) \geq \theta_{\varphi}(\xi)$. Then the next inequality holds true:

$$\nu_{\varphi}(\xi) \leq e^{\frac{329227}{322560}} \theta_{\varphi}(\xi). \quad (13)$$

Proof. We can see that this inequality holds true according to proposition 1, lemma 3, and Stirling's formula (3), stating that $n! = n^n e^{-n} (2\pi n)^{1/2} e^{\theta_n}$ for $n \geq 2$, where $|\theta_n| \leq \frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{(1260(1 + \frac{2}{n(n+1)})n^5)}$. Using the value calculated in (12), we derive the following inequality

$$(n!)^{-\frac{1}{n}} = \frac{1}{n} \frac{e^{1+\frac{\theta_n}{n}}}{(2\pi n)^{\frac{1}{2n}}} \leq \frac{1}{n} e^{1+\frac{\theta_n}{n}} = \frac{1}{n} e^{\frac{329227}{322560}}.$$

□

Lemma 3 improves the following result from (Giuliano-Antonini, Kozachenko, & Nikitina, 2003, Corollary 4.4):

$$\nu_{\varphi}(\xi) \leq \exp \left\{ \frac{49}{48} \right\} \theta_{\varphi}(\xi).$$

Lemma 4. Let ξ be a random variable with $E \xi = 0$, and let φ be a N -function satisfying the condition Q . Let $\lambda_0 > 0$ be some number and

$$c_0 = \inf_{0 < |\lambda| \leq \lambda_0} \varphi(\lambda)/\lambda^2.$$

Let also $\nu_{\varphi}(\xi) < \infty$. Then $\xi \in Sub_{\varphi}(\Omega)$ and the following inequality holds:

$$\tau_{\varphi}(\xi) \leq S_{\varphi} \nu_{\varphi}(\xi), \quad (14)$$

where $S_{\varphi} = \max_{i=1,3} \gamma_i^{-1}$, and γ_1 is the root of the equation $\gamma = \lambda_0 \sqrt{c_0(1-\gamma)}$, γ_2 is the root of the equation $\gamma^3 - 2(1-\gamma) = 0$, and γ_3 is the root of the equation $\gamma = \varphi^{-1}(2) \sqrt{c_0(1-\gamma)}$.

The proof can be found in (Giuliano-Antonini, Kozachenko, & Nikitina, 2003).

Also, note that Lemmas 3 and 4 establish the equivalence of the norms $\nu_{\varphi}(\xi)$ and $\tau_{\varphi}(\xi)$. Furthermore, a random variable ξ belongs to $Sub_{\varphi}(\Omega)$ if and only if $E \xi = 0$ and $\nu_{\varphi}(\xi) < \infty$.

3. Orlicz spaces of exponential type

Let ψ be an arbitrary N -function. The Orlicz space generated by the N -function $U(x) = \{\exp \{\psi(x)\} - 1, x \in \mathbb{R}\}$, is called an Orlicz space of exponential type. We shall be interested in the Orlicz space of exponential type $Exp_{\varphi^*}(\Omega)$ generated by the Young-Fenchel transform φ^* of an N -function φ . The Luxemburg norm in $Exp_{\varphi^*}(\Omega)$ is denoted by σ_{φ^*} , i.e. for a random variable $\xi \in Exp_{\varphi^*}(\Omega)$ we have

$$\sigma_{\varphi^*}(\xi) = \inf \{a > 0: E \{\exp \varphi^*(\xi/a)\} \leq 2\}.$$

The space $Exp_{\varphi^*}(\Omega)$ is a Banach space with respect to the norm $\sigma_{\varphi^*}(\cdot)$.

Lemma 5. Let φ be an N -function and $\xi \in Exp_{\varphi^*}(\Omega)$. Then, for any $p \geq 1$, the following inequality holds:

$$(E |\xi|^p)^{\frac{1}{p}} \leq 2^{\frac{1}{p}} \frac{p}{\varphi^{-1}(p)} \sigma_{\varphi^*}(\xi). \quad (15)$$

The proof can be found in (Giuliano-Antonini, Kozachenko, & Nikitina, 2003).

As a consequence of (13), the following inequalities are also improved.

Lemma 6. Let φ be a N -function satisfying the condition Q , and let ξ be a random variable such that $\xi \in \text{Sub}_\varphi(\Omega)$ and $E\xi = 0$. Then

$$\tau_\varphi(\xi) \leq S_\varphi e^{\frac{329227}{322560}} \sigma_{\varphi^*}(\xi),$$

where S_φ is defined in (14).

Lemma 7. Let ξ be a random variable such that

$$P\{|\xi| \geq x\} \leq C \exp\left\{-\psi\left(\frac{x}{D}\right)\right\},$$

where $\psi(x)$ is an N -function. Then $\xi \in \text{Exp}_\psi(\Omega)$ and $\sigma_\psi(\xi) \leq (1 + C)D$.

Lemma 8. Let $\xi \in \text{Sub}_\varphi(\Omega)$; then $\xi \in \text{Exp}_{\varphi^*}(\Omega)$ and the following inequality holds:

$$\sigma_{\psi^*}(\xi) \leq 3\tau_\varphi(\xi). \quad (16)$$

Considering the lemmas, propositions, and equation (12) mentioned above, we can arrive at the following theorem.

Theorem 1. A random variable ξ belongs to the space $\text{Sub}_\varphi(\Omega)$ only if $E\xi = 0$ and there exist constants $C > 0$ and $D > 0$ such that

$$P\{|\xi| \geq x\} \leq C \exp\left\{-\varphi^*\left(\frac{x}{D}\right)\right\} \quad (17)$$

for any $x > 0$. If (17) holds, then

$$\tau_\varphi(\xi) \leq S_\varphi e^{\frac{329227}{322560}} (1 + C)D,$$

where S_φ is defined in (14).

Example. Consider the Orlicz N -function $\varphi_p(x) = \frac{|x|^p}{p}$, $1 < p < 2$, or, for $p > 2$, the piecewise definition $\varphi_p(x) = \frac{|x|^p}{p}$, $|x| \geq 1$, and $\varphi_p(x) = \frac{|x|^2}{p}$, $|x| < 1$, and examine inequality (16) in this particular setting.

Put $\tau_{\varphi_p}(\xi) = \tau$. Let's apply Lemma 1. Since for $s > 1$

$$\inf_{t \geq 1} \exp\{\varphi_p(t) - s \ln t\} = \left(\frac{e}{s}\right)^{\frac{s}{p}},$$

then

$$E|\xi|^s \leq 2 \left(\frac{s}{e}\right)^{\frac{s}{q}} \tau^s. \quad (18)$$

Since $\varphi_p^*(x) = \frac{|x|^q}{q}$ for the case $1 < p < 2$, and $\varphi_p^*(x) \leq \frac{|x|^q}{q}$ (as $q < 2$) we have by (18) for $a > 0$

$$E \exp\{\varphi^*(\xi/a)\} \leq E \exp\left\{\frac{|\xi|^q}{qa^q}\right\} = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{E|\xi|^k}{(qa^q)^k} \leq 1 + 2 \sum_{k=1}^{\infty} \left(\frac{k^k}{k!e^k}\right) \left(\frac{\tau}{a}\right)^q k. \quad (19)$$

It follows from Stirling's formula and (12) that

$$\frac{k^k}{k!e^k} \leq \frac{1}{\sqrt{2\pi}} e^{\frac{13334}{322560}}.$$

Therefore

$$E \exp\left\{\varphi^*\left(\frac{\xi}{a}\right)\right\} \leq 1 + \frac{2e^{\frac{13334}{322560}}}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \left(\frac{\tau}{a}\right)^{kq}.$$

The above geometric series converges if $\tau/a < 1$ and in such case its sum is equal to

$$1 + \frac{2e^{\frac{13334}{322560}}}{\sqrt{2\pi}} \frac{(\tau/a)^q}{1 - (\tau/a)^q}.$$

Put

$$L = \left(\frac{2e^{\frac{13334}{322560}}}{\sqrt{2\pi}} + 1\right)^{1/q},$$

then the last quantity in (19) is not greater than 2 if $a/\tau > L > 1$. This gives us the following result

$$\sigma_{\varphi^*}(\xi) \leq L\tau_{\varphi_p}(\xi).$$

Discussions and conclusion

By refining Stirling's formula through an optimal choice of remainder terms, we obtained sharper constants in the characterization inequality for norms of φ -sub-Gaussian random variables. This illustrates how classical analytic improvements translate into tighter probabilistic bounds, with potential applications to concentration inequalities and stochastic modeling.

Authors' contribution: Rostyslav Yamnenko — conceptualization and methodology, Dmytro Tykhonenko — analysis of sources, theoretical foundations of research, writing an original draft, Katalin Kuchinka — analysis of results, viewing and editing.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Beesack, P. (1969). Improvements of Stirling's formula by elementary methods. *Publikacije Elektrotehničkog fakulteta. Serija Matematika i fizika*, 274/301, 17–21.
- Buldygin, V. V., & Kozachenko, Y. V. (2000). *Metric characterization of random variables and random processes*. American Mathematical Society, Providence, RI. <https://doi.org/10.1090/mmono/188>
- Feller, W. (1967). A direct proof of Stirling's formula. *American Mathematical Monthly*, 74(10), 1223–1225. <https://doi.org/10.2307/2315671>
- Giuliano-Antonini, R., Kozachenko, Y. V., & Nikitina, T. (2003). Space of φ -sub-Gaussian random variables. *Rendiconti, Accademia Nazionale delle Scienze detta dei XL, Memorie di Matematica e Applicazioni*, XXVII(121°), 92–124.
- Hu, M., Lui, F., & Shi, X. (2006). A new asymptotic series for the Gamma function. *Journal of Computational and Applied Mathematics*, 195(1), 134–154. <https://doi.org/10.1016/j.cam.2005.03.081>
- Hummel, P. (1940). A note on Stirling's formula. *American Mathematical Monthly*, 47(2), 97–99. <https://doi.org/10.2307/2303360>
- Long, L. (2014). On Stirling's formula remainder. *Applied Mathematics and Computation*, 247, 494–500. <https://doi.org/10.1016/j.amc.2014.09.012>
- Macyz, J. (2021). On Stirling's formula. *Lietuvos matematikos rinkinys*, 47, 526–530. <https://doi.org/10.15388/LMR.2007.24256>
- Maria, A. (1965). A remark on Stirling's formula. *American Mathematical Monthly*, 72(10), 1096–1098. <https://doi.org/10.2307/2315957>
- Nanjundiah, T. (1965). Note on Stirling's formula. *American Mathematical Monthly*, 66(8), 701–703. <https://doi.org/10.2307/2309346>
- Schwartz, A. (1967). *Calculus and analytic geometry* (2nd ed.). Holt, Rinehart & Winston.
- Troshki, N. (2017). Upper bounds for supremums of the norms of the deviation between a homogeneous isotropic random field and its model. *Theory of Probability and Mathematical Statistics*, 94, 159–184. <https://doi.org/10.2307/2309346>
- Uspensky, J. (1937). *Introduction to mathematical probability*. McGraw Hill.
- Vasylyk, O., Kozachenko, Y., & Yamnenko, R. (2008). φ -sub-Gaussian random processes. VPC "Kyivskyi Universytet" [in Ukrainian]. [Василик, О., Козаченко, Ю., & Ямненко, Р. (2008). φ -субгауссові випадкові процеси. ВПЦ "Київський університет"].

Отримано редакцією журналу / Received: 10.07.25
Прорецензовано / Revised: 12.09.25
Схвалено до друку / Accepted: 10.10.25

Дмитро ТИХОНЕНКО, асп.
ORCID ID: 0009-0002-1144-6852
e-mail: tykhonenko.dmytro@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

Ростислав ЯМНЕНКО, д-р фіз.-мат. наук
ORCID ID: 0000-0002-9612-7959
e-mail: rostyslav.yamnenko@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

Каталін КУЧІНКА, канд. фіз.-мат. наук
ORCID ID: 0000-0001-8129-9937
e-mail: kucsinka.katalin@unithe.hu
Закарпатський угорський університет імені Ференца Ракоці II, Берегове, Україна; Токайський університет, Шарошпатак, Угорщина

УТОЧНЕННЯ ХАРАКТЕРИЗАЦІЙНОЇ НЕРІВНОСТІ ДЛЯ НОРМ φ -СУБГАУССОВОЇ ВИПАДКОВОЇ ВЕЛИЧИНИ

Здійснено уточнення характеристичної нерівності між нормами φ -субгауссової випадкової величини. Суть цього вдосконалення полягає в більш детальному аналізі залишкових членів у формулі Стірлінга, що використовується для підрахунку наближеного значення для $n!$. У запропонованому дослідженні проведено порівняльний аналіз декількох пар відомих залишкових членів, α_n і β_n , щоб визначити пару, яка мінімізує похибку наближення. Обчислюючи абсолютну різницю для цих пар, продемонстровано, що найтісніші межі досягаються у разі конкретного вибору членів вищого порядку. Це поліпшене наближення для $n!$ потім застосовується для уточнення сталої в характеристичній нерівності для порівняння різних норм φ -субгауссової випадкової величини та теоремі про φ -субгауссовість випадкової величини.

Ключові слова: формула Стірлінга, залишок, характеристична нерівність, норма, φ -субгауссова випадкова величина.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

ДИФЕРЕНЦІАЛЬНІ РІВНЯННЯ, МАТЕМАТИЧНА ФІЗИКА ТА МЕХАНІКА

UDC 539.3

DOI: <https://doi.org/10.17721/1812-5409.2025/2.11>

Natalya VAYSFELD, DSc (Phys. & Math.), Senior Lecturer
ORCID ID: 0000-0001-8082-2503
e-mail: natalya.vaysfeld@kcl.ac.uk
King's College London, London, United Kingdom

Zinaida ZHURAVLOVA, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-3271-8864
e-mail: z.zhuravlova@onu.edu.ua
Odesa I. I. Mechnikov National University, Odesa, Ukraine

AXISYMMETRIC POROELASTICITY PROBLEM FOR A MULTILAYERED CYLINDER

Poroelastic materials are the object of close attention of researchers due to their wide representation both in the natural environment (geological formations, biological tissues) and in technical applications (engineering structures, filtration systems). Among them, cylindrical structures are of particular importance, often characterised by radial layering and heterogeneity. The study of the stress-strain state of such objects is of significant practical importance for calculating their strength, stability, and efficient operation. Despite the considerable amount of scientific work, most of it is based on numerical modelling. At the same time, it is analytical methods that allow us to gain a deeper understanding of physical laws, identify limit regimes, and verify numerical approaches.

In this paper, we consider an analytical solution of an axisymmetric problem for a finite-length poroelastic cylinder with a radially layered structure. A load is applied to the cylindrical surface, which can be either a mechanical or a pressure load. The top and bottom edges of the cylinder are in smooth contact and are impermeable. Perfect mechanical contact is maintained between the layers. The application of the Fourier integral transform method allows us to reduce the original problem to a one-dimensional vector boundary value problem, the general solution of which is found using the matrix differential calculation. The method of recurrence relations is used to find the unknown constants of each layer of the cylinder. As a result, an exact solution was derived, which allows us to study the distribution of normal stress and pore pressure depending on the applied load, geometric characteristics, and physical and mechanical properties of the layers. The obtained results are important for the further development of the analytical mechanics of poroelastic media

Key words: axisymmetric problem, poroelasticity, multilayered cylinder, integral Fourier transform, matrix differential calculation, recurrent correspondences.

AMS 2020 classification: 35Q74, 42A38, 74G05.

Introduction

Relevance of research. The investigation of the stress-strain state of layered finite poroelastic cylinders is an important scientific and practical task due to the growing requirements for the reliability and efficiency of structures made of poroelastic materials. Such cylindrical elements are widely used in engineering practice. The presence of a layered structure along the radius allows for structural optimisation of the distribution of mechanical characteristics, which provides increased resistance to external loads. Despite the existence of a significant number of studies on the mechanics of poroelastic bodies, the issue of stress and strain distribution in finite layered poroelastic cylinders, taking into account boundary conditions and interlayer interaction, remains insufficiently studied. The construction of mathematical models for such objects, as well as the analytical solution of the corresponding boundary value problems, is important for the development of the theory of poroelasticity and the practical application of its results in the design of advanced materials and structures.

The object of research is a poroelastic layered solid cylinder with ideal contact conditions between the layers under the action of a load applied along a cylindrical surface.

The aim and objectives of the research are to construct an analytical solution of the poroelasticity problem for a multilayered solid cylinder with ideal contact conditions between the layers under the action of a load applied along a cylindrical surface, as well as to construct and analyse graphs of the distribution of normal stress and fluid pressure

In recent years, many scientific papers have been published on numerical methods for finding solutions and studying important characteristics of multilayer poroelastic cylinders (Eldeeb, Shabana, & Elsawaf, 2021; Bociu et al., 2021). Some studies combine analytical and numerical approaches (Naccache et al., 2022; Gnadjro, d'Almeida, & Franklin, 2024; Grigorenko, Y., Grigorenko, O., & Rozhok, 2022). At the same time, only analytical methods allow us to study important qualitative properties of solutions (Gohari, Zarastvand, & Talebitooti, 2020; Kubenko et al., 2023; Zhupanska, & Ulitko, 2005; Meleshko, & Tokovyy, 2012). In this paper, we present an analytical approach that allows us to find an exact solution of the axisymmetric poroelasticity problem for a multilayered solid cylinder.

1. Statement of the problem

The finite solid poroelastic cylinder $0 < r < 1, 0 < z < h, -\pi < \varphi < \pi$, which consists of N layers $a_{i-1} < r < a_i, i = \overline{1, N}, a_0 = 0, a_N = 1$ is considered. The ideal contact conditions are fulfilled between the layers (Cheng, 2016)

$$u_l|_{r=a_l-0} = u_{l+1}|_{r=a_l+0}, w_l|_{r=a_l-0} = w_{l+1}|_{r=a_l+0}, p_l|_{r=a_l-0} = p_{l+1}|_{r=a_l+0}, \sigma_r^l|_{r=a_l-0} = \sigma_r^{l+1}|_{r=a_l+0}, \\ \tau_{rz}^l|_{r=a_l-0} = \tau_{rz}^{l+1}|_{r=a_l+0}, k_l \frac{\partial p_l}{\partial r}|_{r=a_l-0} = k_{l+1} \frac{\partial p_{l+1}}{\partial r}|_{r=a_l+0}, l = \overline{1, N-1},$$

where $u_l(r, z), w_l(r, z)$ are dimensionless displacements of l -th layer of solid skeleton regarding axes r, z respectively, $p_l(r, z)$ is dimensionless pore pressure of l -th layer, $\sigma_r^l(r, z), \tau_{rz}^l(r, z)$ are dimensionless normal and tangential stress of l -th layer correspondingly. At the cylindrical surface $r = 1$ the loading is applied $\sigma_r^N|_{r=1} = -L(z) - \alpha_N P(z), \tau_{rz}^N|_{r=1} = T(z), p_N|_{r=1} = P(z)$, where $L(z), T(z), P(z)$ are known functions, α_l is Biot's coefficient of l -th layer. The edges $z = 0, z = h$ are under slide contact conditions, and they are impermeable $w_l|_{z=0,h} = 0, \tau_{rz}^l|_{z=0,h} = 0, \frac{\partial p_l}{\partial z}|_{z=0,h} = 0$. The displacements and pore pressure

© Vaysfeld Natalya, Zhuravlova Zinaida, 2025

should satisfy the system of equations (Verruijt, 2010). The stress-deformable state of the multilayered cylinder with indicated boundary and transmission conditions should be found.

2. The solving of the boundary value problem

The original problem is reduced to the one-dimensional boundary value problem with the help of finite sin-, cos- Fourier transform regarding variable z (Vaysfeld, & Zhuravlova, 2023). The derived one-dimensional problem in transform domain is written in a vector form

$$\{L_2 \vec{y}_{l,\beta}(r) = 0, a_{l-1} < r < a_l, A_\beta \vec{y}'_{N,\beta}(1) + B_\beta \vec{y}_{N,\beta}(1) = \vec{g}_\beta, \vec{y}_{l,\beta}|_{r=a_{l-0}} = \vec{y}_{l+1,\beta}|_{r=a_{l+0}}, (S_{l,\beta} \vec{y}'_{l,\beta} + T_{l,\beta} \vec{y}_{l,\beta})|_{r=a_{l-0}} = (S_{l+1,\beta} \vec{y}'_{l+1,\beta} + T_{l+1,\beta} \vec{y}_{l+1,\beta})|_{r=a_{l+0}}\} \quad (1)$$

where differential operator

$$L_2 = \left(\frac{1}{r} \frac{d}{dr} \left(r \frac{d}{dr} \right) - \frac{1}{r^2} - \frac{\kappa_l - 1}{\kappa_l + 1} \beta^2 \right) \frac{2\beta}{\kappa_l + 1} \frac{d}{dr} - \alpha_l \frac{\kappa_l - 1}{\kappa_l + 1} \frac{d}{dr} - \frac{2\beta}{\kappa_l - 1} \frac{1}{r} \frac{d}{dr} \left(r \frac{d}{dr} \right) - \frac{\kappa_l + 1}{\kappa_l - 1} \beta^2 \alpha_l \beta - \frac{\alpha_l}{\kappa_l} \frac{1}{r} \frac{d}{dr} \left(r \frac{d}{dr} \right) - \frac{\alpha_l \beta}{\kappa_l} \frac{1}{r} \frac{d}{dr} \left(r \frac{d}{dr} \right) - \beta^2 - \frac{S_{pl}}{\kappa_l},$$

vectors $\vec{y}_{l,\beta}(r) = (u_{l,\beta}(r), w_{l,\beta}(r), p_{l,\beta}(r))^T$, $\vec{g}_\beta = \left(\frac{\kappa_N - 1}{2} (\alpha_N P_\beta - L_\beta), T_\beta, P_\beta \right)^T$, matrices

$$A_\beta = (\kappa_N + 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0), B_\beta = (3 - \kappa_N \ (3 - \kappa_N) \beta \ 0 \ -\beta \ 0 \ 0 \ 0 \ 0 \ 1), S_{l,\beta} = \left(\frac{\kappa_l + 1}{\kappa_l - 1} G_l \ 0 \ 0 \ 0 \ G_l \ 0 \ 0 \ 0 \ K_l \right), T_{l,\beta} = \left(\frac{3 - \kappa_l}{\kappa_l - 1} \frac{G_l}{\alpha_l} \frac{3 - \kappa_l}{\kappa_l - 1} \beta G_l \ 0 \ -\beta G_l \ 0 \ 0 \ 0 \ 0 \ 0 \right),$$

$\kappa_l = 3 - 4\mu_l$ is Muskhelishvili's constant, μ_l is Poisson ratio, S_{pl}, K_l – are dimensionless values of storativity of the pore space S_{pl} and permeability coefficient k_l of l -th layer respectively.

The apparatus of matrix differential calculation (Gantmacher, 1959) is applied for the find of general solution of the boundary value problem (1). According to it, the corresponding matrix differential equation $L_2 Y_{l,\beta}(r) = 0$ is considered. The following correspondences take place (Popov, & Protserov, 2016) $L_2 Q_i(r, \xi) = -Q_i(r, \xi) M_{l,\beta}(\xi)$, $i = 1, 2$, where

$$Q_1(r, \xi) = (J_1(\xi r) \ 0 \ 0 \ 0 \ J_0(\xi r) \ 0 \ 0 \ 0 \ J_0(\xi r)), Q_2(r, \xi) = (N_1(\xi r) \ 0 \ 0 \ 0 \ N_0(\xi r) \ 0 \ 0 \ 0 \ N_0(\xi r)),$$

$J_1(\xi r), J_0(\xi r)$ are Bessel functions, $N_1(\xi r), N_0(\xi r)$ are Neumann functions, $M_{l,\beta}(\xi)$ is known matrix. The solution of the matrix equation is found by the formula $Y_{l,\beta,j}(r) = \frac{1}{2\pi i} \oint_C Q_j(r, \xi) M_{l,\beta}^{-1}(\xi) d\xi$, $j = 1, 2$, where closed contour C covers all singularity points of the integrand. The general solution of the vector boundary value problem (1) has the form $\vec{y}_{l,\beta}(r) = Y_{l,\beta}(r) \vec{C}_l$, where

$$Y_{l,\beta}(r) = \{Y_{1,\beta,1}(r), l = 1, (Y_{l,\beta,1}(r), Y_{l,\beta,2}(r)), 2 \leq l \leq N, \vec{C}_l = \{(c_{1,1}, c_{1,2}, c_{1,3})^T, l = 1, (c_{l,1}, c_{l,2}, c_{l,3}, c_{l,4}, c_{l,5}, c_{l,6})^T, 2 \leq l \leq N, H_{l,\beta}(r) = (Y_{l,\beta}(r) S_{l,\beta} Y'_{l,\beta}(r) + T_{l,\beta} Y_{l,\beta}(r))\}.$$

The transmission conditions in (1) can be written in the following matrix form $H_{l,\beta}(a_l) \vec{C}_l = H_{l+1,\beta}(a_l) \vec{C}_{l+1}$, $l = \overline{1, N-1}$, from which, according to (Popov, 2013) the constants of each layer $\vec{C}_l$ are recurrently expressed through the constants of the first layer $\vec{C}_1$:

$$\vec{C}_l = R_{l,\beta} \vec{C}_1 = H_{l,\beta}^{-1}(a_{l-1}) H_{l-1,\beta}(a_{l-1}) \dots H_{2,\beta}^{-1}(a_1) H_{1,\beta}(a_1) \vec{C}_1, l = \overline{2, N}. \quad (2)$$

Substituting (2) in the boundary condition in (1), we derive $\vec{C}_1 = \left[(A_\beta Y'_{N,\beta}(1) + B_\beta Y_{N,\beta}(1)) R_{N,\beta} \vec{C}_1 \right]^{-1} \vec{g}_\beta$. The case when $\beta = 0$ is solved analogically.

3. The results of the numerical analysis

The derived exact solution of the original problem allows to conduct the numerical investigation of normal stress and pore pressure change inside the cylinder. This analysis was carried out for the applied concentrated mechanical loading $L(z) = \delta(z - \frac{h}{2})$, $T(z) = 0, P(z) = 0$ ($\delta(z)$ is Dirac delta function). Geometrical sizes of the cylinder were chosen as: $a_1 = 0.5 \text{ m}, h = 1 \text{ m}$. The two-layered cylinder with the materials Ohio sandstone for the inner layer and Boise sandstone for the outer layer (Cheng, 2016) was studied.

At Fig. 1, the normal stress and pore pressure are symmetrical, which is explained by the same boundary conditions at the edges $z = 0, z = h$ and the fact that the load is applied in the middle of the cylinder. The normal stress is negative, and no tensile stress is observed. The maximum is reached at the point of load application. The highest values are seen closer to the loaded face $r = 1$. The pore pressure is nonzero, which is due to fluid leakage on the cylindrical surface, which also affects the compression of the cylinder.

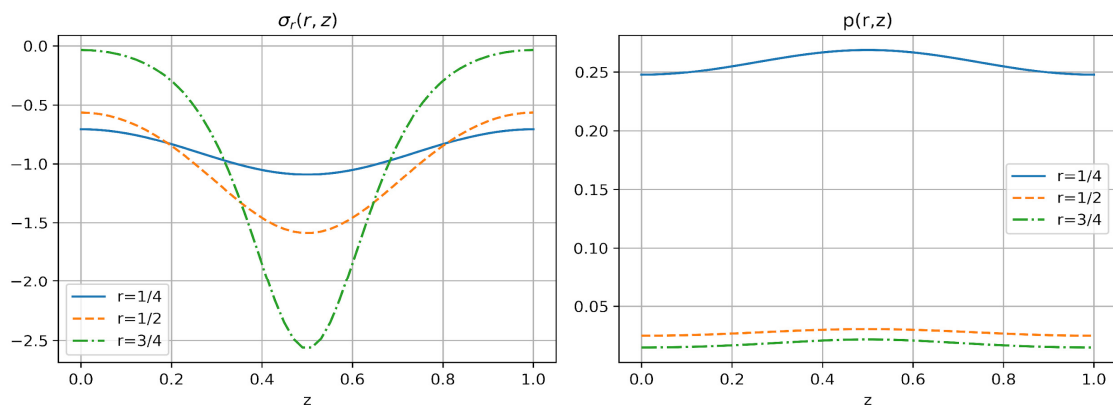


Fig. 1. Normal stress and pore pressure inside the cylinder

As can be seen at Fig. 2, an increase in the thickness of the inner layer of the cylinder increases the pore pressure inside this layer. This is due to the fact that the inner layer of the cylinder (Ohio sandstone) has a higher permeability coefficient. Changing the height of the cylinder also has a significant effect on the pore pressure. Thus, the highest pore pressure values are observed for a longer cylinder. This can be explained by the fact that the outer surface area through which the fluid can flow increases.

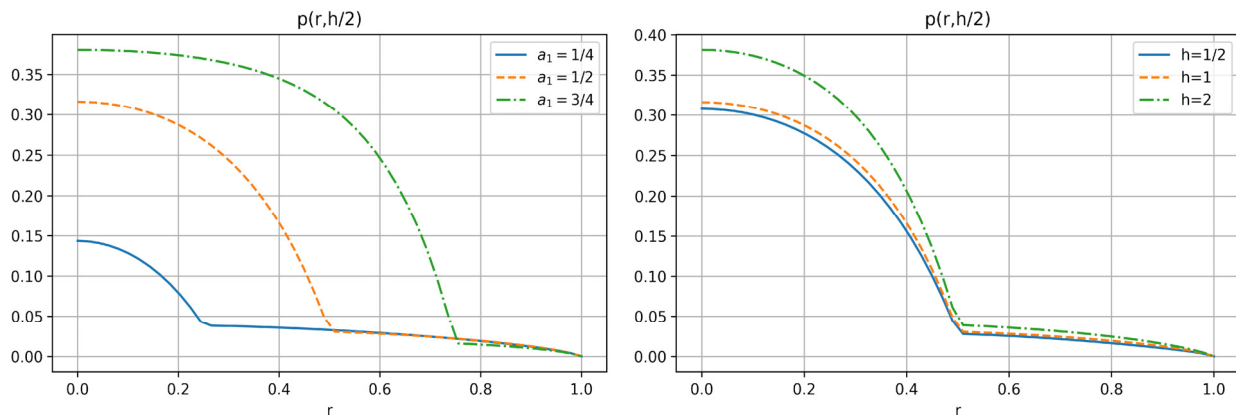


Fig. 2. The change of pore pressure regarding the change of thickness of the inner layer (left) and the change of the cylinder's height (right)

Discussion and conclusions

In the present study, an exact analytical solution of the axisymmetric problem for a finite multilayered poroelastic cylinder was obtained using the methods of integral transforms, matrix differential calculation, and recurrence relations. The numerical analysis revealed the nature of changes in normal stress and pore pressure depending on geometric parameters.

The highest absolute values of normal stress and pore pressure were recorded at the point of concentrated mechanical loading's application, which indicates a localised increase in the interaction between the solid matrix and the pore medium. It is established that an increase in the thickness of the inner layer causes an increase in the pore pressure in this area, while an increase in the height of the cylinder generally contributes to an increase in pore pressure along its entire length.

The proposed methodology demonstrates high versatility and can be adapted to analyse cylindrical structures with more complex conditions of interlayer communication, which opens up prospects for its further application in engineering practice and biomechanics.

Authors' contribution: Natalya Vaysfeld – preparation of theoretical foundations of research; conceptualization; Zinaida Zhuravlova – methodology; software.

Acknowledgments. N. Vaysfeld acknowledges support from the Royal Society Wolfson Visiting Fellowship R3/233003.

Sources of funding. The research is supported by European project funded by Horizon 2020 Framework Programme for Research and Innovation (2014–2020) (H2020-MSCA-RISE-2020) Grant Agreement number 101008140 EffectFact "Effective Factorisation techniques for matrix-functions: Developing theory, numerical methods and impactful applications".

References

- Bociu, L., Canic, S., Muha, B., & Webster, J. T. (2021). Multilayered Poroelasticity Interacting with Stokes Flow. *SIAM Journal on Mathematical Analysis*, 53(6), 6243–6279. <https://doi.org/10.1137/20M1382520>
- Cheng, A. H.-D. (2016). *Poroelasticity. Theory and applications of transport in porous media*. Springer.
- Eldeeb, A., Shabana, Y., & Elsawaf, A. (2021). Investigation of the thermoelastoplastic behaviors of multilayer FGM cylinders. *Composite Structures*, 276, 114523. <https://doi.org/10.1016/J.COMPSTRUCT.2021.114523>
- Gantmacher, F. R. (1959). *The theory of matrices*. Chelsea Publishing Company.
- Gnadjro, D., d'Almeida, A., & Franklin, H. (2024). Effective properties of randomly distributed poroelastic cylinders in a poroelastic matrix. *Journal of Sound and Vibration*, 593, 118670. <https://doi.org/10.1016/j.jsv.2024.118670>
- Gohari, H., Zarastvand, M., & Talebitooti, R. (2020). Acoustic performance prediction of a multilayered finite cylinder equipped with porous foam media. *Journal of Vibration and Control*, 26, 899–912. <https://doi.org/10.1177/1077546319890025>
- Grigorenko, Y. M., Grigorenko, O. Y., & Rozhok, L. S. (2022). Stress State of Non-Thin Nearly Circular Cylindrical Shells Made of Continuously Inhomogeneous Materials. *International Applied Mechanics*, 58, 381–388. <https://doi.org/10.1007/s10778-022-01163-0>
- Kubenko, V. D., Yanchevskiy, I. V., Zhuk, Ya. O., & Liskin, V. O. (2023). Hydrodynamic Characteristics of Interaction of the Plane Wave with a Spherical Body in the Filled by Compressive Liquid Semi-infinite Circular Cylindrical Cavity. *International Applied Mechanics*, 59(2), 3–18. <https://doi.org/10.1007/s10778-023-01207-z>
- Meleshko, V., & Tokovyy, Yu. (2013). Equilibrium of an elastic finite cylinder under axisymmetric discontinuous normal loadings. *Journal of Engineering Mathematics*, 78, 143–166. <https://doi.org/10.1007/s10665-011-9524-y>
- Naccache, F., Hamdi, M., Chazot, J., & Anciant, M. (2022). Semi-analytical model of noise transmission through finite length cylinder with poroelastic core using Biot's and Shell theories. *Journal of Sound and Vibration*, 535, 117102. <https://doi.org/10.1016/j.jsv.2022.117102>
- Popov, G. Ya. (2013). *Exact solutions of some boundary value problems of deformable solid mechanic*. Astroprint.
- Popov, G. Ya., & Protserov, Yu. S. (2016). Axisymmetric problem for an elastic cylinder of finite length with fixed lateral surface with regard for its weight. *Journal of Mathematical Sciences*, 212(1), 67–82.
- Vaysfeld, N. D., & Zhuravlova, Z. Yu. (2023). Investigation of porous stress state of semi-infinite hollow cylinder under axisymmetric loading. *Bulletin of the Taras Shevchenko National University of Kyiv. Physics and Mathematics*, 2023(2), 92–95. <https://doi.org/10.17721/1812-5409.2023/2.11>

Verruijt, A. (2010). *An introduction to soil dynamics. Theory and applications of transport in porous media*. Springer.

Zhupanska, O. I., & Uliitko, A. F. (2005). Contact with friction of a rigid cylinder with an elastic half-space. *Journal of the Mechanics and Physics of Solids*, 53(5), 975–999.

Отримано редакцією журналу / Received: 28.06.25

Прорецензовано / Revised: 06.10.25

Схвалено до друку / Accepted: 10.10.25

Наталя ВАЙСФЕЛЬД, д-р фіз.-мат. наук, проф.

ORCID ID: 0000-0001-8082-2503

e-mail: natalya.vaaysfeld@kcl.ac.uk

Королівський коледж у м. Лондоні, Лондон, Об'єднане Королівство

Зінаїда ЖУРАВЛЮВА, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0002-3271-8864

e-mail: z.zhuravlova@onu.edu.ua

Одеський національний університет імені І. І. Мечникова, Одеса, Україна

ОСЕСИМЕТРИЧНА ЗАДАЧА ПОРОПРУЖНОСТІ ДЛЯ ШАРУВАТОГО ЦИЛІНДРУ

Поропружні матеріали є об'єктом пильної уваги дослідників завдяки широкій представленості як у природному середовищі (геологічні формації, біологічні тканини), так і в технічних застосуваннях (інженерні конструкції, фільтраційні системи). Серед них особливої ваги набувають циліндричні структури, що нерідко характеризуються радіальною шаруватістю та неоднорідністю. Вивчення напружено-деформованого стану таких об'єктів має суттєве практичне значення для розрахунку їхньої міцності, стійкості й ефективного функціонування. Незважаючи на значний обсяг наукових праць, більшість з них базується на чисельному моделюванні. Водночас саме аналітичні методи дають змогу отримати глибше розуміння фізичних закономірностей, виявити граничні режими та забезпечити перевірку чисельних підходів.

У пропонованій роботі розглянуто аналітичне розв'язання осесиметричної задачі для скінченного за довжиною поропружного циліндра з радіально-шаруватою структурою. До циліндричної поверхні прикладено навантаження, що може бути як механічним, так і навантаженням тиску. Верхній і нижній торці циліндру перебувають в умовах гладкого контакту та є непроникними. Між шарами виконуються умови ідеального механічного контакту. Застосування методу інтегрального перетворення Фур'є надає можливість звести вихідну задачу до одновимірної векторної крайової задачі, загальний розв'язок якої отримано за допомогою апарату матричного диференціального числення. Для знаходження невідомих сталих кожного шару циліндру використано метод рекурентних співвідношень. У підсумку отримано точний розв'язок, що дає змогу дослідити розподіл нормальних напружень і тиску рідини залежно від прикладеного навантаження, геометричних характеристик і фізико-механічних властивостей шарів. Одержані результати є важливими для подальшого розвитку аналітичної механіки поропружних середовищ.

Ключові слова: осесиметрична задача, поропружність, шаруватий циліндр, інтегральне перетворення Фур'є, матричне диференціальне числення, рекурентні співвідношення.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 534

DOI: <https://doi.org/10.17721/1812-5409.2025/2.12>

Віктор ГРІНЧЕНКО¹, д-р фіз.-мат. наук, акад.
ORCID ID: 0000-0003-3229-1810
e-mail: grinchenko@nas.gov.ua

Ірина ЛЕБЕДЄВА², канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0001-7150-1310
e-mail: lebedevaiv@knu.ua

Володимир МАЦИПУРА², д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0002-0136-6659
e-mail: mnivtt@gmail.com

Тихомир ТРИФОНОВ³, канд. фіз.-мат. наук, проф.
ORCID ID: 0000-0002-9649-4396
e-mail: TihomirTrifonov@ieee.org

¹Інститут гідромеханіки НАН України, Київ, Україна

²Київський національний університет імені Тараса Шевченка, Київ, Україна

³Великотирновський університет імені святих Кирила і Мефодія, Велико Тирново, Болгарія

ВПЛИВ ЗГЛАДЖУВАННЯ ПОВЕРХНІ СКІНЧЕННОГО КЛИНОПОДІБНОГО ОБ'ЄКТА НА ХАРАКТЕРИСТИКИ РОЗСІЮВАННЯ ПЛОСКОЇ ХВИЛІ. ЧАСТИНА II. АНАЛІЗ ЧИСЕЛЬНИХ РЕЗУЛЬТАТІВ

Пропонована стаття є другою частиною роботи (Грінченко, Лебедєва, Маципура, 2024), що присвячена задачі розсіювання плоскої гармонічної хвилі на скінченному клиноподібному об'єкті зі згладженою поверхнею (іншими словами, із заокругленням). Такий об'єкт являє собою спрощену модель профілю крила літака. Для цієї моделі, базуючись на розв'язку, отриманому в межах методу часткових областей, визначено вираз для зворотного перерізу розсіювання. Наведено приклад виконання граничних умов на межі часткових областей. Побудовано графіки зворотного перерізу розсіювання для різного радіуса заокруглення передньої кромки профілю моделі і різної провідності поверхні заокруглення. Показано, що зворотний переріз розсіювання істотно залежить від радіуса заокруглення профілю і нормальної провідності поверхні заокруглення. При цьому покриття поверхні згладжування поглинаючим матеріалом істотно зменшує, у певному діапазоні кутів падіння хвилі, розсіяне поле об'єкта.

Побудований чисельно-аналітичний розв'язок дає змогу здійснювати оціночні розрахунки зворотного перерізу розсіювання об'єкта у разі зміни в широкому діапазоні його геометричних і фізичних параметрів.

Ключові слова: зворотний переріз розсіювання, метод часткових областей, плоска хвиля, поточкове спряження полів.

Класифікація відповідно до AMS 2020: 76Q05.

Вступ

Задача про дифракцію хвиль на нескінченному або скінченному клині є фундаментальною проблемою, на результатах розв'язку якої базуються багато фізичних уявлень про процеси взаємодії хвиль із тілами. Цій проблемі присвячено велику кількість джерел (Daniele, & Lombardi, 2020; Felsen, & Marcuvitz, 1973; Sommerfeld, 2004; Grinchenko, Vovk, & Matsypura, 2018), у яких досліджують як нескінчений клин з ідеальними (Monzon, Forester, & Loschialpo, 2005) та діелектричними поверхнями (Emig, 2024; Grzesik, 2019), так і скінчений клин із різною геометрією та властивостями поверхонь (Grinchenko, Vovk, & Matsypura, 2018). Варто відзначити великий інтерес до проблеми розсіювання електромагнітних хвиль на клиноподібних об'єктах у фахівців у галузі електродинаміки (Emig, 2024; Grzesik, 2019; Nacivelioglu, Sevgi, & Ufimtsev, 2011). Загалом підвищений інтерес до процесу дифракції хвилі на клині зумовлений його глибоким фізичним змістом і важливим практичним значенням.

Пропонована стаття є другою частиною роботи (Грінченко, Лебедєва, & Маципура, 2024), що присвячена задачі розсіювання плоскої гармонічної хвилі на скінченному клиноподібному об'єкті зі згладженою поверхнею (іншими словами, із заокругленням) (рис. 1).

Такий об'єкт являє собою спрощену модель профілю крила літака. У задачах розсіювання хвиль найважливішими характеристиками є перерізи розсіювання об'єкта. Ці енергетичні характеристики визначаються в далекому полі розсіювання.

Спираючись на результати роботи (Грінченко, Лебедєва, & Маципура, 2024), отримуємо вираз для зворотного перерізу розсіювання:

$$\sigma_L(\theta_0) = \frac{4}{k} L(\theta_0) L^*(\theta_0), \quad (1)$$

де

$$\begin{aligned} L(\theta) &= \sum_{n=0}^{\infty} \frac{A_n}{H_n^{(1)}(ka)} \cos(n\theta) \exp\left(-in\frac{\pi}{2}\right) + \sum_{n=1}^{\infty} \frac{B_n}{H_n^{(1)}(ka)} \sin(n\theta) \exp\left(-in\frac{\pi}{2}\right) + \\ &+ \sum_{n=0}^{\infty} \frac{C_n}{H_n^{(1)}(kR)} \cos(n\theta) \exp\left(ikd \cos \theta - in\frac{\pi}{2}\right) + \sum_{n=1}^{\infty} \frac{D_n}{H_n^{(1)}(kR)} \sin(n\theta) \exp\left(ikd \cos \theta - in\frac{\pi}{2}\right). \end{aligned} \quad (2)$$

© Грінченко Віктор, Лебедєва Ірина, Маципура Володимир, Трифонов Тихомир, 2025

ISSN 1812-5409 (Print), ISSN 2218-2055 (Online)

Одиниця вимірювання зворотного перерізу розсіювання для плоскої задачі – метр.

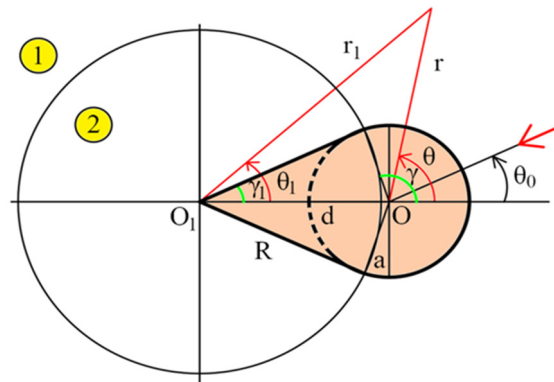


Рис. 1. Геометрія моделі

Щоб зробити вираз (1) безрозмірним, його ділять на мідель об'єкта. Міделевий переріз (мідель) – це площа поперечного перерізу зони геометричної тіні об'єкта. Мідель залежить від кута падіння хвилі θ_0 на об'єкт дослідження, тому будемо ділити вираз (1) на максимальне значення міделя M_{max} . Для нашої плоскої задачі (рис. 1) величина $M_{max} = d + a$.

1. Побудова чисельного алгоритму і перевірка виконання граничних умов

Необхідно зауважити, що побудовані в межах методу часткових областей аналітичні уявлення звукового поля завжди точно задовольняють рівняння Гельмгольца за будь-якої кількості складових, що утримуються в рядах. З огляду на це в основу оцінки точності розв'язку задачі варто покласти оцінку точності виконання умов спряження полів на межах часткових областей. Очевидно, що для підвищення точності оцінки звукових полів необхідно збільшувати кількість утримуваних невідомих комплексних коефіцієнтів під час розв'язку алгебраїчної системи рівнянь методом редуцції. Особливо це важливо за високих частот. Чисельний експеримент дає змогу оцінити ступінь виконання умов спряження звукових полів для вибраних параметрів моделі. Така процедура є стандартною у разі використання методу часткових областей і не раз апробована нами (Grinchenko, Vovk, & Matsypura, 2018). Згідно з міркуваннями, зазначеними у першій частині роботи (Грінченко, Лебедева, & Маципура, 2024), будемо використовувати поточкове спряження полів на межі часткових областей.

Отже, обмежимо нескінченні ряди у формулі (2), задаючи кількість коефіцієнтів A_n , B_n рівним N_1 , коефіцієнтів C_n , D_n – N_2 , коефіцієнтів E_n , F_n – N_3 (Грінченко, Лебедева, & Маципура, 2024). Відправною точкою в розрахунках за поточкового спряження полів є нормована до довжини хвилі λ відстань Δ між вузловими точками, тобто величина Δ/λ . Згідно з величиною Δ/λ , встановлюємо кількість вузлових точок: $N_1 = [2\pi a/\Delta]$, $N_2 = [2\pi R/\Delta]$, $N_3 = [2(\pi - \gamma_1)R/\Delta]$. Наведемо приклад виконання граничних умов. Згідно з рис. 1 задаємо геометричні параметри об'єкта a і R , коефіцієнт відбиття згладженої поверхні V і кут падіння плоскої хвилі θ_0 . Визначаємо також такі геометричні параметри:

$$d = OO_1 = \sqrt{R^2 + a^2}, \quad \gamma = \pi - \arccos(a/d), \quad \gamma_1 = \pi/2 - \arccos(a/d). \quad (3)$$

Оскільки для нас цікавою є модель профілю крила у вигляді гострого клиноподібного об'єкта з наявністю плоских поверхонь скінченного клину великого хвильового розміру, то оберемо такі геометричні параметри моделі: $\frac{R}{\lambda} = 10,2$, $a/\lambda = 0,8$. Нехай $V = 1$, тоді відносна нормальна провідність згладженої поверхні $\tilde{Y} = 0$. Згідно з формулами (3), кути $\gamma \approx 94,5^\circ$ і $\gamma_1 \approx 4,5^\circ$, величина $d/\lambda \approx 10,23$. Нехай кут падіння плоскої хвилі $\theta_0 = 90^\circ$. Функції $v_1(a, \theta)$ і $v_2(R, \theta_1)$ у граничних умовах (див. формули (8) і (10) в першій частині роботи (Грінченко, Лебедева, & Маципура, 2024)) покладемо рівними нулю.

Виберемо відстань між вузловими точками на колі радіуса a , що дорівнює $\Delta/\lambda = 0,05$, а на колі радіуса R – яка дорівнює $\Delta/\lambda = 0,15$. Тоді кількість вузлових точок і, відповідно, кількість мод, що враховуються в часткових областях, визначиться величинами $N_1 = 102$, $N_2 = 214$, $N_3 = 209$.

Графіки на рис. 2 ілюструють виконання умов спряження полів на межі часткових областей 1 і 2, тобто на колі радіусом R за зміни кута θ_1 в межах $\theta_1 = (\gamma_1, \pi; -\pi, -\gamma_1)$ для амплітуд тиску $|p|$ і швидкості $|v|$. Графіки на рис. 3 показують хорошу якість виконання умов спряження полів.

Можна відзначити такі особливості в поведінці кривих на рис. 2 ($\theta_0 = 90^\circ$). Для амплітуди тиску $|p|$ чітко виділяється інтервал кутів $\theta_1 = (\gamma_1, 90^\circ)$, у якому спостерігається яскрава інтерференція між хвилею, що падає, і хвилею, що відбивається, від плоскої поверхні клина великого хвильового розміру.

В інтервалі зміни кутів $\theta_1 = (120^\circ, 180^\circ)$, $(-180^\circ, -120^\circ)$ інтерференція між зазначеними хвилями практично відсутня і тут домінує плоска хвиля одиначної амплітуди, що падає. Інтервал кутів $\theta_1 = (-90^\circ, -\gamma_1)$ визначає зону геометричної тіні.

Подібний характер має і крива для амплітуди швидкості $|v|$ з корекцією на плавний спад до нуля за кутів $\theta_1 = \pm 180^\circ$ (для цих кутів нормальна складова швидкості до поверхні поділу часткових областей дорівнює нулю).

2. Розрахунок зворотного перерізу розсіювання

На рис. 3 показано графіки залежностей зворотного перерізу розсіювання $\sigma_L(\theta_0)$ для трьох варіантів моделі.

У всіх трьох варіантах розмір плоскої поверхні моделі залишається незмінним, а саме $R/\lambda = 10,2$. Для кривої 1 заокруглення профілю є акустично жорстким із радіусом кола $a/\lambda = 0,8$. Для кривої 2 заокруглення також є акустично жорстким, але з радіусом округлення вдвічі меншим, тобто $a/\lambda = 0,4$. Крива 3 відповідає заокругленню з поглинаючою поверхнею з коефіцієнтом поглинання $V = 0,1$ і величиною $a/\lambda = 0,8$.

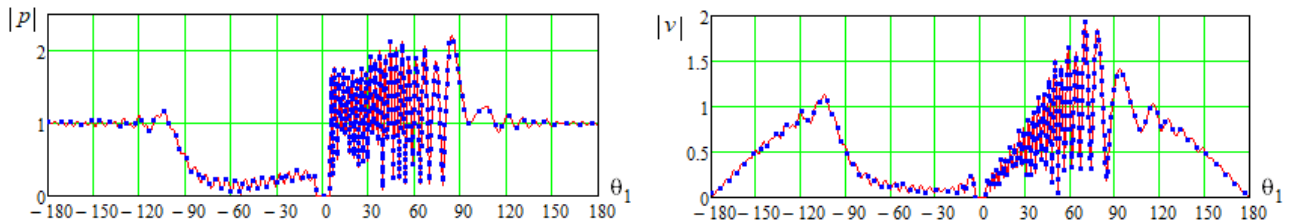


Рис. 2. Приклад виконання граничних умов на межі часткових областей 1 і 2 для амплітуд тиску $|p|$ і швидкості $|v|$; $\theta_0 = 90^\circ$: лінія – поле з боку області 1, точки – поле з боку області 2; інтервал кутів $\theta_1 = (-\gamma_1, \gamma_1)$, $\gamma_1 \approx 4,5^\circ$ є нефізичним, тому всередині цього інтервалу значення поля не обчислюємо

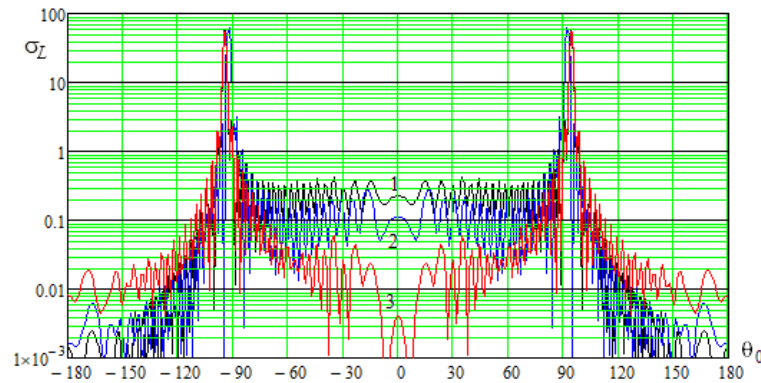


Рис. 3. Залежність зворотного перерізу розсіювання σ_L від кута падіння плоскої хвилі θ_0 за різних параметрів моделі, $R/\lambda = 10,2$: 1 – $a/\lambda = 0,8$, $V = 1$, $\bar{Y} = 0$; 2 – $a/\lambda = 0,4$, $V = 1$, $\bar{Y} = 0$; 3 – $a/\lambda = 0,8$, $V = 0,1$, $\bar{Y} = 0,818$

Як бачимо, для зворотного перерізу розсіювання σ_L усі три характеристики мають два відблиски, обумовлені нормальним падінням плоскої хвилі на плоску поверхню значного хвильового розміру $R/\lambda = 10,2$. Порівнюючи криві 1 і 2, можна відзначити, що зменшення радіуса заокруглення знижує величину зворотного перерізу розсіювання в діапазоні кутів $-70^\circ \leq \theta_0 \leq 70^\circ$, що є цілком очікуваним результатом. Покриття поверхні заокруглення поглинаючим матеріалом із коефіцієнтом поглинання $V = 0,1$ істотно знижує величину перерізу зворотного розсіювання в діапазоні кутів $-70^\circ \leq \theta_0 \leq 70^\circ$ (крива 3) порівняно з кривою 1. Ми знайшли значення $N = 14 \cdot 10^9$, що відповідає інтервалу $[-A, A] = [-1000, 1000]$.

Дискусія і висновки

На основі методу часткових областей побудовано точний (строгий) розв'язок задачі розсіювання плоскої хвилі на скінченному клиноподібному об'єкті. Достовірність розв'язку обумовлена контролем якості виконання умов спряження полів на межах часткових областей і граничної умови на згладженій поверхні об'єкта.

Проведено розрахунки зворотного перерізу розсіювання об'єкта. Показано, що зворотний переріз розсіювання істотно залежить від радіуса заокруглення профілю і нормальної провідності поверхні заокруглення. При цьому покриття поверхні згладжування поглинаючим матеріалом істотно зменшує, в певному діапазоні кутів падіння хвилі, розсіяне поле об'єкта.

Побудований чисельно-аналітичний розв'язок дає змогу проводити оціночні розрахунки зворотного перерізу розсіювання об'єкта у разі зміни в широкому діапазоні його геометричних і фізичних параметрів.

Внесок авторів: Віктор Грінченко – постановка задачі й обговорення висновків; Ірина Лебедева – робота з літературними джерелами, побудова аналітичного розв'язку й обговорення висновків; Володимир Маципура – побудова аналітичного розв'язку й обговорення висновків; Тихомир Трифонов – побудова аналітичного розв'язку й обговорення висновків.

Джерела фінансування. Фінансування здійснюється власним коштом авторів і частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Грінченко, В. Т., Лебедева, І. В., & Маципура, В. Т. (2024). Вплив згладжування поверхні скінченного клиноподібного об'єкта на характеристики розсіювання плоскої хвилі. Частина I. Побудова розв'язку. *Вісник Київського національного університету імені Тараса Шевченка. Фізико-математичні науки*, 79(2), 29–32. <https://doi.org/10.17721/1812-5409.2024/2.5>
- Daniele, V., & Lombardi, G. (2020). *Scattering and Diffraction by Wedges 1: the Wiener-Hopf Solution-Theory*. John Wiley & Sons.
- Emig, T. (2024). Surface scattering expansion for the Casimir-Polder interaction of a dielectric wedge. *Physical Review A*, 110(6), 062809. <https://doi.org/10.1103/PhysRevA.110.062809>
- Felsen, L., & Marcuvitz, N. (1973). *Radiation and Scattering of Waves*. Prentice-Hall, Inc.
- Grinchenko, V. T., Vovk, I. V., & Matsypura, V. T. (2018). *Acoustic wave problems*. Begell House, Inc.
- Grzesik, J. A. (2019). Dielectric Wedge Scattering: An Analytic Inroad. *Progress In Electromagnetics Research B*, 84, 43–60. <https://doi.org/10.2528/PIERB19013001>
- Hacivelioglu, F., Sevgi, L., & Ufimtsev, P. Y. (2011). Electromagnetic wave scattering from a wedge with perfectly reflecting boundaries: Analysis of asymptotic techniques. *IEEE Antennas & Propagation Magazine*, 53(3), 232–253. <https://doi.org/10.1109/MAP.2011.6028472>

- Monzon, C., Forester, D., & Loschialpo, P. (2005). Exact solution to line source scattering by an ideal left-handed wedge. *Physical Review E*, 72(5), 056606. <https://doi.org/10.1103/PhysRevE.72.056606>
- Sommerfeld, A. (2004). *Mathematical Theory of Diffraction*. Springer Science+Business. <https://doi.org/10.1007/978-0-8176-8196-8>

References

- Daniele, V., & Lombardi, G. (2020). *Scattering and Diffraction by Wedges 1: the Wiener-Hopf Solution-Theory*. John Wiley & Sons.
- Emig, T. (2024). Surface scattering expansion for the Casimir-Polder interaction of a dielectric wedge. *Physical Review A*, 110(6), 062809. <https://doi.org/10.1103/PhysRevA.110.062809>
- Felsen, L., & Marcuvitz, N. (1973). *Radiation and Scattering of Waves*. Prentice-Hall, Inc.
- Grinchenko, V. T., Lebedyeva, I. V., & Matsypura, V. T. (2024). Effect of smoothing of a finite wedge-shaped object on the scattering characteristics of a plane wave. Part I. Solution construction. *Bulletin of the Taras Shevchenko National University of Kyiv. Physics and Mathematics*, 79(2), 29–32. <https://doi.org/10.17721/1812-5409.2024/2.5>
- Grinchenko, V. T., Vovk, I. V., & Matsypura, V. T. (2018). *Acoustic wave problems*. Begell House, Inc.
- Grzesik, J. A. (2019). Dielectric Wedge Scattering: An Analytic Inroad. *Progress In Electromagnetics Research B*, 84, 43–60. <https://doi.org/10.2528/PIERB19013001>
- Hacivelioglu, F., Sevgi, L., & Ufimtsev, P. Y. (2011). Electromagnetic wave scattering from a wedge with perfectly reflecting boundaries: Analysis of asymptotic techniques. *IEEE Antennas & Propagation Magazine*, 53(3), 232–253. <https://doi.org/10.1109/MAP.2011.6028472>
- Monzon, C., Forester, D., & Loschialpo, P. (2005). Exact solution to line source scattering by an ideal left-handed wedge. *Physical Review E*, 72(5), 056606. <https://doi.org/10.1103/PhysRevE.72.056606>
- Sommerfeld, A. (2004). *Mathematical Theory of Diffraction*. Springer Science+Business. <https://doi.org/10.1007/978-0-8176-8196-8>

Отримано редакцією журналу / Received: 25.07.25

Прорецензовано / Revised: 29.09.25

Схвалено до друку / Accepted: 10.10.25

Viktor GRINCHENKO¹, DSc (Phys. & Math.), Active Member of the NAS of Ukraine
ORCID ID: 0000-0003-3229-1810
e-mail: grinchenko@nas.gov.ua

Iryna LEBEDYEVA², PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0001-7150-1310
e-mail: lebedyevaiv@knu.ua

Volodymyr MATSYUPURA², DSc (Phys. & Math.), Prof.
ORCID ID: 0000-0002-0136-6659
e-mail: mnivtt@gmail.com

Tihomir TRIFONOV³, PhD (Phys. & Math.), Prof.
ORCID ID: 0000-0002-9649-4396
e-mail: TihomirTrifonov@ieee.org

¹Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine

²Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

³University of Veliko Turnovo St Cyril and St Methodius, Veliko Turnovo, Bulgaria

EFFECT OF SMOOTHING OF A FINITE WEDGE-SHAPED OBJECT ON THE SCATTERING CHARACTERISTICS OF A PLANE WAVE. PART II. ANALYSIS OF NUMERICAL RESULTS

This article is the second part of the work (Grinchenko, Lebedyeva, & Matsypura, 2024), which is devoted to the problem of scattering of a plane harmonic wave on a finite wedge-shaped object with a smoothed surface (in other words, with rounding). Such an object represents a simplified model of an aircraft wing profile. For this model, based on the solution obtained within the limits of the partial domains method, an expression for the inverse scattering cross section is determined. An example of the fulfillment of boundary conditions at the boundary of partial domains is given. Plots of the backscattering cross-section were constructed for different rounding radii of the front edge of the model profile and different conductivities of the rounding surface. It is shown that the backscattering cross-section significantly depends on the radius of the profile rounding and the normal conductivity of the rounding surface. In this case, covering the smoothing surface with an absorbing material significantly reduces, in a certain range of wave incidence angles, the scattered field of the object. The developed numerical-analytical solution makes it possible to perform approximate calculations of the backscattering cross section of the object under variation of its geometric and physical parameters over a wide range.

Key words: inverse scattering cross section; partial domains method; plane wave; point conjugation of fields.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 534.83:519.6

DOI: <https://doi.org/10.17721/1812-5409.2025/2.13>

Андрій ДРЕУС¹, д-р техн. наук, проф.
ORCID ID: 0000-0002-1234-5678
e-mail: dreus_a@365.dnu.edu.ua

Віталій ДЕРБАБА², канд. техн. наук, доц.
ORCID ID: 0000-0002-3918-2177
e-mail: derbaba.v.a@nmu.one

Дмитро ГАРКАВЕНКО², асп.
ORCID ID: 0009-0004-5011-9015
e-mail: harkavenko.d.v@nmu.one

Сергій АЛЕКСЄЄНКО², д-р техн. наук, проф.
ORCID ID: 0000-0003-0320-989X
e-mail: aleksieienko.s.v@nmu.one

¹Дніпровський національний університет імені Олеся Гончара, Дніпро, Україна
²Національний технічний університет "Дніпровська політехніка", Дніпро, Україна

АНАЛІЗ АКУСТИЧНОГО ВИПРОМІНЮВАННЯ МОДИФІКОВАНОГО СЕГМЕНТА ЛОПАТІ ВІТРОВОЇ ТУРБІНИ З ПРОФІЛЕМ НАСА 0012

Актуальність дослідження зумовлена потребою зменшення шумового забруднення, що виникає під час експлуатації вітроенергетичних установок, зокрема в умовах густонаселених або екологічно чутливих регіонів. Одним із перспективних підходів до зниження акустичного випромінювання лопатей вітрових турбін є модифікація їхньої геометрії для контролю турбулентних структур потоку. У пропонованій роботі представлено методологію та результати експериментального дослідження аеродинамічного шуму, що генерується профілем лопаті типу НАСА0012. Досліджено вплив періодичних зазубрин на задній крайці на акустичні характеристики профілю. Метою роботи є оцінка ефективності зазубреної задньої крайки у зниженні рівня аеродинамічного шуму. Дослідження проведено в безлунній камері з використанням фізичних моделей профілю, виготовлених методом 3D-друку, а акустичні вимірювання виконано за допомогою шумоміра з подальшим обробленням даних методом швидкого перетворення Фур'є (FFT). Експерименти охоплювали різні швидкості набігаючого потоку (від 10 до 20 м/с) і кути атаки (0° та 5°), а результати подано у вигляді спектрів звукового тиску. Встановлено, що модифікований профіль забезпечує зниження рівня звукового тиску до 7 дБ порівняно зі стандартним профілем. Ефект особливо виражений у частотному діапазоні до 8 кГц, характерному для вихрових джерел шуму за відносно низьких швидкостей потоку. Запропонований підхід демонструє високу ефективність і може бути застосований для оптимізації конструкцій вітрових турбін, а також інших обертових елементів, таких як пропелери БПЛА або вентилятори.

Ключові слова: аеродинамічний шум, вітрова турбіна, профіль НАСА0012, зазубрена крайка, експериментальне дослідження, акустичні вимірювання, FFT, акустичне випромінювання.

Класифікація відповідно до AMS 2020: 76Q05.

Вступ

Аеродинамічний шум, що виникає під час роботи вітрових турбін, є негативним фактором, який обмежує можливості розвитку наземної вітрової енергетики, зокрема в районах із великою щільністю населення, або в районах, вразливих з погляду біорізноманіття. Детальний огляд соціальних й екологічних проблем, пов'язаних з акустичним випромінюванням, подано в роботах (Omodaro et al., 2023; Sander, Jung, & Schindler, 2024). Природа виникнення аеродинамічного шуму полягає у взаємодії задньої та передньої крайок профіля лопаті з турбулентними структурами під час обтікання потоком повітря. Основні фізичні механізми генерації шумового випромінювання аеродинамічним профілем розглянуто, наприклад, у роботі (Bertagnolio, & Fischer, 2022). Варто зауважити, що проблема аеродинамічного шуму лопатей є актуальною не тільки для вітрових турбін, а також і для промислових та побутових вентиляторів, лопатей пропелерів БПЛА та інших літальних апаратів.

Методи боротьби з аеродинамічним шумом поділяються на активні і пасивні (Ahmed et al., 2025). Активні методи використовують зовнішні впливи на структуру примежового шару задля керування станом примежового шару і турбулентністю. До таких методів належать методи вдування повітря або відсмоктування примежового шару з поверхні (Yang et al., 2023), керування формою лопаті за допомогою закріпків (Hu et al., 2022), використання електроплазмових методів (Patel et al., 2021) тощо. Незважаючи на те, що активні методи дають змогу досягти кращих результатів зі зниження рівня шуму, проте вони є досить складними, вартісними, потребують додаткового споживання енергії і менш надійні. Натепер такі методи використовуються переважно лише в експериментальних вітрових турбінах. На практиці набули поширення пасивні методи зниження шуму. Такі методи передбачають використання шумопоглинальних або текстурних матеріалів (Yang et al., 2024a) для лопатей турбін, пористих (Shige et al., 2024) або демпфірувальних вставок у конструкції лопаті (Arif et al., 2024), а також різні модифікації задньої крайки лопаті, що дає змогу руйнувати турбулентні структури, які виникають (Liu et al., 2022). Саме останній підхід є одним з найперспективніших, оскільки допомагає зменшити рівень шуму без втрати аеродинамічних характеристик профіля лопаті (Zhou et al., 2022).

Упровадження таких модифікацій потребує здійснення відповідних акустичних досліджень. Експериментальні акустичні дослідження вимагають проведення дослідів у спеціальних безлунних кімнатах або аеродинамічних трубах з організацією складної інформаційно-вимірювальної системи (Firoozi et al., 2024). Основною проблемою в таких дослідженнях є складність відокремлення фонового шуму або шуму від зовнішніх джерел (механічного шуму експериментальної установки). Основним інструментом дослідження акустичних процесів й оптимізації профілю турбін є математичне

© Дреус Андрій, Дербабa Віталій, Гаркавенко Дмитро, Алексєєнко Сергій, 2025

моделювання. Нині запропоновано велику кількість моделей і методів розрахунку аеродинаміки профілів від простих, що використовують теорію ідеального вітроколеса (Goman et al., 2022), до складних числових методів обчислювальної аеродинаміки (Alekseyenko et al., 2022) і методів машинного навчання (Yang et al., 2024b). Хоча сучасні числові підходи, зокрема методи обчислювальної аероакустики (CAA), демонструють значний прогрес у передбаченні шумових характеристик, вони залишаються залежними від моделювальних припущень, як-от: турбулентність, граничні умови, точність чисельної сітки. Експериментальні вимірювання, зокрема мікрофонні масиви, забезпечують незалежне джерело даних, що дає змогу не лише перевірити точність числових моделей, але й виявити фізичні явища, які важко піддаються моделюванню). Експериментальні методи посідають ключове місце в аероакустичних дослідженнях, доповнюючи числове моделювання шляхом верифікації й уточнення розрахункових результатів. Отже, інтеграція експериментальних і числових методів є критично важливою для достовірного аналізу аероакустичних процесів та розроблення ефективних засобів шумозниження. У пропонованій роботі використано експериментальний метод й алгоритм обробки експериментальних даних на основі швидкого перетворення Фур'є (FFT) для дослідження акустичних процесів, що перебігають під час обтікання стандартного і модифікованого профілів лопаті турбіни.

Об'єкт дослідження – аеродинамічний профіль NACA0012, який найчастіше використовують у тестових задачах з дослідження аеродинаміки лопатей турбін.

Мета роботи – визначення впливу модифікації задньої крайки профіля на рівень аеродинамічного шуму шляхом математичного й комп'ютерного моделювання. Для досягнення мети сформульовано такі завдання: визначити методологію дослідження, виконати розрахунки акустичних полів навколо профілів різної геометрії, оцінка впливу модифікації задньої крайки на рівень аеродинамічного шуму профіля.

1. Постановка задачі і методологія дослідження

Розглянуто два варіанти виконання лопаті вітрової турбіни, що мають аеродинамічний профіль NACA0012: стандартний і модифікований. Експериментальні моделі, що були виготовлені методом 3D-друку представлені на рис. 1.

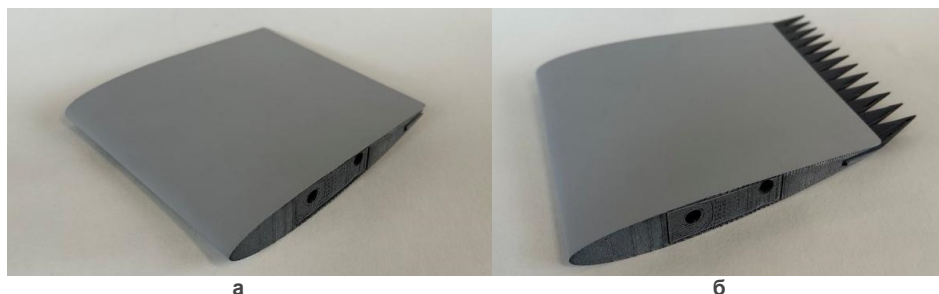


Рис. 1. Схеми виконання аеродинамічного профілю:
а – стандартна; б – модифікована

Стандартне виконання профілю передбачає гостру лінійну у плані задню крайку, як показано на рис. 1 а. У другому варіанті задню крайку виконано у вигляді зазубрин, як показано на рис. 1 б. Передбачається, що така модифікація сприятиме руйнуванню сильних когерентних турбулентних вихорів у зоні відриву потоку, що має приводити до зменшення інтенсивності аеродинамічного шуму. Розглянуто сегмент профіля довжиною 200 мм і хордою 215 мм, на якому розташовано 13 однакових загострених зазубрин розміром 43,3 мм.

Для дослідження аеродинамічного шуму сегмента лопаті вітрової турбіни розроблено оригінальний експериментальний підхід, реалізований у безлунній камері розміром 3,8×3,8×3,2 м, де стіни вкрито звукопоглинальними матеріалами. Експериментальна установка складалась з обертової штанги із закріпленою моделлю сегмента лопаті профілю NACA0012, що рухалась зі сталою лінійною швидкістю (до 24 м/с) за нульового кута атаки. Для уникнення впливу кінцевих вихорів модель розміщували між торцевими пластинами.

Акустичні вимірювання проводили шумоміром SV 971A1 з мікрофоном ACO SV 7152, розташованим на фіксованій відстані від траєкторії моделі. Для зменшення впливу шуму від приводу встановлено звукоізоляційний екран. Обробка сигналів виконували із застосуванням алгоритму швидкого перетворення Фур'є із частотною роздільною здатністю 12 Гц. Для виділення корисного сигналу застосовували спектральне віднімання фонового шуму, що фіксували окремо до запуску моделі. Збіжність результатів вимірювань забезпечено шляхом контролю повторності й усереднення результатів за множиною інтервалів спостереження (до 20), відповідно до вимог стандарту ДСТУ ГОСТ ISO 5725-1:2005. Отримані результати дали змогу ідентифікувати тональні компоненти шуму, а також встановити залежність рівня звукового тиску.

Аналіз акустичних сигналів проводили з використанням алгоритму швидкого перетворення Фур'є, що надає можливість перейти від часової області до частотної з роздільною здатністю 12 Гц. Основною метою було виділення спектра аеродинамічного шуму, зумовленого рухом моделі лопаті, із сумарного сигналу, що містив також фонові компоненти.

Акустичний сигнал $x(t)$ надходить із шумоміру у вигляді дискретної послідовності значень $x[n]$, де n – індекс відліку часу, а частота дискретизації f визначає часовий крок $t = 1/f$. Швидке перетворення Фур'є дає можливість обчислити оцінку спектра сигналу $X[k]$ у частотній області:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-i\frac{2\pi}{N}kn}, \quad k = 0, 1, \dots, N-1, \quad (1)$$

де N – довжина аналізованого вікна (кількість точок FFT), k – індекс частоти, що відповідає $f_k = \frac{kf}{N}$.

З (1) визначаємо амплітуду $|X[k]|$ та $\arg X[k]$. Нехай у часовій області сумарний сигнал, що отримуємо в результаті вимірювань,

$$x(t) = s(t) + n(t), \quad (2)$$

де $s(t)$ – корисний сигнал (шум моделі), який потрібно відокремити, $n(t)$ – фоновий шум, виміряний окремо в аналогічних умовах, але без моделі. У частотній області запишемо

$$X[f] = S(f) + N(f). \quad (3)$$

Оцінку спектра корисного сигналу $S(f)$ отримаємо через спектральне віднімання

$$|S(f)|^2 = \max(|X(f)|^2 - \alpha|N(f)|^2, 0), \quad (4)$$

де α – коефіцієнт компенсації шуму ($1 \leq \alpha \leq 2$), від'ємні значення усуваємо через обмеження на ≥ 0 . Реконструкція спектра з урахуванням фази

$$S(f) = \sqrt{|S(f)|^2} \cdot e^{i \arg X(f)}. \quad (5)$$

Після очищення спектра застосовуємо обернене перетворення Фур'є для повернення в часову область

$$s[n] = \frac{1}{N} \sum_{k=0}^{N-1} S[k] \cdot e^{i \frac{2\pi}{N} kn}. \quad (6)$$

Вираз (6) дає змогу отримати очищений сигнал. Спектральна щільність потужності конвертується в рівень звукового тиску SPL

$$SPL(f) = 10 \lg \left(\frac{p(f)^2}{p_0^2} \right) \frac{1}{N} \sum_{k=0}^{N-1} S[k] \cdot e^{i \frac{2\pi}{N} kn}, \quad (7)$$

де $p_0 = 20$ мкПа – референтне значення тиску для повітря. Для зменшення випадкових флуктуацій спектри усереднюються за кількома незалежними інтервалами спостереження

$$SPL = \frac{1}{M} \sum_{j=1}^M SPL_j(f), \quad (8)$$

де M – кількість незалежних інтервалів спостережень. Оцінка стабільності проводиться за відносним відхиленням

$$\Delta_{rel} = \frac{1}{M} \sum_{j=1}^M \frac{|SPL_j(f) - SPL|}{SPL}. \quad (9)$$

Для інтегральної оцінки шуму використовуємо середньозважений рівень шуму

$$L_{eq} = 10 \lg \left(\frac{1}{T} \int_0^T 10^{L(t)/10} dt \right), \quad (10)$$

де T – тривалість сигналу, $L(t)$ – миттєвий рівень шуму в децибелах.

3. Результати дослідження

Відповідно до запропонованого алгоритму було проведено низку досліджень акустичного випромінювання сегментів стандартної (без зазубрин) і модифікованої (із зазубринами) лопатей вітрової турбіни, що встановлені під різними кутами атаки і на різних швидкостях вітру. На рис. 2 представлено порівняльний аналіз рівнів аеродинамічного шуму SPL для тестового (стандартного профіля NACA0019 без зазубрин) сегмента лопаті і модифікованої лопаті (із зазубринами). Результати представлені у вигляді спектрів, що отримані внаслідок швидкого перетворення Фур'є. Розрахунки проводили для швидкостей набігаючого потоку від 10 м/с до 20 м/с, результати представлено для нульового кута атаки. Отримані спектральні характеристики дають змогу простежити частотну структуру акустичного випромінювання, що, зі свого боку, допомагає локалізувати джерела шуму, пов'язані з турбулентними збуреннями на задній крайці. Це забезпечує надійний критерій для порівняння ефективності різних геометричних конфігурацій профілю в контексті зниження аеродинамічного шуму.

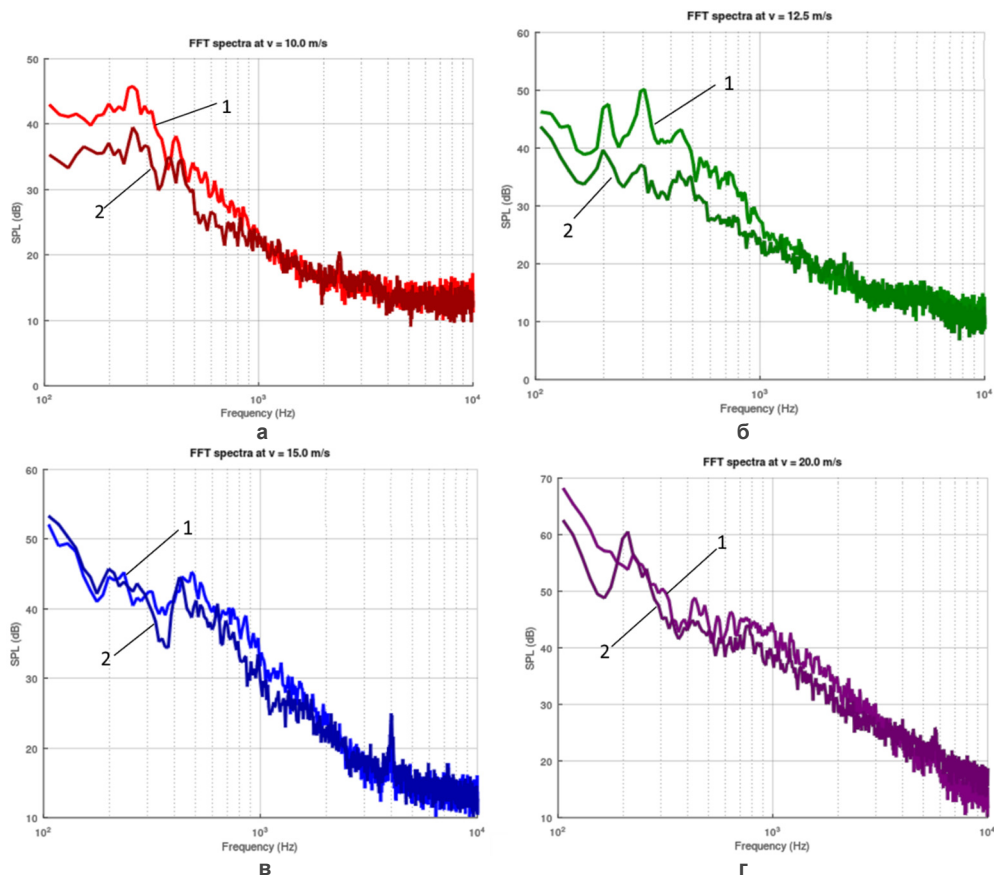


Рис. 2. FFT-спектри шуму від тестової (лінія 1) і модифікованої (лінія 2) лопатей за нульового кута атаки і швидкостей набігаючого потоку: а – 10 м/с; б – 12,5 м/с; в – 15 м/с; г – 20 м/с

Окрім швидкості, на акустичне випромінювання значно впливає кут атаки. На рис. 3 представлено вплив кута атаки на рівень шуму для модифікованого профілю.

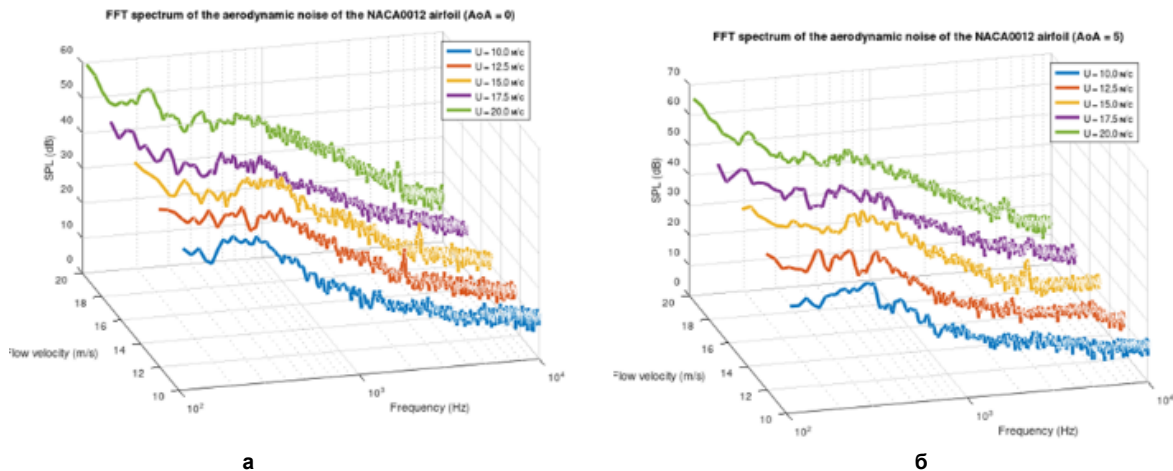


Рис. 4. FFT спектри шуму для модифікованого профілю на різних кутах атаки: а – 0°; б – 5°

На рис. 4 представлено результати моделювання акустичного випромінювання як для стандартного, так і модифікованого профілю для кута атаки 0°.

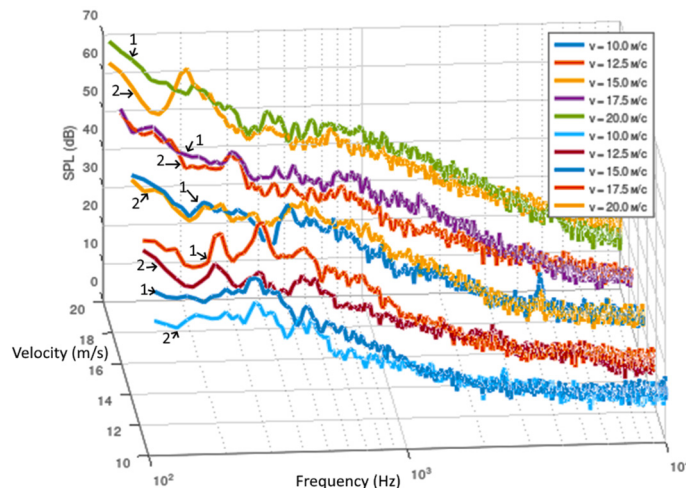


Рис. 4. FFT спектри шуму для кута атаки 5° і різних швидкостях набігаючого потоку :
1 – тестовий профіль; 2 – модифікований профіль

Отже, отримані спектри дають змогу наочно порівняти рівні аеродинамічного шуму між стандартним і модифікованим профілями за різних умов обтікання. Виявлені відмінності в частотному складі шуму створюють підґрунтя для подальшого аналізу ефективності запропонованої геометричної модифікації.

Дискусія і висновки

Результати дослідження дали змогу отримати надійні дані щодо розподілу акустичного тиску та спектральних характеристик шуму. Наведені результати числового аналізу спектральних характеристик аеродинамічного шуму підтверджують ефективність запропонованої модифікації задньої крайки профілю NACA0012. Модифікований профіль із зазубреними крайками продемонстрував зниження рівня звукового тиску в широкому частотному діапазоні порівняно з традиційним профілем, особливо в діапазоні низьких і середніх частот (до 8 кГц), що є типовими для аеродинамічного випромінювання від вихрових структур. У височастотній області акустичні характеристики не демонструють значної розбіжності для обох випадків. Показано, що для нульового кута атаки максимальне зниження рівня SPL становить до 5–7 дБ за швидкостей набігаючого потоку 10–15 м/с. Цей ефект дещо зменшується за збільшення швидкості набігаючого потоку і підвищення кута атаки. Це пояснюємо зміною структури відриву потоку та зсувом джерел шуму. Водночас для всіх розглянутих розрахункових варіантів модифікація профілю дає змогу зменшити акустичне випромінювання. Зауважимо, що збільшення кута атаки призводить до збільшення рівня аеродинамічного шуму від лопаті.

Проведене дослідження підтвердило, що модифікація задньої крайки профілю лопаті вітрової турбіни у вигляді періодичних зазубрин сприяє суттєвому зменшенню аеродинамічного шумового випромінювання без втрати геометричної цілісності або аеродинамічної ефективності. Найбільший ефект зниження рівня шуму спостерігаємо за нульового кута атаки та середніх швидкостей набігаючого потоку. Це означає, що запропонована модифікація є перспективною для використання у вітрових турбінах, а також оптимізації інших типів лопатей (напр., пропелерів БПЛА або промислових вентиляторів). Розроблена методика аналізу може бути використана для подальших досліджень й оптимізації профілю лопаті вітрової турбіни з метою зменшення аеродинамічного шуму.

Внесок авторів: Андрій Дреус – концептуалізація, аналіз джерел, розробка методики обробки даних, аналіз результатів; Віталій Дербаба – проведення експериментів; Дмитро Гаркавенко – підготовка та проведення експериментів, обробка результатів досліджень; Сергій Алексєнко – розробка фізичних моделей тестових профілів, проведення акустичних вимірювань у безлунній камері, інтерпретація й аналіз отриманих даних.

Джерела фінансування. Роботу виконано за підтримки зовнішнього інструменту допомоги Європейського Союзу для виконання зобов'язань України у Рамковій програмі Європейського Союзу з наукових досліджень та інновацій "Горизонт 2020", договір PH/54-2024 з Міністерством освіти і науки України, № державної реєстрації: 0124U004405.

Список використаних джерел

- Ahmed, M. M., Robin, H. M., Shahadat, M. Z., & Masud, M. H. (2025). Innovative approaches for reducing wind turbine noise: A review from mechanical and aerodynamic perspective. *Energy Reports*, 13, 728–746. <https://doi.org/10.1016/j.egy.2024.12.049>
- Alekseyenko, S., Dreus, A., Dron, M., & Brazaluk, O. (2022). Numerical Study of Aerodynamic Characteristics of a Pointed Plate of Variable Elongation in Subsonic and Supersonic Gas Flow. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, 96(2), 2022, 88–97. <https://doi.org/10.37934/arfm.96.2.889>
- Arif, I., Lam, G., & Leung, R. (2022). Distributed surface compliance for airfoil tonal noise reduction at various loading conditions. *Physics of Fluid*, 34, 046113. <https://doi.org/10.1063/5.0087350>
- Bertagnolio, F., & Fischer, A. (2022). Wind Turbine Aerodynamic Noise Sources. In *Handbook of Wind Energy Aerodynamics*. Springer. Cham. https://doi.org/10.1007/978-3-030-31307-4_70
- Firoozi, A. A., Hejazi, F., & Firoozi, A. A. (2024). Advancing Wind Energy Efficiency: A Systematic Review of Aerodynamic Optimization in Wind Turbine Blade Design. *Energies*, 17(12), 2919. <https://doi.org/10.3390/en17122919>
- Goman, O., Dreus, A., Rozhkevych, A., Heti, K., & Karplyuk, V. (2022). Improving the efficiency of Darier rotor by controlling the aerodynamic design of blades. *Energy Reports*, 8, 788–794. <https://doi.org/10.1016/j.egy.2022.10.162>
- Hu Z., Xu G., Shi Y., & Xia R. (2022). Airfoil–Vortex Interaction Noise Control Mechanism Based on Active Flap Control. *Journal of Aerospace Engineering*, 35(1). [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0001356](https://doi.org/10.1061/(ASCE)AS.1943-5525.0001356)
- Liu, X., Zang, B., & Azarpeyvand, M. (2022). Wake-aerofoil interaction noise control with trailing-edge serrations. *Experimental Thermal and Fluid Science*, 130, 110510. <https://doi.org/10.1016/j.expthermflusci.2021.110510>
- Omodara, L., Saavalainen, P., Pitkäaho, S., Pongrácz, E., & Keiski, R. L. (2023). Sustainability assessment of products – Case study of wind turbine generator types. *Environmental Impact Assessment Review*, 98, 106943. <https://doi.org/10.1016/j.eiar.2022.106943>
- Patel, T. K., Lilley, A. J., Shen, W. et al. (2021). Fundamental investigation using active plasma control to reduce blade–vortex interaction noise. *International Journal of Aeroacoustics*, 20(8), 870–900. <https://doi.org/10.1177/1475472X211052699>
- Sander, L., Jung, C., & Schindler, D. (2024). Global Review on Environmental Impacts of Onshore Wind Energy in the Field of Tension between Human Societies and Natural Systems. *Energies*, 17(13), 3098. <https://doi.org/10.3390/en17133098>
- Shige, K., Terashima, J., Inasawa, A., Komatsuzaki, T., Sakai, Y., & Kusano T. (2024). On the reduction of the flow-induced noise using porous material plates with high acoustic transmissibility. *Journal of Sound and Vibration*, 568, 117967. <https://doi.org/10.1016/j.jsv.2023.117967>
- Yang, C., Arcondoulis E. J. G., Yang Y., Guo H., Maryami, R., Bi C., & Liu, Y., (2023). Active control of airfoil turbulent boundary layer noise with trailing-edge blowing. *Journal of the Acoustical Society of America*, 153, 2115. <https://doi.org/10.1121/10.001778>
- Yang, Y., Han S.-H., Ahn, S., Kim, J., Lee, S., & Park, K. (2024a). Design and manufacturing of aeroacoustic metamaterial: Textured rotor blades with enhanced acoustic and aerodynamic performance. *Additive Manufacturing*, 84, 104109. <https://doi.org/10.1016/j.addma.2024.104109>
- Yang, H., Yuan, W., Zhu, W., Sun, Z., Zhang, Y., Zhou, Y. (2024b). Wind turbine airfoil noise prediction using dedicated airfoil database and deep learning technology. *Applied Energy*, 364, 123165. <https://doi.org/10.1016/j.apenergy.2024.123165>
- Zhou, T., Cao, H., Zhang, V., & Liao, C. (2022). Performance simulation of wind turbine with optimal designed trailing-edge serration. *Energy*, 243, 122998. <https://dx.doi.org/10.2139/ssrn.3914850>

References

- Ahmed, M. M., Robin, H. M., Shahadat, M. Z., & Masud, M. H. (2025). Innovative approaches for reducing wind turbine noise: A review from mechanical and aerodynamic perspective. *Energy Reports*, 13, 728–746. <https://doi.org/10.1016/j.egy.2024.12.049>
- Alekseyenko, S., Dreus, A., Dron, M., & Brazaluk, O. (2022). Numerical Study of Aerodynamic Characteristics of a Pointed Plate of Variable Elongation in Subsonic and Supersonic Gas Flow. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, 96(2), 2022, 88–97. <https://doi.org/10.37934/arfm.96.2.889>
- Arif, I., Lam, G., & Leung, R. (2022). Distributed surface compliance for airfoil tonal noise reduction at various loading conditions. *Physics of Fluid*, 34, 046113. <https://doi.org/10.1063/5.0087350>
- Bertagnolio, F., & Fischer, A. (2022). Wind Turbine Aerodynamic Noise Sources. In *Handbook of Wind Energy Aerodynamics*. Springer. Cham. https://doi.org/10.1007/978-3-030-31307-4_70
- Firoozi, A. A., Hejazi, F., & Firoozi, A. A. (2024). Advancing Wind Energy Efficiency: A Systematic Review of Aerodynamic Optimization in Wind Turbine Blade Design. *Energies*, 17(12), 2919. <https://doi.org/10.3390/en17122919>
- Goman, O., Dreus, A., Rozhkevych, A., Heti, K., & Karplyuk, V. (2022). Improving the efficiency of Darier rotor by controlling the aerodynamic design of blades. *Energy Reports*, 8, 788–794. <https://doi.org/10.1016/j.egy.2022.10.162>
- Hu Z., Xu G., Shi Y., & Xia R. (2022). Airfoil–Vortex Interaction Noise Control Mechanism Based on Active Flap Control. *Journal of Aerospace Engineering*, 35(1). [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0001356](https://doi.org/10.1061/(ASCE)AS.1943-5525.0001356)
- Liu, X., Zang, B., & Azarpeyvand, M. (2022). Wake-aerofoil interaction noise control with trailing-edge serrations. *Experimental Thermal and Fluid Science*, 130, 110510. <https://doi.org/10.1016/j.expthermflusci.2021.110510>
- Omodara, L., Saavalainen, P., Pitkäaho, S., Pongrácz, E., & Keiski, R. L. (2023). Sustainability assessment of products – Case study of wind turbine generator types. *Environmental Impact Assessment Review*, 98, 106943. <https://doi.org/10.1016/j.eiar.2022.106943>
- Patel, T. K., Lilley, A. J., Shen, W. et al. (2021). Fundamental investigation using active plasma control to reduce blade–vortex interaction noise. *International Journal of Aeroacoustics*, 20(8), 870–900. <https://doi.org/10.1177/1475472X211052699>
- Sander, L., Jung, C., & Schindler, D. (2024). Global Review on Environmental Impacts of Onshore Wind Energy in the Field of Tension between Human Societies and Natural Systems. *Energies*, 17(13), 3098. <https://doi.org/10.3390/en17133098>
- Shige, K., Terashima, J., Inasawa, A., Komatsuzaki, T., Sakai, Y., & Kusano T. (2024). On the reduction of the flow-induced noise using porous material plates with high acoustic transmissibility. *Journal of Sound and Vibration*, 568, 117967. <https://doi.org/10.1016/j.jsv.2023.117967>
- Yang, C., Arcondoulis E. J. G., Yang Y., Guo H., Maryami, R., Bi C., & Liu, Y., (2023). Active control of airfoil turbulent boundary layer noise with trailing-edge blowing. *Journal of the Acoustical Society of America*, 153, 2115. <https://doi.org/10.1121/10.001778>
- Yang, Y., Han S.-H., Ahn, S., Kim, J., Lee, S., & Park, K. (2024a). Design and manufacturing of aeroacoustic metamaterial: Textured rotor blades with enhanced acoustic and aerodynamic performance. *Additive Manufacturing*, 84, 104109. <https://doi.org/10.1016/j.addma.2024.104109>

Yang, H., Yuan, W., Zhu, W., Sun, Z., Zhang, Y., Zhou, Y. (2024b). Wind turbine airfoil noise prediction using dedicated airfoil database and deep learning technology. *Applied Energy*, 364, 123165. <https://doi.org/10.1016/j.apenergy.2024.123165>

Zhou, T., Cao, H., Zhang, V., & Liao, C. (2022). Performance simulation of wind turbine with optimal designed trailing-edge serration. *Energy*, 243, 122998. <https://dx.doi.org/10.2139/ssrn.3914850>

Отримано редакцією журналу / Received: 04.07.25
Прорецензовано / Revised: 19.09.25
Схвалено до друку / Accepted: 10.10.25

Andrii DREUS¹, DSc (Engin.), Prof.
ORCID ID: 0000-0003-0598-9287
e-mail: dreus_a@365.dnu.edu.ua

Vitalii DERBABA², PhD (Engin.), Assoc. Prof.
ORCID ID: 0000-0002-3918-2177
e-mail: derbaba.v.a@nmu.one

Dmytro HARKAVENKO², PhD Student
ORCID ID: 0009-0004-5011-9015
e-mail: harkavenko.d.v@nmu.one

Serhii ALEKSIENKO², DSc (Engin.), Prof.
ORCID ID: 0000-0003-0320-989X
e-mail: aleksieienko.s.v@nmu.one

¹Oles Honchar Dnipro National University, Dnipro, Ukraine

²Dnipro University of Technology, Dnipro, Ukraine

ANALYSIS OF THE ACOUSTIC EMISSION OF A MODIFIED SEGMENT OF A WIND TURBINE BLADE WITH A NACA 0012 AIRFOIL

The relevance of the study is due to the need to reduce the noise pollution that occurs during the operation of wind power plants, in particular, in densely populated or environmentally sensitive regions. One of the promising approaches to reducing the acoustic radiation of wind turbine blades is the modification of their geometry to control turbulent flow structures. This paper presents the methodology and results of an experimental study of aerodynamic noise generated by a NACA0012 blade profile. The effect of periodic notches on the trailing edge on the acoustic characteristics of the profile is studied. The aim of the work is to assess the effectiveness of the notched trailing edge in reducing the level of aerodynamic noise. The study was conducted in an anechoic chamber using physical models of the profile manufactured by 3D printing, and acoustic measurements were performed using a sound level meter with subsequent data processing using the Fast Fourier Transform (FFT) method. The experiments covered different incident flow velocities (from 10 to 20 m/s) and angles of attack (0° and 5°), and the results are presented in the form of sound pressure spectra. It was found that the modified profile provides a reduction in the sound pressure level of up to 7 dB compared to the standard profile. The effect is especially pronounced in the frequency range up to 8 kHz, typical of vortex noise sources at relatively low flow velocities. The proposed approach demonstrates high efficiency and can be applied to optimize the designs of wind turbines, as well as other rotating elements, such as UAV propellers or fans.

Keywords: aerodynamic noise, wind turbine, NACA0012 airfoil, serrated edge, experimental investigation, acoustic measurements, FFT, acoustic emission.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 539.3

DOI: <https://doi.org/10.17721/1812-5409.2025/2.14>

Yaroslav ZHUK, DSc (Phys. & Math.), Prof.,
Corresponding Member of the National Academy of Sciences of Ukraine
ORCID ID: 0000-0002-2726-8395
e-mail: yaroslavzhuk@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Arash Soleiman FALLAH, PhD, Prof.
ORCID ID: 0000-0002-0382-633X
e-mail: arashsol@oslomet.no
Oslo Metropolitan University, Oslo, Norway

TECHNOLOGY FOR LASER THERMOFORMING OF THIN-WALLED METAL STRUCTURAL ELEMENTS

The article investigates the thermoforming of thin-walled metal structural elements using pulsed laser irradiation with a physically nonlinear thermoviscoplastic finite element model. The research focuses on the deformation mechanisms caused by transient thermal loads and their impact on the geometric accuracy of the molded parts. Simulation shows that pulsed laser irradiation creates localized thermal stresses and residual plastic deformations, which in turn cause controlled bending of thin metal sheets without the need for external dies or mechanical tools.

Numerical analysis demonstrates that the temperature-dependent viscoplastic properties of alloys play a decisive role in determining the efficiency of deformation and the stability of the resulting geometry. This article proposes a numerical model for simulating the laser thermoforming (LTF) process. It is based on the thermodynamically consistent theory of coupled thermoviscoplasticity. It is shown that it is suitable for modeling LTF for thin-walled metal structural elements. Within this model, the problem formulation consists of the Cauchy equation, the equations of motion, and the heat conduction equation, as well as mechanical and thermal boundary conditions and initial conditions. A generalized model of physically nonlinear temperature-dependent thermoviscoplasticity is used to describe the behavior of the material. Spatial discretization of the axisymmetric problem of laser pulse loading of the disk is performed using the finite element method. The unsteady process of LTF of the deformed disk configuration is modeled. The final profile of the disk is obtained as a result of the thermally induced state of residual stresses and deformations caused by rapid heating and subsequent gradual cooling of the material in the laser irradiation area. The results confirm that laser thermoforming allows localized shaping of thin-walled parts with reduced material waste, lower energy consumption, and higher flexibility compared to conventional forming methods. In particular, the modeling results align with experimental trends reported in the literature, where bending angle control, curvature reversal, and multi-pass strategies are feasible for both aerospace-scale panels and micro-electromechanical system (MEMS) components.

Key words: laser thermoforming, thin-wall metal elements, thermoviscoplasticity theory, thermomechanical model, finite element simulation, residual stress-strain state.

AMS 2020 classification: 35Q79, 74F05.

Introduction

The demand for lightweight yet high-strength structural elements in mechanical engineering, aerospace, automotive, and microsystems industries continues to grow due to the superior performance of metallic materials such as aluminum, titanium, and their alloys. Traditional forming processes - deep drawing, hot stamping, and superplastic forming - remain widely used for manufacturing thin-walled components of complex geometry. However, these conventional methods suffer from several drawbacks, especially when processing materials with high elasticity and strength, such as titanium alloys (Hu et al., 2016; Langer, Spradlin, & Fitzpatrick, 2016). Laser-based processing offers an alternative tool-free approach that relies on localized, non-contact heating to achieve shape modification of metallic sheets. Among these methods, laser thermoforming (LTF) has proven particularly attractive for the precise bending and shaping of thin-walled elements. By controlled irradiation with pulsed or continuous laser beams, residual thermal stresses are generated, enabling permanent deformation without external forces or dies (Ocaña et al., 2011). Compared to conventional processes, LTF provides high design flexibility, low-cost adaptability, and the possibility to process materials that are otherwise difficult to form mechanically, such as nickel-based superalloys and refractory metals (Hu et al., 2016).

Recent advances in experimental and numerical studies have demonstrated that LTF enables smooth curvature transitions, reversal of bending direction, and multi-scale applications from large aerospace panels to micro-electromechanical systems (MEMS) (Ocaña et al., 2011). Numerical modeling based on coupled thermo-viscoplastic formulations has also become an indispensable tool for predicting deformation, stress distribution, and microstructural changes during pulsed heating and subsequent cooling (Wakchaure, Misra, & Menezes, 2024; Langer, Spradlin, & Fitzpatrick, 2020). In this paper, the interaction between a thermal pulse and a structural element is studied, making use of a dynamic problem statement involving a generalized model of physically nonlinear behavior of materials in a wide temperature range consistent with the thermodynamics of irreversible processes.

1. Problem statement

The formulation of the problem of irradiation of a metal structural element by a short-term thermal pulse based on a coupled model of thermos-inelastic behavior was discussed in the literature (Hu et al., 2016; Ocaña et al., 2011; Zhuk, Senchenkov, & Boichuk, 2009). In this paper, the axisymmetric version of the problem statement is used. Below, the equations of the thermo-viscoplastic model of material behavior, along with the problem equations, are briefly presented. Then, the formulation of the initial and boundary conditions usually used in modeling the LTF process is also listed to complete the problem statement.

Let a metal disk of radius and thickness h be considered. Its geometry in the cylindrical coordinate system $Orz\phi$ is given as follows: $|r| \leq R$, $0 \leq |z| \leq h$.

The formulation of the dynamic axisymmetric coupled problem of thermo-viscoplasticity consists of the equations of kinematics and dynamic equilibrium/motion in terms of the Cauchy stress tensor, the energy balance equation, which is reduced to the heat conduction equation, mechanical and thermal boundary conditions, as well as initial conditions that are listed below

$$\varepsilon_z = \frac{\partial u_z}{\partial z}, \quad \varepsilon_r = \frac{\partial u_r}{\partial r}, \quad \varepsilon_\phi = \frac{u_r}{r}, \quad \varepsilon_{rz} = \frac{1}{2} \left(\frac{\partial u_z}{\partial r} + \frac{\partial u_r}{\partial z} \right), \quad (1)$$

$$\frac{\partial \sigma_r}{\partial r} + \frac{1}{r} (\sigma_r - \sigma_\phi) + \frac{\partial \sigma_{rz}}{\partial z} = \rho \frac{\partial^2 u_r}{\partial t^2}, \quad \frac{\partial \sigma_{rz}}{\partial r} + \frac{1}{r} \sigma_{rz} + \frac{\partial \sigma_z}{\partial z} = \rho \frac{\partial^2 u_z}{\partial t^2}, \quad (2)$$

$$c_v \dot{\theta} + 3\alpha\theta K_V (\dot{\varepsilon}_{kk} - 3\alpha\dot{\theta}) - D' - k\Delta\theta = r_s, \quad (3)$$

$$\sigma_{ij} n_j = 0 \quad \text{on } S, \quad (4)$$

$$-k \frac{\partial \theta}{\partial z} = q_s \quad \text{on } S_p \text{ i } 0 \leq t \leq t_p; \quad \frac{\partial \theta}{\partial n} = 0 \quad \text{on } S - S_p, \quad (5)$$

$$u_r = u_z = \dot{u}_r = \dot{u}_z = 0, \quad \theta = \theta_0 \quad \text{at } t = 0 \quad (6)$$

where u_r , u_z are the displacements along the corresponding axes; ε_r , ε_z , ε_{rz} , and ε_ϕ are the components of the strain tensor; σ_r , σ_z , σ_{rz} , and σ_ϕ are the components of the stress tensor; θ is temperature; α , c_v , and k are coefficients of linear thermal expansion, heat capacity at constant volume, and thermal conductivity, respectively; K_V is volume modulus of the material; r_s is power of the given internal heat sources; D' is mechanical energy dissipation rate; θ_0 is the initial temperature; n_j 's are the component of the outward unit normal vector to the corresponding boundary surface; S_p is part of the surface ($r \leq r_p$, $z = 0$) subjected to the thermal pulse; and $\dot{\varepsilon}_{kk} = \dot{\varepsilon}_r + \dot{\varepsilon}_z + \dot{\varepsilon}_\phi$ is the rate of the dilatation.

To describe the physically nonlinear thermo-viscoplastic behavior of the disk material, a generalized model developed in (Zhuk et al., 2001; Zhuk, Senchenkov, & Boichuk, 2009) is used. It is consistent with the thermodynamics of irreversible processes and consists of a representation of the total strain as a sum of elastic, inelastic, and thermal components,

$$\varepsilon_{ij} = \varepsilon_{ij}^e + \varepsilon_{ij}^p + \varepsilon_{ij}^\theta, \quad ij \leftrightarrow r, z, r\phi, \quad (7)$$

where thermal deformation is represented as:

$$\varepsilon_{ij}^\theta = \delta_{ij} \int_{\theta_0}^{\theta} \alpha(\theta') d\theta'; \quad (8)$$

- Hooke's law written for the deviatoric and volumetric parts of the stress tensor

$$s_{ij} = 2G(e_{ij} - \varepsilon_{ij}^p), \quad \sigma_{kk} = 3K_V(\varepsilon_{kk} - \varepsilon_{kk}^\theta), \quad (9)$$

where s_{ij} and e_{ij} are the deviators of the stress and strain tensors, respectively; G is the shear modulus; the summation over the repeating indices is assumed;

- the flow rule accompanied by the condition of plastic incompressibility

$$\dot{\varepsilon}_{ij}^p = \lambda s_{ij}, \quad \dot{\varepsilon}_{kk}^p = 0; \quad (10)$$

- kinetic equation

$$D_2^p = D_0^2 \exp \left[- \left(Z^2 / 3J_2 \right)^n \right], \quad Z = K + D, \quad J_2 = s_{ij}s_{ij}/2, \quad D_2^p = \dot{\varepsilon}_{ij}^p \dot{\varepsilon}_{ij}^p / 2, \quad \lambda^2 = D_2^p / J_2 \quad (11)$$

where $Z = K + D$, $J_2 = s_{ij}s_{ij}/2$, $D_2^p = \dot{\varepsilon}_{ij}^p \dot{\varepsilon}_{ij}^p / 2$, $\lambda^2 = D_2^p / J_2$;

- and the evolution equations for the internal variables of isotropic and kinematic hardening

$$\dot{K} = m_1(K_1 - K)\dot{W}_p, \quad K(0) = K_0; \quad \dot{\beta}_{ij} = m_2(D_1 u_{ij} - \beta_{ij})\dot{W}_p, \quad \beta_{ij}(0) = 0; \quad D = \beta_{ij}u_{ij}, \quad u_{ij} = \sigma_{ij} / (\sigma_{mn}\sigma_{mn})^{1/2}, \quad \dot{W}_p = \sigma_{ij}\dot{\varepsilon}_{ij}^p. \quad (12)$$

The values D_0 , D_1 , K_0 , K_1 , m_1 , m_2 and n are constants of the constitutive model. The expression for the dissipative function D' used was derived in (Zhuk, Senchenkov, & Boichuk, 2009) in the frame of thermodynamics of irreversible processes

$$D' = \sigma_{ij}\dot{\varepsilon}_{ij}^p - K\dot{\delta} - \beta_{ij}\dot{\alpha}_{ij} = \dot{W}_p - \dot{W}_{sk} - \dot{W}_{s\beta}. \quad (13)$$

The values of δ and α_{ij} are some internal variables conjugate to the thermodynamic forces K and β_{ij} (parameters of isotropic and kinematic hardening, respectively).

We assume that a laser pulse irradiates the center of the surface $z = 0$ over a circular spot of radius r_p . This single thermal pulse is modeled by a specified heat flux q_s of duration t_p across the disk boundary. The time-space law describing the pulse is given as follows:

$$q_s = \begin{cases} q_0 \cos \frac{\pi r}{2r_p} \sin \frac{\pi t}{t_p}; & r \leq r_p, \quad t \leq t_p, \\ 0; & r > r_p, \quad t > t_p. \end{cases} \quad (14)$$

A thermal insulation condition is assumed on the rest of the disk surface. After the pulse has ceased, the irradiated part of the surface is also considered to be thermally insulated. Equations (1)–(13) with initial and boundary conditions (14) constitute the formulation of a coupled problem of thermomechanics of physically nonlinear solids under thermal loading.

2. Material properties and solution technique

Steel 35CrMo was chosen as the disk material. The physical and mechanical properties for this material and their dependencies on temperature were taken from (Senchenkov, & Tabieva, 1996; Zhuk, Senchenkov, & Boichuk, 2009).

The problem statement, as formulated by equations (1)–(14), represents an essentially nonlinear problem which may only be solved numerically. Here, the same approach as developed in (Zhuk, Senchenkov, & Boichuk, 2009; Zhuk et al., 2001) for solving dynamic plane and axisymmetric problems of thermo-viscoplasticity is adopted.

The numerical implementation of this technique is carried out as a double iterative process. The spatial discretization of the problem is carried out by making use of the finite element method (FEM). The calculations were performed for a fine mesh, especially in the irradiated area, to correctly simulate the thermomechanical response caused by large temperature gradients. The application of FEM within the framework of the external iterative process leads to the problem of the dynamics of viscoelastic solids. The second time derivatives in the equations of motion for an arbitrary moment of time are represented by Newmark's formulas. At each time increment, the problem is solved by the iteration method, while each iteration consists of solving a linearized problem of motion.

3. Calculation results and analysis

Calculations were performed for a disk of radius $R = 5 \cdot 10^{-3}$ m and thickness $h = 10^{-4}$ m, which are typical for micromechanical systems. The radius of the irradiation zone was $r_p = 1.5 \cdot 10^{-3}$ m, the pulse durations t_p have been varied from 10^{-8} s to 10^{-7} s, and the heat flux parameter q_0 was also varied from $6 \cdot 10^7$ to $2 \cdot 10^8$ kW/m². The initial temperature θ_0 of the disk was assumed to be the room temperature of 20 °C.

The mechanism of laser thermoforming of thin-walled parts is to create laser-induced thermal residual strain without the use of tools or external forces. When a disk is irradiated, a region of high temperature gradients occurs on the surface and in the near-surface zone, which changes its configuration over time. After the end of the laser pulse, in the absence of heat exchange with the environment, the temperature gradually equalizes over the disk volume. As a result of the fast expansion of the material over the irradiated zone, significant compressive stresses occur, which form a quasi-static component of the stress field. After a significant time has elapsed compared to the pulse duration (for the considered configuration and conditions, calculations showed that the stabilization time can be chosen as $t = 0.4 \cdot 10^{-4}$ s), the compressive stresses disappear, and a region of quasi-static residual tensile stresses is formed in the center of the disk, which can subsequently compromise the strength and durability of the structure.

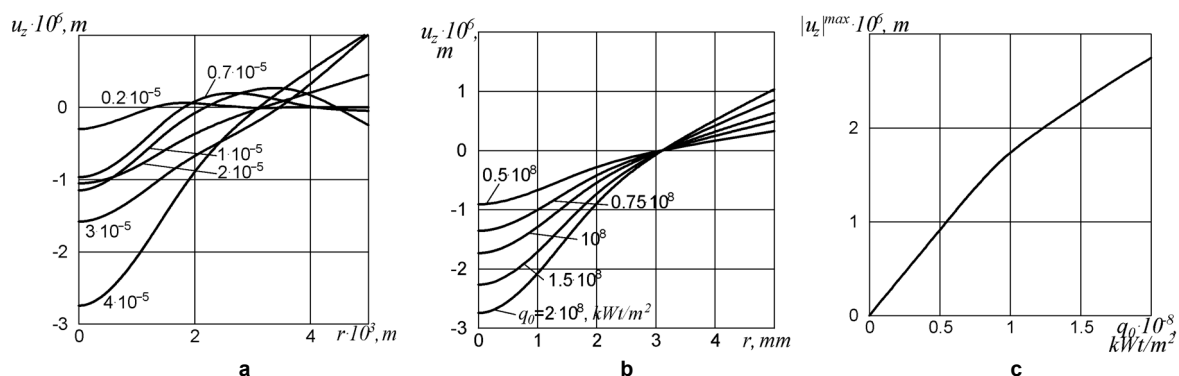


Fig. 1. Results of disk forming simulation

The developed problem statement, the method of its solution, and the obtained modeling results were used to study the applicability of the LTF technology to treat the structural elements by short thermal pulses for the purpose of stamping or forming the desired profile (configuration). The presence of significant inelastic strains induced by thermal processing, which are asymmetric relative to the median surface of the disk, leads to a modification of the surface curvature of metal plates, shells, and structural elements without the use of an external force load. The results of modeling this process are illustrated in Fig. 1.

Figure 1 a shows the evolution of displacements of the surface points of a disk irradiated by a laser pulse for the case of a free edge. The time moments for which the surface profiles were drawn are marked with numbers. The maximum deflections are achieved in the center of the disk. Fig. 1 b shows the residual displacements of the disk surface under the influence of pulses of different power. For higher laser pulse powers, which are simulated by an increase of the heat flux parameter q_0 , an increase in surface deflections is observed both in the center and at the edge. The dependence of the maximum displacement magnitude on the heat flux parameter is shown in Fig. 1 c.

The maximum deflection is always observed in the center, with the bend directed toward the irradiated side. The edge of the disk moves in the opposite direction. An additional factor that affects the quantitative characteristics of the thermoforming is the conditions applied at the disk edge. In the case of a fixed contour, the residual deflections for the given problem parameters are predicted to be 7.5 % smaller than in the case of a free contour.

Discussion and conclusions

The model developed for studying the process of laser thermoforming of a thin-walled steel structural element allows for simulating the complex coupled thermomechanical behavior of the material under the assumptions made. A dynamic formulation of the problem was applied with making use of a generalized model of physically nonlinear response of materials in a wide temperature

range consistent with the thermodynamics of irreversible processes. The obtained quantitative and qualitative results of modeling the laser thermoforming of metal sheet parts show that the creation of laser-induced thermal residual inelastic strains is the main mechanism of shape change. The evolution of inelastic deformation fields in a disk, the dependence of residual inelastic deformation distributions along the radius and thickness of the disk, the effect of laser thermal irradiation level on the magnitudes of induced deflections, and the influence of the conditions of disk contour fixation on the efficiency of thermoforming were studied. It was found that a greater variation of the disk profile can be achieved when the disk edge is free.

Authors' contribution: Yaroslav Zhuk – conceptualization; methodology, software; Arash Soleiman Fallah – analysis of sources, empirical data collection and validation, empirical research, preparation of literature review, and theoretical foundations of research.

Sources of funding. The research was partially funded by Taras Shevchenko National University of Kyiv.

References

- Hu, Y., Luo, M., & Yao, Z. (2016). Increasing the capability of laser peen forming to bend titanium alloy sheets with laser-assisted local heating. *Materials and Design*, 90, 364–372. <https://doi.org/10.1016/j.matdes.2015.10.128>
- Langer, K., Spradlin, T. J., & Fitzpatrick, M. E. (2020). Finite Element Analysis of Laser Peening of Thin Aluminum Structures. *Metals*, 10, 93. <https://doi.org/10.3390/met10010093>
- Ocaña, J. L., Morales, M., Porro, J. A., García-Ballesteros, J. J., & Correa, C. (2011). Laser Shock Microforming of Thin Metal Sheets with ns Lasers. *Physics Procedia*, 12, 201–206. <https://doi.org/10.1016/j.phpro.2011.03.123>
- Senchenkov, I. K., & Tabieva, G. A. (1996). Determination of the parameters of the Bodner-Partom model for thermoviscoplastic deformation of materials. *International Applied Mechanics*, 32, 132–139. <https://doi.org/10.1007/BF02086653>
- Wakchaure, M. B., Misra, M., & Menezes, P. L. (2024). Comprehensive Review on Finite Element Analysis of Laser Shock Peening. *Materials*, 17, 4174. <https://doi.org/10.3390/ma17174174>
- Zhuk, Y. A., Senchenkov, I. K., & Boichuk, O. V. (2009). Residual stress-strain state of a steel disk under thermal pulsed irradiation. *Journal of Mathematical Sciences*, 160, 478–491. <https://doi.org/10.1007/s10958-009-9512-1>
- Zhuk, Y. A., Senchenkov, I. K., Kozlov, V. I., & Tabieva, G. A. (2001). Axisymmetric dynamic problem of coupled thermoviscoplasticity. *International Applied Mechanics*, 37(10), 1311–1317. <https://doi.org/10.1023/A:1013976219144>

Отримано редакцією журналу / Received: 30.08.25

Прорецензовано / Revised: 03.10.25

Схвалено до друку / Accepted: 16.10.25

Ярослав ЖУК, д-р фіз.-мат. наук, проф., чл.-кор. НАН України

ORCID ID: 0000-0002-2726-8395

e-mail: yaroslavzhuk@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Араш Солейман ФАЛЛАХ, д-р філософії, проф.

ORCID ID: 0000-0002-0382-633X

e-mail: arashsol@oslomet.no

Університет Осло Метрополітен, Осло, Норвегія

ТЕХНОЛОГІЯ ЛАЗЕРНОГО ТЕРМОФОРМУВАННЯ ТОНКОСТІННИХ МЕТАЛЕВИХ КОНСТРУКЦІЙНИХ ЕЛЕМЕНТІВ

Запропоновано чисельну модель для моделювання процесу лазерного термоформування (LTF). Вона розроблена на основі термодинамічно узгодженої теорії сполученої термовіскопластичності і підходить для моделювання LTF для тонкостінних металевих конструкційних елементів. У межах цієї моделі постановка задачі складається з рівняння Коші, рівнянь руху та рівняння енергетичного балансу, що зведено до рівняння теплопровідності, а також механічних і теплових граничних умов та початкових умов. Для опису поведінки матеріалу використано узагальнену модель фізично нелінійної температурно-залежної термовіскопластичності. Просторову дискретизацію осесиметричної задачі лазерного імпульсного навантаження диска виконано методом скінченних елементів. Змодельовано нестационарний процес LTF деформованої конфігурації диска. Кінцевий профіль диска отримано внаслідок термічно індукованого стану залишкових напружень і деформацій, викликаного швидким нагріванням і подальшим поступовим охолодженням матеріалу в ділянці, опроміненій лазером.

Ключові слова: лазерне термоформування, тонкостінні металеві елементи, теорія термов'язкопластичності, термомеханічна модель, метод скінченних елементів, залишковий напружено-деформований стан.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 519.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.15>

Oleksiy KAPUSTYAN¹, DSc (Phys. & Math.), Prof.
ORCID ID: 0000-0002-9373-6812
e-mail: alexkap@univ.kiev.ua

Taras YUSYPIV¹, PhD, Assist.
ORCID ID: 0000-0003-2798-9472
e-mail: taras.yusypiv@knu.ua

Yuriy PERESTYUK¹, PhD (Phys. & Math.), Researcher
ORCID ID: 0000-0001-7651-1649
e-mail: perestyuk@gmail.com

Zoia KHALETSKA², PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0001-6272-2189
e-mail: khaletskazoya@gmail.com

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

²Volodymyr Vynnychenko Central Ukrainian State University, Kropyvnytskyi, Ukraine

UNIFORM ATTRACTORS AND ASYMPTOTIC GAIN PROPERTY FOR THE PDE-ODE SYSTEM

The paper investigates the qualitative behavior of weak solutions to a dissipative infinite-dimensional non-autonomous perturbed system, which consists of a parabolic reaction-diffusion system and a nonlinear system of ordinary differential equations. Perturbations are modeled by bounded functions included in the right-hand side of both systems. The considered system is known to possess a global attractor in the absence of perturbations, which determines the long-term dynamics of all trajectories. However, the robustness of such attractors under external disturbances remains a challenging issue, especially in the context of nonlinear and infinite-dimensional systems.

In this study, we adopt an approach based on the theory of uniform attractors for non-autonomous dynamical systems (semi-processes). A corresponding family of semi-processes associated with the perturbed PDE-ODE system is constructed. The existence of a uniform attractor is proven, and its convergence to the global attractor of the unperturbed system is established as the amplitude of disturbances tends to zero. This allows us to derive a nonlocal robust estimate of the asymptotic gain (AG) type, which complements the local ISS-based robustness result and provides an upper bound on the deviation of perturbed trajectories from the unperturbed attractor in terms of the perturbation magnitude. The results contribute to the theoretical foundation for robustness analysis in dissipative infinite-dimensional systems with complex long-term dynamics.

Key words: asymptotic gain, uniform attractor, infinite-dimensional system, PDE-ODE system.

AMS 2020 classification: 34D45.

Introduction

It is known that the global asymptotically stable equilibrium position of a linear system is robust in the sense that for any bounded perturbations, the solution eventually finds itself in an equilibrium position, and the size of this neighborhood depends only on the norm of the perturbations. For nonlinear systems, this is generally not the case (Khalil, 1992). An estimate of the trajectory divergences for such a system in the case of its perturbation can be obtained within the framework of the Input to State Stability (ISS) theory (Sontag, 1998; Atamas, Dashkovskiy, & Slynko, 2023a; Zheng, Zhu, & Dashkovskiy, 2022; Mironchenko, 2023b, 2023a; Mironchenko, Prieur, & Wirth, 2021). The main tool in both the finite-dimensional and infinite-dimensional cases is the ISS Lyapunov function, in terms of which the desired robust estimates are obtained (Sontag, 1998; Dashkovskiy, & Mironchenko, 2013; Kasyanov, & Zgurovsky, 2018a; Atamas, Dashkovskiy, & Slynko, 2023b; Mironchenko, 2023c). However, for many infinite-dimensional dissipative dynamical systems, a characteristic property is the presence of a non-trivial attracting set in the phase space – a global attractor (Temam, 1998; Robinson, 2001; Kasyanov, Toscano, & Zadoianchuk, 2012; Kapustyan et al., 2014; Chueshov, & Rezounenko, 2015; Chueshov, & Ryzhkova, 2012). The transfer of the basic ISS constructions to parabolic and hyperbolic systems with a non-trivial attractor was carried out in work (Kapustyan, & Dashkovskiy, 2022). At the same time, to prove the property of Asymptotic Gain (AG), an approach related to the analysis of uniform attractors of non-autonomous dynamical systems (semi-processes) and their dependence on perturbations turned out to be productive (Kapustyan et al., 2023; Kasyanov, & Zgurovsky, 2018c, 2018b; Zgurovsky et al., 2017; Dashkovskiy, & Slynko, 2022; Xu, Caraballo, & Valero, 2022b, 2022a).

In this work, this approach is applied to the PDE-ODE type system, the robustness of which in the sense of local ISS was established in the article (Kapustyan, & Yusypiv, 2023). Here, a corresponding family of semi-processes will be constructed, the existence of a uniform attractor will be proven, and (based on the fact of its convergence to the attractor of the unperturbed system) a robust estimate will be established in the sense of AG.

1. Problem statement and main result

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. Consider the PDE-ODE system

$$\begin{cases} u_t = A\Delta u - f(u) + B(x)v(t) + d_1(t, x), \\ u|_{\partial\Omega} = 0 \end{cases} \quad (1)$$

$$\frac{dv}{dt} = -F(v) + \int_{\Omega} G(x)u(x,t)dx + d_2(t) \quad (2)$$

where A is given $N \times N$ -matrix for which $\frac{1}{2}(A + A^*) \geq \nu_1 I$, $u = u(x, t) = (u^1, \dots, u^N)$, $v = v(t) = (v^1, \dots, v^M)$ are unknown functions, $B, G \in L^2(\Omega)$ are given matrices of corresponding dimensions, $d_1 \in L^\infty(0, +\infty; (L^2(\Omega))^N)$, $d_2 \in L^\infty(0, +\infty; \mathbb{R}^M)$ are incoming "disturbing" signals, and for all $y, w \in \mathbb{R}^N$, $u \in \mathbb{R}^M$ the following conditions hold:

$$\begin{aligned} f &\in C^1(\mathbb{R}^N; \mathbb{R}^N), F \in C^1(\mathbb{R}^M; \mathbb{R}^M), \\ \sum_{i=1}^N f^i(y)y^i &\geq \nu_2 \cdot \sum_{i=1}^N |y^i|^{p_i} - c_1, \\ \sum_{i=1}^N |f^i(y)|^{\frac{p_i}{p_i-1}} &\leq c_2 \left(\sum_{i=1}^N |y^i|^{p_i} + 1 \right), \\ (Df(u)w, w)_{\mathbb{R}^N} &\geq -c_3 \cdot \|w\|_{\mathbb{R}^N}^2, \\ \sum_{i=1}^M F^i(u)u^i &\geq \nu_3 \|u\|_{\mathbb{R}^M}^2 - c_4, \end{aligned} \quad (3)$$

where $\nu_1, \nu_2, \nu_3, c_1, c_2, c_3, c_4$ are given constants, $p_i \geq 2, i = \overline{1, N}$.

Hereafter, we will use the notation:

$$p = (p_1, \dots, p_N), q = (q_1, \dots, q_N), q_i = \frac{p_i}{p_i - 1},$$

$$L^p(\Omega) = L^{p_1}(\Omega) \times \dots \times L^{p_N}(\Omega),$$

$$L^p(0, T; L^p(\Omega)) = L^{p_1}(0, T; L^{p_1}(\Omega)) \times \dots \times L^{p_N}(0, T; L^{p_N}(\Omega)),$$

$$H = (L^2(\Omega))^N, V = (H_0^1(\Omega))^N$$

$AC([0, T]; \mathbb{R}^M)$ is the space of absolutely continuous functions from $[0, T]$ to $\mathbb{R}^M$.

Using Galerkin approximations (Robinson, 2001; Kapustyan et al., 2014), it is standardly established that under conditions (3), the problem (1),(2) for arbitrary perturbations $d = \{d_1, d_2\} \in L^\infty(0, +\infty; (L^2(\Omega))^N) \times L^\infty(0, +\infty; \mathbb{R}^M)$ and for arbitrary initial data $z_0 = \{u_0, v_0\}$ from the phase space $X = (L^2(\Omega))^N \times \mathbb{R}^M$ has a unique weak solution

$$z = \{u, v\} \in (L^2(0, T; V) \cap L^p(0, T; L^p(\Omega))) \times AC([0, T]; \mathbb{R}^M)$$

on $(0, T)$ with $z(0) = z_0$. Extending each such solution to $[0, +\infty)$, we define

$$S_d(t, z_0) := \{z(t), z(\cdot) \text{ is a solution to (1),(2) with perturbation } d, z(0) = z_0\}.$$

In the paper (Kapustyan, & Yussyipiv, 2023), it was proven that in the absence of perturbations (i.e., $d \equiv 0$), the semigroup S_0 has a global attractor $\Theta \subset X$, meaning there exists a compact set $\Theta \subset X$ that is invariant:

$$S_0(t, \Theta) = \Theta \quad \forall t > 0,$$

and uniformly attracting:

$$\text{dist}(S_0(t, B), \Theta) \rightarrow 0, t \rightarrow \infty \text{ for any bounded } B \subset X,$$

where, here and hereafter,

$$\text{dist}(A, B) = \inf_{a \in A} \sup_{b \in B} \|a - b\|_X.$$

In the presence of perturbations, S_d is no longer a semigroup, and the question arises as to how the deviation of the trajectories of the perturbed system S_d from the set Θ depends on the magnitude

$$\|d\|_\infty := \max\{esssup_{t \geq 0} \|d_1(t)\|_H, esssup_{t \geq 0} \|d_2(t)\|_{\mathbb{R}^M}\}.$$

The deviation of a point $z \in X$ from the set $\Theta \subset X$ will be evaluated using the quantity:

$$\|z\|_\Theta := \inf_{\theta \in \Theta} \|z - \theta\|_X.$$

In the paper (Kapustyan, & Yussyipiv, 2023), by constructing an ISS-Lyapunov function in a neighborhood of the attractor, the local stability of Θ with respect to perturbations d in the local ISS sense was established, that is,

$$\exists r > 0, \exists \beta \in \mathcal{KL}, \exists \gamma \in \mathcal{K} \text{ such that}$$

$$\forall \|z_0\|_\Theta \leq r \quad \forall \|d\|_\infty \leq r, \forall t \geq 0,$$

$$\|S_d(t, z_0)\|_\Theta \leq \beta(\|z_0\|_\Theta, t) + \gamma(\|d\|_\infty),$$

where $\mathcal{K}$ is the class of continuous, strictly increasing functions $\gamma : [0, +\infty) \mapsto [0, +\infty)$ with $\gamma(0) = 0$, and $\mathcal{KL}$ is the class of continuous functions $\beta : [0, +\infty) \times [0, +\infty) \mapsto [0, +\infty)$, such that $\forall s \geq 0 \beta(\cdot, s) \in \mathcal{K}, \forall t \geq 0 \beta(t, s) \rightarrow 0, s \rightarrow +\infty$.

In this paper, we will establish the nonlocal property of AG only under the conditions (3):

$$\begin{aligned} \exists \gamma \in \mathcal{K}, \forall z_0 \in X, \\ \overline{\lim}_{t \rightarrow \infty} \|S_d(t, z_0)\|_{\Theta} \leq \gamma(\|d\|_{\infty}). \end{aligned} \quad (4)$$

To prove (4), we use the semiprocess technique: we immerse problem (1),(2) into the family of semiprocesses $\{(1)_{\sigma}\}_{\sigma \in \Sigma}$, where $(1)_{\{0\}} = (1)$, and Σ is a specially constructed parameter space that depends on the perturbation d . Next, we will prove that the family of semiprocesses $\{(1)_{\sigma}\}_{\sigma \in \Sigma}$ has a uniform attractor Θ_{Σ} and

$$\text{dist}(\Theta_{\Sigma}, \Theta) \rightarrow 0, \|d\|_{\infty} \rightarrow 0.$$

The last convergence just allows us to prove estimate (4).

2. Semiprocesses and their attractors

Let Σ be a translation-invariant set of functional parameters, i.e.

$$\forall h \geq 0, \forall \sigma(\cdot) \in \Sigma, \sigma(\cdot + h) \in \Sigma.$$

Let a given family $S_{\sigma}(t, \tau, x)$, where $\sigma \in \Sigma, t \geq \tau \geq 0, x \in X$.

Definition 1. A family $\{S_{\sigma}\}_{\sigma \in \Sigma}$ is called a family of semiprocesses if for all $\sigma \in \Sigma, x \in X$ the following conditions are satisfied:

$$\begin{aligned} S_{\sigma}(\tau, \tau, x) &= x, \forall \tau \geq 0, \\ S_{\sigma}(t, s, S_{\sigma}(s, \tau, x)) &= S_{\sigma}(t, \tau, x), \forall t \geq s \geq \tau \geq 0, \\ S_{\sigma}(t + h, \tau + h, x) &= S_{\sigma(\cdot + h)}(t, \tau, x), \forall t \geq \tau \geq 0, \forall h \geq 0. \end{aligned}$$

Remark 1. When $\sigma = 0$ (undisturbed case) we have

$$S_0(t_1 + t_2, 0, x) = S_0(t_1, 0, S_0(t_2, 0, x)).$$

In other words, S_0 satisfies the semigroup property.

Let us put

$$S_{\Sigma}(t, 0, x) := \bigcup_{\sigma \in \Sigma} S_{\sigma}(t, 0, x).$$

Definition 2. A compact set $\Theta_{\Sigma} \subset X$ is called a uniform attractor of the family $\{S_{\sigma}\}_{\sigma \in \Sigma}$, if for an arbitrary bounded $B \subset X$

$$\text{dist}(S_{\Sigma}(t, 0, B), \Theta_{\Sigma}) \rightarrow 0, t \rightarrow \infty,$$

and Θ_{Σ} is the minimal closed set satisfying this property.

Lemma 1. (Kapustyan et al., 2014) Let Σ be a compact set, and the following properties hold:

1) exists bounded $B_0 \subset X$ such that $\forall$ bounded $B \subset X$

$$\exists T = T(B), \forall t \geq T, S_{\Sigma}(t, 0, B) \subset B_0;$$

2) $\forall \{\sigma_n\} \subset \Sigma$, bounded $\{x_n\} \subset X, \{t_n\} \nearrow \infty$

the sequence $\{S_{\sigma_n}(t, 0, x_n)\}$ is precompact in X .

Then $\{S_{\sigma}\}_{\sigma \in \Sigma}$ has a uniform attractor $\Theta_{\Sigma} \subset B_0$.

If an additional condition is met:

3) $\forall t > 0$ mapping $(x, \sigma) \rightarrow S_{\sigma}(t, 0, x)$ is continuous,

then Θ_{Σ} is invariant in the following sense:

$$\Theta_{\Sigma} \subset S_{\Sigma}(t, 0, \Theta_{\Sigma}), \forall t \geq 0.$$

Remark 2. For $\Sigma = \{0\}$ conditions 1)-3) guarantee the existence of a global attractor Θ for the semigroup S_0 .

Lemma 2. (Kapustyan, & Dashkovskiy, 2022) Let $\Sigma = \Sigma(d), \|d\|_{\infty} \leq R$ satisfies the conditions of the Lemma 1 with the same set $B_0, \Sigma(0) = \{0\}, d \in \Sigma(d)$ and the following conditions are fulfilled:

1) $\forall \|d_n\|_{\infty} \rightarrow 0, \forall \{x_n\} \subset B_0, \forall t_n \nearrow \infty, \forall \sigma_n \in \Sigma(d_n)$

sequence $\{S_{\sigma_n}(t_n, 0, x_n)\}$ is precompact in X ;

2) $\forall t > 0, \forall \|d_n\|_{\infty} \rightarrow 0, \forall \sigma_n \in \Sigma(d_n), \forall x_n \rightarrow x$ up to a subsequence

$$S_{\sigma_n}(t, 0, x_n) \rightarrow S_0(t, 0, x).$$

Then $\exists \gamma \in K, \forall x \in X, \forall \|d\|_\infty \leq R$

$$\overline{\lim}_{t \rightarrow \infty} \text{dist}(S_{\Sigma(d)}(t, 0, x), \Theta) \leq \gamma(\|d\|_\infty). \quad (5)$$

Remark 3. Since $d \in \Sigma(d)$, then (5) implies

$$\overline{\lim}_{t \rightarrow \infty} \|S_d(t, 0, x)\|_\Theta \leq \gamma(\|d\|_\infty).$$

3. Uniform attractors of the semiprocesses generated by (1),(2)

Let $d \in L^\infty(\mathbb{R}_+; X)$ with $\|d\|_\infty \leq R_0$. We consider the set

$$\Sigma(d) = cl_{L_{loc}^{2,w}(\mathbb{R}_+; X)} \{d(\cdot + \tau) \mid \tau \geq 0\},$$

where cl means the closure, and $L_{loc}^{2,w}(\mathbb{R}_+; X)$ is the space of locally integrable X -valued functions supplied with the local weak convergence topology. It is known that $\Sigma(d)$ is translation-invariant and compact in $L_{loc}^{2,w}(\mathbb{R}_+; X)$. Moreover,

$$\forall \sigma \in \Sigma(d) \quad \|\sigma\|_+^2 := \sup_{t \geq 0} \int_t^{t+1} \|\sigma(s)\|_X^2 ds \leq \|d\|_\infty^2. \quad (6)$$

As in the case of bounded perturbations, using Galerkin approximations (Robinson, 2001; Kapustyan et al., 2014), it is standardly established that under conditions (3), the problem (1),(2) for arbitrary perturbations of the form $\sigma = \{\sigma_1, \sigma_2\} \in L_{loc}^2(0, +\infty; X)$ and for arbitrary $\tau \geq 0$ and $z_\tau \in X$ has a unique weak solution on $[\tau, +\infty)$, satisfying the condition $z(\tau) = z_\tau$. Moreover, the family of mappings

$$S_\sigma(t, \tau, z_\tau) := \{z(t), z(\cdot)\} \text{ is solution to (1),(2) on } (\tau, +\infty) \text{ with perturbation } \sigma \in \Sigma(d), z(\tau) = z_\tau\}$$

satisfies the conditions of Definition 1, thus forming a family of semi-processes.

Theorem 1. For every perturbation d , $\|d\|_\infty \leq R_0$, the family of mappings $\{S_\sigma\}_{\sigma \in \Sigma(d)}$ possesses the uniform attractor $\Theta_{\Sigma(d)}$, and

$$\Theta_{\Sigma(d)} \subset S_{\Sigma(d)}(t, 0, \Theta_{\Sigma(d)}) \quad \forall t \geq 0.$$

Proof. Let us verify the fulfillment of the conditions of Lemma 1. To this end, we establish some a priori estimates for $z(t) = \{u(t), v(t)\} = S_\sigma(t, 0, z_0)$, where $\sigma = \{\sigma_1, \sigma_2\} \in \Sigma(d)$ satisfies (6). Since for $u \in L^2(0, T; V) \cap L^p(0, T; L^p(\Omega))$, $v \in C([0, T]; \mathbb{R}^M)$, by virtue of (3),

$$A\Delta u - f(u) + B \cdot v + D \cdot d \in L^2(0, T; V^*) + L^q(0, T; L^q(\Omega)) + L^\infty(0, T; L^2(\Omega)) \subset L^2(0, T; V^*) + L^q(0, T; L^q(\Omega)),$$

then, according to (Robinson, 2001), $z : [0, T] \mapsto X$ is an absolutely continuous function, and after multiplying (1) by u scalarly in H , and (2) by v scalarly in $\mathbb{R}^m$, we obtain: for almost all (a.e.) $t > 0$

$$\frac{1}{2} \frac{d}{dt} \|u(t)\|_H^2 + \nu_1 \|u(t)\|_V^2 + \nu_2 \|u(t)\|_{L^p}^p \leq c_1 |\Omega| + \|B\|_{L^2} \cdot \|v(t)\|_{\mathbb{R}^M} + \|\sigma_1(t)\|_H \|u(t)\|_H, \quad (7)$$

$$\frac{1}{2} \frac{d}{dt} \|v(t)\|_{\mathbb{R}^M}^2 + \nu_3 \|v(t)\|_{\mathbb{R}^M}^2 \leq c_4 + \|G\|_{L^2} \cdot \|u(t)\|_H + \|\sigma_2(t)\|_{\mathbb{R}^M} \|v(t)\|_{\mathbb{R}^M}. \quad (8)$$

From (7), (8), Young's inequality, and the Poincaré inequality, we deduce that there exist $\nu > 0$, $c > 0$, depending only on the constants of the problem, such that for a.a. $t > 0$

$$\begin{aligned} \frac{d}{dt} (\|u(t)\|_H^2 + \|v(t)\|_{\mathbb{R}^M}^2) + \nu (\|u(t)\|_V^2 + \|u(t)\|_{L^p}^p + \|u(t)\|_H^2 + \|v(t)\|_{\mathbb{R}^M}^2) \leq \\ \leq c(1 + \|\sigma_1(t)\|_H^2 + \|\sigma_2(t)\|_{\mathbb{R}^M}^2). \end{aligned}$$

Next, using the inequality

$$\int_0^t \|h(s)\|^2 e^{-a(t-s)} ds \leq (1 - e^{-a})^{-1} \sup_{t \geq 0} \int_t^{t+1} \|h(s)\|^2 ds,$$

and due to the continuity of $z \cdot = \{u(\cdot), v(\cdot)\} : [0, +\infty) \mapsto X$, we deduce:

$$\forall t \geq 0 \quad \|z(t)\|_X^2 \leq \|z_0\|_X^2 e^{-\nu t} + \frac{c}{\nu} + c\|d\|_\infty^2. \quad (9)$$

From (9), we obtain condition 1) of Lemma 1 with the set

$$B_0 = \{z \in X \mid \|z\|_X^2 \leq 1 + \frac{c}{\nu} + cR_0^2\}.$$

Let us verify condition 2): the precompactness of the sequence $\{\xi_n = S_{\sigma_n}(t_n, 0, z_0^n)\}$, where $\{\sigma_n\} \subset \Sigma(d)$, $t_n \rightarrow +\infty$,

and $\{z_0^n\}$ is bounded in X .

Since, by the properties of the semiprocesses and condition 1), for sufficiently large $n \geq 1$, we have

$$\xi_n = S_{\sigma_n}(t_n - 1 + 1, t_n - 1, S_{\sigma_n}(t_n - 1, 0, z_0^n)) = S_{\sigma_n(\cdot + t_n - 1)}(1, 0, S_{\sigma_n}(t_n - 1, 0, z_0^n)) \in S_{\Sigma}(1, 0, B_0),$$

it is sufficient to show the precompactness of the $\{S_{\sigma_n}(1, 0, z_0^n)\}$, where $\{\sigma_n\} \subset \Sigma(d)$, $\|z_0^n\|_X \leq r$.

Let $z_n(t) = \{u_n(t), v_n(t)\} = S_{\sigma_n}(t, 0, z_0^n)$, $\|z_0^n\|_X \leq r$.

The boundedness, and hence the precompactness, of the $\{v_n(1)\}$ in $\mathbb{R}^M$ follows from (9).

Let us prove that

$$\|u_n(1)\|_V \leq K(r) \quad (10)$$

for some constant $K(r)$ independent of n . Then, by the compactness of the embedding $V \subset H$, we obtain the precompactness of the $\{u_n(1)\}$ in H .

Let us derive (10) following the arguments from (Temam, 1998), which are justified by passing to Galerkin approximations of problem (1),(2). We multiply equation (1) by Δu_n in H . Taking into account conditions (3) and the smoothness of f , we obtain the inequality:

$$\frac{1}{2} \frac{d}{dt} \|u_n(t)\|_V^2 + \nu_1 \|\Delta u_n(t)\|_H^2 \leq c_3 \|u_n(t)\|_V^2 + \|B\|_{L^2} \|v_n(t)\|_{\mathbb{R}^M} \|\Delta u_n(t)\|_H + \|\sigma_1^n(t)\|_H \|\Delta u_n(t)\|_H. \quad (11)$$

Since, by (9)

$$\sup_{t \geq 0} \|v_n(t)\|_{\mathbb{R}^M} \leq \sqrt{r^2 + \frac{c}{\nu} + cR_0^2},$$

we deduce from (11):

$$\frac{d}{dt} \|u_n(t)\|_V^2 + \lambda \|u_n(t)\|_V^2 \leq c(r)(\|u_n(t)\|_V^2 + 1 + \|\sigma_1^n(t)\|_H^2) \quad (12)$$

for some positive constants λ and $c(r)$.

Multiplying (12) by $te^{\lambda t}$ yields:

$$\frac{d}{dt} (t \|u_n(t)\|_V^2 e^{\lambda t}) \leq c(r)(\|u_n(t)\|_V^2 + 1 + \|\sigma_1^n(t)\|_H^2) te^{\lambda t} + \|u_n(t)\|_V^2 e^{\lambda t}.$$

Integrating the last inequality over $[0, 1]$, we obtain:

$$\|u_n(1)\|_V^2 \leq (c(r) + 1) \int_0^1 \|u_n(t)\|_V^2 dt + c(r)R_0^2. \quad (13)$$

From estimates (7) and (9) we have:

$$\nu_1 \int_0^1 \|u_n(t)\|_V^2 dt \leq \frac{1}{2} r^2 + c_1 |\Omega| + (\|B\|_{L^2} + R_0^2) \sqrt{r^2 + \frac{c}{\nu} + R_0^2}. \quad (14)$$

Combining (13) and (14), we derive estimate (10).

Thus, the family of semiprocesses $\{S_{\sigma}\}_{\sigma \in \Sigma(d)}$ possesses the uniform attractor $\Theta_{\Sigma(d)}$.

The theorem is proved. $\square$

4. Robust stability of the global attractor of the system (1),(2)

Now we turn to the formulation of the main result regarding the robust stability of the global attractor.

Theorem 2. *The attractor Θ of the unperturbed system (1),(2) (when $d \equiv 0$) is stable in the sense of AG, i.e., $\exists \gamma \in \mathcal{K}$ such that $\forall z_0 \in X \forall \|d\|_{\infty} \leq R_0$*

$$\lim_{t \rightarrow \infty} \|S_d(t, z_0)\|_{\Theta} \leq \gamma(\|d\|_{\infty}),$$

where we denote $S_d(t, z_0) = S_d(t, 0, z_0)$.

Proof. Suppose that the limit equality holds

$$\text{dist}(\Theta_{\Sigma(d)}, \Theta) \rightarrow 0, \|d\|_{\infty} \rightarrow 0. \quad (15)$$

Let us prove that (15) implies (4). Indeed, by construction $\Sigma(0) = \{0\}$ and $d \in \Sigma(d)$. Therefore, for $z_0 \in X$, $z \in \Theta_{\Sigma(d)}$, $t > 0$, $z(t) \in S_d(t, z_0)$ we have the following implications

$$\begin{aligned} \|z(t) - \theta\|_X &\leq \|z(t) - z\|_X + \|z - \theta\|_X \Rightarrow \\ &\Rightarrow \inf_{\theta \in \Theta} \|z(t) - \theta\|_X \leq \|z(t) - z\|_X + \inf_{\theta \in \Theta} \|z - \theta\|_X \Rightarrow \\ &\Rightarrow \inf_{\theta \in \Theta} \|z(t) - \theta\|_X \leq \inf_{z \in \Theta_{\Sigma(d)}} \|z(t) - z\|_X + \sup_{z \in \Theta_{\Sigma(d)}} \inf_{\theta \in \Theta} \|z - \theta\|_X \Rightarrow \end{aligned}$$

$$\Rightarrow \|z(t)\|_{\Theta} \leq \text{dist}(z(t), \Theta_{\Sigma(d)}) + \text{dist}(\Theta_{\Sigma(d)}, \Theta) \Rightarrow$$

$$\Rightarrow \|S_d(t, z_0)\|_{\Theta} \leq \text{dist}(S_{\Sigma(d)}(t, z_0), \Theta_{\Sigma(d)}) + \text{dist}(\Theta_{\Sigma(d)}, \Theta).$$

Since $\Theta_{\Sigma(d)}$ is the uniform attractor, then from Definition 2 for each d we have:

$$\text{dist}(S_{\Sigma(d)}(t, z_0), \Theta_{\Sigma(d)}) \rightarrow 0, t \rightarrow +\infty.$$

Let us put $\gamma(s) := \sup_{\|d\|_{\infty} \leq s} \text{dist}(\Theta_{\Sigma(d)}, \Theta) + s$. Then $s \mapsto \gamma(s)$ is increasing and $\gamma(s) \rightarrow 0$ as $s \rightarrow 0+$. Therefore (Clarke, Ledyev, & Stern, 1998), Lemma 5 implies the existence of $\gamma_1 \in K$ such that $\text{dist}(\Theta_{\Sigma(d)}, \Theta) \leq \gamma_1(\|d\|_{\infty})$, which means that (4) is satisfied.

So, according to the previous arguments, we need to verify assumptions 1), 2) of Lemma 2.

Since $\forall \|d_n\|_{\infty} \rightarrow 0, \forall \sigma_n \in \Sigma(d_n)$ satisfies $\|\sigma_n\|_+ \leq \|d_n\| \leq R_0$, condition 1) follows from the arguments of Theorem 1 and the estimate (10).

Now consider $z_1 - z_2$, where $z_1 = \{u_1, v_1\}, z_2 = \{u_2, v_2\}$ are solutions to (1),(2) with perturbation $d = \{d_1, d_2\}, \|d\|_{\infty} \rightarrow 0$, and without perturbations, respectively. Using the estimate for each solution

$$\sup_{t \geq 0} \|z(t)\|_X^2 \leq \|z(0)\|_X^2 + \frac{c}{\nu} + cR_0^2,$$

we derive for a.a. $t > 0$:

$$\begin{aligned} & \frac{1}{2} \frac{d}{dt} \|u_1(t) - u_2(t)\|_H^2 + \nu_1 \|u_1(t) - u_2(t)\|_V^2 \leq c_3 \|u_1(t) - u_2(t)\|_H^2 + \\ & + \|B\|_{L^2} \|v_1(t) - v_2(t)\|_{\mathbb{R}^M} \|u_1(t) - u_2(t)\|_H + \|d_1(t)\|_H \|u_1(t) - u_2(t)\|_H, \\ & \frac{1}{2} \frac{d}{dt} \|v_1(t) - v_2(t)\|_{\mathbb{R}^M}^2 \leq \sup_{\|v\|_{\mathbb{R}^M} \leq r_0} \|DF(v)\| \cdot \|v_1(t) - v_2(t)\|_{\mathbb{R}^M}^2 + \\ & + \|G\|_{L^2} \|u_1(t) - u_2(t)\|_H \|v_1(t) - v_2(t)\|_{\mathbb{R}^M} + \|d_2(t)\|_{\mathbb{R}^M} \|v_1(t) - v_2(t)\|_{\mathbb{R}^M}, \end{aligned}$$

where $r_0 = \sqrt{\max\{\|z_1(0)\|_X^2, \|z_2(0)\|_X^2\} + \frac{c}{\nu} + cR_0^2}$.

Thus, with some constant $c(r_0) > 0$, we obtain

$$\forall t \geq 0 \|z_1(t) - z_2(t)\|_X^2 \leq e^{c(r_0)t} (\|z_1(0) - z_2(0)\|_X^2 + \|d\|_{\infty}^2).$$

From the last inequality, the desired property 2) of Lemma 2 follows.

The theorem is proved. □

Discussions and conclusions

This study provides important insights into the behavior of complex, infinite-dimensional PDE-ODE systems under external disturbances. We constructed a family of semiprocesses for the perturbed system and proved the existence of a uniform attractor, showing that it approaches the global attractor of the unperturbed system as the disturbance size decreases. This establishes a nonlocal asymptotic gain property, which measures how far perturbed trajectories deviate from the unperturbed attractor. Building on earlier local stability results, our findings use mathematical techniques like Galerkin approximations to confirm the system's robustness, enhancing our understanding of its long-term dynamics under perturbations.

These results are valuable for applications like reaction-diffusion processes, where such systems are common. The AG property offers a way to predict and control system behavior despite external influences, aiding in the design of effective control strategies. Future work could explore systems with time-varying or random disturbances and test these findings in practical settings. The methods used here, combining semiprocesses and stability analysis, can also be applied to other complex systems, opening new avenues for research into their robustness and stability.

Authors' contribution: Oleksiy, Kapustyan — conceptualization and methodology; Taras, Yusypiv — preparation of theoretical foundations of research and empirical research; Yuriy, Perestyuk — empirical data collection, validation; Zoia, Khaletska — analysis of sources, preparation of literature review.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Atamas, I., Dashkovskiy, S., & Slynko, V. (2023a). Impulsive input-to-state stabilization of an ensemble. *Set-Valued and Variational Analysis*, 31, 1–11. <https://doi.org/10.1007/s11228-023-00688-x>
- Atamas, I., Dashkovskiy, S., & Slynko, V. (2023b). Lyapunov functions for linear hyperbolic systems. *IEEE Transactions on Automatic Control*, 68, 6496–6508. <https://doi.org/10.1109/TAC.2023.3247879>
- Chueshov, I., & Rezounenko, A. (2015). Finite-dimensional global attractors for parabolic nonlinear equations with state-dependent delay. *Communications on Pure and Applied Analysis*, 14, 1685–1704. <https://doi.org/10.3934/cpaa.2015.14.1685>
- Chueshov, I., & Ryzhkova, I. (2012). A global attractor for a fluid-plate interaction model. *Communications on Pure and Applied Analysis*, 14, 1635–1656. <https://doi.org/10.3934/cpaa.2013.12.1635>

- Clarke, F., Ledyae, Y., & Stern, R. (1998). Asymptotic stability and smooth lyapunov functions. *Journal of Differential Equations*, 149, 69–114. <https://doi.org/10.1006/jdeq.1998.3476>
- Dashkovskiy, S., & Mironchenko, A. (2013). Input-to-state stability of infinite-dimensional control systems. *Mathematics of Control, Signals, and Systems*, 25, 1–35. <https://doi.org/10.1007/s00498-012-0090-2>
- Dashkovskiy, S., & Slynko, V. (2022). Stability conditions for impulsive dynamical systems. *Mathematics of Control, Signals, and Systems*, 34, 95–128. <https://doi.org/10.1007/s00498-021-00305-y>
- Kapustyan, O., & Dashkovskiy, S. (2022). Robustness of global attractors: abstract framework and application to dissipative wave equation. *SIAM Journal on Mathematical Analysis*, 11(5), 1565–1577. <https://doi.org/10.3934/eect.2021054>
- Kapustyan, O., Kasyanov, P., Valero, J., & Zgurovsky, M. (2014). Structure of uniform global attractor for general non-autonomous reaction-diffusion system. *Continuous and distributed systems*, 211, 221–237. https://doi.org/10.1007/978-3-319-03146-0_12
- Kapustyan, O., Kurylko, O., Yushpiv, T., & Pankov, A. (2023). Stability w.r.t. disturbances for the global attractor of multi-valued semiflow generated by nonlinear wave equation. *Journal of optimization, differential equations and their applications*, 31(1), 111–124. <https://doi.org/10.15421/142306>
- Kapustyan, O., & Yushpiv, T. (2023). Stability under perturbations for the attractor of a dissipative pdf-odf-type system. *Journal of Mathematical Sciences*, 272(2), 236–243. <https://doi.org/10.1007/s10958-023-06413-1>
- Kasyanov, P., Toscano, L., & Zadoianchuk, N. (2012). Long-time behaviour of solutions for autonomous evolution hemivariational inequality with multidimensional reaction-displacement law. *Abstract and Applied Analysis*, 2012(3), 450984. <https://doi.org/10.1155/2012/450984>
- Kasyanov, P., & Zgurovsky, M. (2018a). *Indirect lyapunov method for autonomous dynamical systems*. Studies in Systems, Decision and Control. https://doi.org/10.1007/978-3-319-59840-6_9
- Kasyanov, P., & Zgurovsky, M. (2018b). *Uniform global attractors for non-autonomous dissipative dynamical systems*. Studies in Systems, Decision and Control. https://doi.org/10.1007/978-3-319-59840-6_7
- Kasyanov, P., & Zgurovsky, M. (2018c). *Uniform trajectory attractors for non-autonomous nonlinear systems*. Studies in Systems, Decision and Control. https://doi.org/10.1007/978-3-319-59840-6_8
- Khalil, K. (1992). *Nonlinear systems*. Macmillan Publishing Company.
- Mironchenko, A. (2023a). *Input-to-state stability*. Communications and Control Engineering. https://doi.org/10.1007/978-3-031-14674-9_2
- Mironchenko, A. (2023b). *Integral input-to-state stability*. Communications and Control Engineering. https://doi.org/10.1007/978-3-031-14674-9_4
- Mironchenko, A. (2023c). Lyapunov criteria for robust forward completeness of distributed parameter systems. *Systems and Control Letters*, 180(4), 105618. <https://doi.org/10.1016/j.sysconle.2023.105618>
- Mironchenko, A., Prieur, C., & Wirth, F. (2021). Local stabilization of an unstable parabolic equation via saturated controls. *IEEE Transactions on Automatic Control*, 66, 2162–2176. <https://doi.org/10.1109/TAC.2020.3007733>
- Robinson, J. (2001). *Infinite-dimensional dynamical systems*. Cambridge University Press.
- Sontag, E. (1998). *Mathematical control theory: deterministic finite-dimensional systems*. Springer. <https://doi.org/10.1007/978-1-4612-0577-7>
- Temam, R. (1998). *Infinite-dimensional dynamical systems in mechanics and physics*. Springer. <https://doi.org/10.1007/978-1-4612-0645-3>
- Xu, J., Caraballo, T., & Valero, J. (2022a). Asymptotic behavior of a semilinear problem in heat conduction with long time memory and non-local diffusion. *Journal of Differential Equations*, 327, 418–447. <https://doi.org/10.1016/j.jde.2022.04.033>
- Xu, J., Caraballo, T., & Valero, J. (2022b). Asymptotic behavior of nonlocal partial differential equations with long time memory. *Discrete and Continuous Dynamical Systems - Series S*, 15, 3059–3078. <https://doi.org/10.3934/dcdss.2021140>
- Zgurovsky, M., Gluzman, M., Gorban, N., Kasyanov, P., Paliichuk, L., & Khomenko, O. (2017). Uniform global attractors for non-autonomous dissipative dynamical systems. *Discrete and Continuous Dynamical Systems*, 22, 2053–2065. <https://doi.org/10.3934/dcdsb.2017120>
- Zheng, J., Zhu, G., & Dashkovskiy, S. (2022). Relative stability in the sup-norm and input-to-state stability in the spatial sup-norm for parabolic PDEs. *IEEE Transactions on Automatic Control*, 67, 5361–5375. <https://doi.org/10.1109/TAC.2022.3192325>

Отримано редакцією журналу / Received: 24.05.25
Прорецензовано / Revised: 24.09.25
Схвалено до друку / Accepted: 21.10.25

Олексій КАПУСТЯН¹, д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0002-9373-6812
e-mail: alexkap@univ.kiev.ua

Тарас ЮСИПІВ¹, д-р філософії, асист.
ORCID ID: 0000-0003-2798-9472
e-mail: taras.yusypiv@knu.ua

Юрій ПЕРЕСТЮК¹, канд. фіз.-мат. наук, наук. співроб.
ORCID ID: 0000-0001-7651-1649
e-mail: perestyuk@gmail.com

Зоя ХАЛЕЦЬКА², канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0001-6272-2189
e-mail: khaletskazoya@gmail.com

¹Київський національний університет імені Тараса Шевченка, Київ, Україна

²Центральноукраїнський державний університет імені Володимира Винниченка, Кропивницький, Україна

РІВНОМІРНІ АТРАКТОРИ ТА ВЛАСТИВІСТЬ АСИМПТОТИЧНОГО ПІДСИЛЕННЯ ДЛЯ СИСТЕМИ ТИПУ PDE-ODE

У статті досліджено якісну поведінку слабких розв'язків нескінченновимірної дисипативної неавтономної системи із зовнішніми збуреннями, що складається з параболічної системи типу реакції-дифузії та нелінійної системи звичайних диференціальних рівнянь. Збурення моделюються обмеженими функціями, що входять у праві частини обох підсистем. Відомо, що в незбуреному випадку така система має глобальний аттрактор, який визначає довготривалу динаміку всіх траєкторій. Водночас питання робастності таких аттракторів за наявності зовнішніх збурень залишається складним, особливо у випадку нелінійних та нескінченновимірних систем.

У запропонованому дослідженні застосовано підхід, що базується на теорії рівномірних аттракторів для неавтономних динамічних систем (теорія напівпроцесів). Побудовано відповідну сім'ю напівпроцесів, що відповідає збуреній системі типу PDE-ODE. Доведено

існування рівномірного атрактора, а також встановлено його збіжність до глобального атрактора незбуреної системи за наближення амплітуди збурень до нуля. Це дає змогу вивести нелокальну робастну оцінку типу асимптотичного підсилення (AG), що доповнює робастну оцінку в сенсі локальної ISS і забезпечує верхню межу відхилення збурених траєкторій від незбуреного атрактора в термінах норми збурень. Отримані результати сприяють розвитку теоретичних засад аналізу робастності дисипативних нескінченновимірних систем зі складною довготривалою поведінкою.

Ключові слова: асимптотичне підсилення, рівномірний атрактор, нескінченновимірна система, система типу PDE-ODE.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

УДК 534.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.16>

Володимир КУЗЬМЕНКО, студ.
ORCID ID: 0009-0005-7922-430X
e-mail: vovakuzmenko0200@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

Ярослав ТРОЦЕНКО, д-р філософії, асист.
ORCID ID: 0009-0000-8387-9118
e-mail: yaroslav.trotsenko@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

МОДЕЛЮВАННЯ ПОШИРЕННЯ ЗВУКУ У ПЛОСКОМУ ХВИЛЕВОДІ З ОКРУГЛИМИ ВИГИНАМИ

в' Актуальність проведеного дослідження зумовлена практичною значущістю хвильоводів складної геометрії та суттєвою залежністю характеристик поширення хвиль у них від фізичних і геометричних параметрів системи. Мета роботи – визначення закономірностей поширення звукових хвиль у плоскому хвильоводі, заповненому ідеальною стислою рідиною, який містить два округлі вигини. На основі методу часткових областей побудовано чисельно-аналітичну модель, що дає змогу досліджувати поведінку акустичних хвиль у хвильоводі залежно від обраних параметрів. Виконано аналіз особливостей проходження хвилі крізь геометричні неоднорідності залежно від хвильового розміру для двох варіантів значень кутів вигинів. Обчислено енергетичні коефіцієнти проникнення звукової хвилі крізь вигини, а також коефіцієнти збурення однорідних мод у широкому діапазоні параметрів. Проведено верифікацію отриманих результатів шляхом контролю виконання закону збереження енергії. Похибка обчислення зростає зі збільшенням хвильового розміру та кутів вигинів, проте може бути зменшена завдяки врахуванню більшої кількості неоднорідних мод. Показано, що зародження вищої однорідної моди пов'язане зі зменшенням коефіцієнта проникнення. При цьому можливе повне запирання потоку енергії хвилі, що відповідає зародженню першої моди. Для кутів вигинів 45° домінують нульова мода майже на всьому діапазоні хвильових розмірів. Для кутів вигинів 135° виявлено додаткові області зниження коефіцієнта проникнення, що зумовлені резонансними ефектами геометрії хвильоводу. У цьому випадку на першій половині розглянутого діапазону хвильових розмірів домінує нульова мода, а на другій – перша. Отримані результати можуть бути використані під час проєктування повітряних звукопроводів, частотно-селективних пристроїв і систем шумопоглинання. Перспективним напрямом подальших досліджень є аналіз механізмів перетворення енергії між модами, що відкриває можливості для керування просторовою структурою акустичного поля.

Ключові слова: акустична хвиля, хвильовід, звукове поле, однорідні моди, неоднорідні моди, енергетичні коефіцієнти, метод часткових областей.

Класифікація відповідно до AMS 2020: 34B60, 65L10.

Вступ

Сучасна акустика є однією із ключових галузей фізики, а її прикладні аспекти охоплюють широкий спектр практичних застосувань. Особливе значення мають хвильоводи, геометрія яких у багатьох інженерних і наукових задачах є складною, що суттєво впливає на характер поширення хвиль (Grinchenko, Vovk, & Matsypura, 2018). Зокрема, у роботі (Маципура, Вовк, & Трунов, 2013) встановлено, що власні форми однорідних мод у криволінійних і плоскопаралельних хвильоводах можуть суттєво відрізнятися. Крім того, у дослідженні (Трунов, 2014) показано, що структура поля за вигином хвильоводу значно залежить від його геометричних параметрів. У зв'язку із цим виникає потреба в детальній аналізі закономірностей поширення звукових хвиль у хвильоводах складної форми.

Одним з ефективних сучасних підходів до аналізу хвильових процесів у складних геометріях є метод часткових областей. Наприклад, у роботі (Naida et al., 2023) продемонстровано ефективність розширення цього методу для розв'язання складних просторових задач із практичними застосуваннями, зокрема, у біомедицинській інженерії. У дослідженні (Грінченко, & Маципура, 2025) продемонстровано можливість одночасного використання різних форм загального розв'язку в окремих часткових областях – як аналітичних, так і чисельних – з метою зменшення обсягів обчислень за збереження точності результатів.

Об'єктом пропонованого дослідження є акустичне поле в нескінченному плоскому хвильоводі з двома округлими вигинами, заповненому ідеальною стислою рідиною. Мета роботи – побудова чисельно-аналітичного розв'язку задачі поширення звукових хвиль у такому хвильоводі із застосуванням методу часткових областей, а також визначення залежностей між характеристиками акустичного поля та параметрами хвильоводу.

1. Постановка задачі та математична модель

Авторами розглянуто задачу про поширення акустичних хвиль у нескінченному плоскому хвильоводі з двома округлими вигинами (рис. 1). Величини вигинів визначено кутами α та β . Хвильовід заповнений ідеальною стислою рідиною, поверхні вважалися абсолютно жорсткими. Втрати звукової енергії у середовищі відсутні.

Дослідження акустичного поля у хвильоводі проводили на основі методу часткових областей. Загальну область існування звукового поля природно розділити на п'ять областей, де області I та V – це напівнескінченні хвильоводи з висотою h , область III – скінченний хвильовід з висотою h та довжиною l , області II і IV – обмежені дугами кіл з радіусами R , $R + h$ та R' , $R' + h$ відповідно. Уводилися чотири декартові системи координат xOy , $x'Oy'$, $x''O'y''$ та $x'''O'y'''$ так, щоб їхні осі ординат були паралельними до поверхонь областей I, III та V, а також дві полярні системи координат із центрами в точках O та O' .

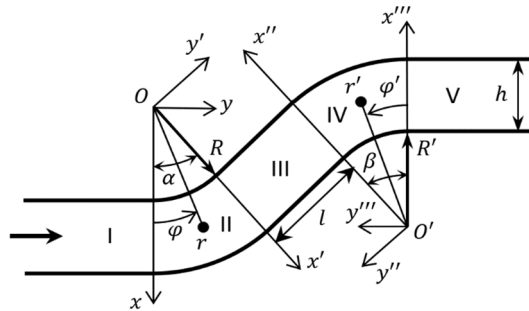


Рис. 1. Схема хвильоводу з вигинами

Нехай ліворуч у хвильоводі поширюється гармонічна плоска хвиля з одиничною амплітудою тиску $p_0 = \exp \exp(iky)$, де $k = 2\pi/\lambda$ – хвильове число, λ – довжина хвилі. Часовий множник $\exp(-i\omega t)$, де ω – частота хвилі, не враховуємо для скорочення викладок. Для аналізу акустичного поля хвильоводу необхідно представити функції тиску p в кожній частковій області так, щоб вони задовольняли рівняння Гельмгольца та крайові умови на поверхнях хвильоводу, тобто нормальна складова швидкості коливань на стінках дорівнювала нулю.

Унаслідок взаємодії плоскої хвилі з вигином утворюється відбита хвиля та хвиля, що проходить в область II. Тому тиск в області I можна записати в системі координат xOy як суму падаючої хвилі та відбитої, яку представлено як суперпозицію нормальних мод області I:

$$p_I = \exp \exp(iky) + \sum_{n=0}^{\infty} A_n \cos(\eta_n(x-R)) \exp(-i\xi_n y), \quad \text{де } \eta_n = \frac{n\pi}{h}, \quad \xi_n = \sqrt{k^2 - \eta_n^2}.$$

Аналогічно, поле тиску в області V в системі координат $x'''O'y'''y'''$ визначає прониклу хвилю крізь другий вигин:

$$p_V = \sum_{n=0}^{\infty} B_n \cos(\eta_n(x''' - R')) \exp(-i\xi_n y''').$$

В області III функцію тиску можна представити в системі координат $x'O'y'$ як суперпозицію відбитих хвиль від першого та другого вигину:

$$p_{III} = \sum_{n=0}^{\infty} C_n \cos \cos(\eta_n(x' - R)) \exp \exp(i\xi_n y') + \sum_{n=0}^{\infty} D_n \cos \cos(\eta_n(x' - R)) \exp \exp(-i\xi_n (y' - l)).$$

Побудуємо розв'язок для області II:

$$p_{II} = \sum_{n=0}^{\infty} E_n \cos \cos(\eta_n(x - R)) \exp \exp(i\xi_n y) + \sum_{n=0}^{\infty} F_n \cos \cos(\eta_n(x' - R)) \exp \exp(-i\xi_n y') + \\ + \sum_{n=0}^{\infty} G_n \cos \cos(v_n \varphi) \frac{J_{v_n}(kr)}{J_{v_n}'(k(R+h))} + \sum_{n=0}^{\infty} H_n \cos \cos(v_n \varphi) \frac{Y_{v_n}(kr)}{Y_{v_n}'(kR)}, \quad \text{де } v_n = \frac{n\pi}{\alpha}.$$

Тут $J_{v_n}(kr)$ та $Y_{v_n}(kr)$ – функції Бесселя 1-го та 2-го роду відповідно, штрих при функції Бесселя позначає похідну за аргументом.

Аналогічно побудуємо розв'язок для області IV:

$$p_{IV} = \sum_{n=0}^{\infty} P_n \cos \cos(\eta_n(x'' - R')) \exp \exp(-i\xi_n y'') + \sum_{n=0}^{\infty} Q_n \cos \cos(\eta_n(x''' - R')) \exp \exp(i\xi_n y''') + \\ + \sum_{n=0}^{\infty} R_n \cos \cos(\mu_n \varphi') \frac{J_{\mu_n}(kr')}{J_{\mu_n}'(k(R'+h))} + \sum_{n=0}^{\infty} S_n \cos \cos(\mu_n \varphi') \frac{Y_{\mu_n}(kr')}{Y_{\mu_n}'(kR')}, \quad \text{де } \mu_n = \frac{n\pi}{\beta}.$$

Для визначення невідомих коефіцієнтів необхідно задовольнити умови спряження на межах часткових областей, тобто забезпечити рівність тисків і нормальних складових коливальних швидкостей, а також крайові умови на жорстких поверхнях в областях II та IV. Підстановка виразів, що описують поля тиску в кожній з областей, у відповідні умови спряження та крайові умови, з урахуванням формул перетворень між системами координат, приведе до системи функціональних рівнянь. Застосування властивості ортогональності відповідних систем функцій дає змогу алгебраїзувати функціональну систему. У підсумку отримаємо нескінченну систему лінійних алгебраїчних рівнянь другого роду відносно невідомих коефіцієнтів, яку розв'язують методом редукції (Grinchenko, Vovk, & Matsypura, 2018).

2. Результати роботи програми

Важливими характеристиками поширення звукової хвилі у хвильоводі є енергетичні коефіцієнти, тобто коефіцієнт проникнення W падаючої хвилі крізь вигини та коефіцієнт відбиття V від першого вигину. Коефіцієнт W визначимо як відношення середнього потоку потужності хвилі в області V до середнього потоку потужності падаючої хвилі в області I. Використовуючи відповідні вирази для функцій p_0 та p_V , отримаємо

$$W = \sum_{n=0}^{N_1} W_n, \quad \text{де } W_n = \frac{\varepsilon_n}{k} \operatorname{Re}(\xi_n) |B_n|^2.$$

Тут N_1 – кількість однорідних мод в області V. Для хвильоводу із жорсткими межами $\varepsilon_0 = 1$, $\varepsilon_n = 0.5$ для $n > 0$ (Grinchenko, Vovk, & Matsypura, 2018). Аналогічно, коефіцієнт відбиття V визначаємо як відношення середнього потоку потужності у відбитій хвилі до середнього потоку потужності падаючої хвилі в області I:

$$V = \sum_{n=0}^{N_2} V_n, \quad \text{де } V_n = \frac{\varepsilon_n}{k} \operatorname{Re}(\xi_n) |A_n|^2,$$

N_2 – кількість однорідних мод в області I.

Для представлених у цій роботі результатів кількість доданків у рядах Фур'є становила $n = 15$. Довжину області III обирали такою, що дорівнює $l = h - R \operatorname{tg} \left(\frac{\alpha}{2} \right) - R' \operatorname{tg} \left(\frac{\beta}{2} \right)$, а радіуси вигинів – $R = R'$. На рис. 2–3 наведено залежності енергетичних коефіцієнтів $W + V, W$ та коефіцієнтів збурення нормальних хвиль W_n області V від хвильових розмірів $\frac{h}{\lambda}$. За кутів вигинів $\alpha = 45^\circ, \beta = 45^\circ$ закон збереження енергії виконується з точністю не нижче 0.03 % (рис. 2 а). За $\frac{h}{\lambda} \approx 0.5$ коефіцієнт проникнення W знижується до нуля, тобто спостерігається запирання потоку енергії хвилі, що пов'язане із зародженням першої моди хвильоводу. За $\frac{h}{\lambda} \approx 1$ і $\frac{h}{\lambda} \approx 1.5$, що відповідає зародженню другої та третьої мод відповідно, запирання потоку енергії не відбувається (рис. 2 б). Варто зауважити, що практично на всьому розглянутому інтервалі $\frac{h}{\lambda}$ домінантною є нульова мода (рис. 2 в).

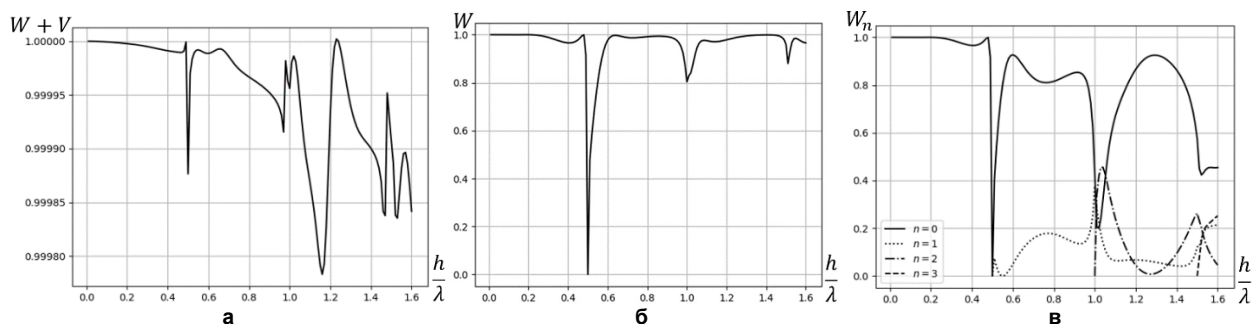


Рис. 2. Залежність енергетичних коефіцієнтів від хвильових розмірів за $\alpha = 45^\circ, \beta = 45^\circ$: а – $W + V$; б – W ; в – W_n

Для кутів $\alpha = 135^\circ, \beta = 135^\circ$ закон збереження енергії виконується з меншою точністю, найвища похибка обчислення становить близько 3.9 % за $\frac{h}{\lambda} \approx 1.44$ (рис. 3 а). Зменшити похибку можна, урахувавши більшу кількість доданків n у рядах Фур'є. Запирання потоку енергії за $\frac{h}{\lambda} \approx 0.5$ не відбувається, проте наявні додаткові області зниження коефіцієнта проникнення W , пов'язані з резонансними властивостями геометрії. Найбільше зниження коефіцієнта проникнення W спостерігається за $\frac{h}{\lambda} \approx 1.04$ (рис. 3 б). У цьому випадку за $\frac{h}{\lambda} < 0.82$ домінантною є нульова мода, тоді як за $\frac{h}{\lambda} > 0.82$ переважає перша мода (рис. 3 в).

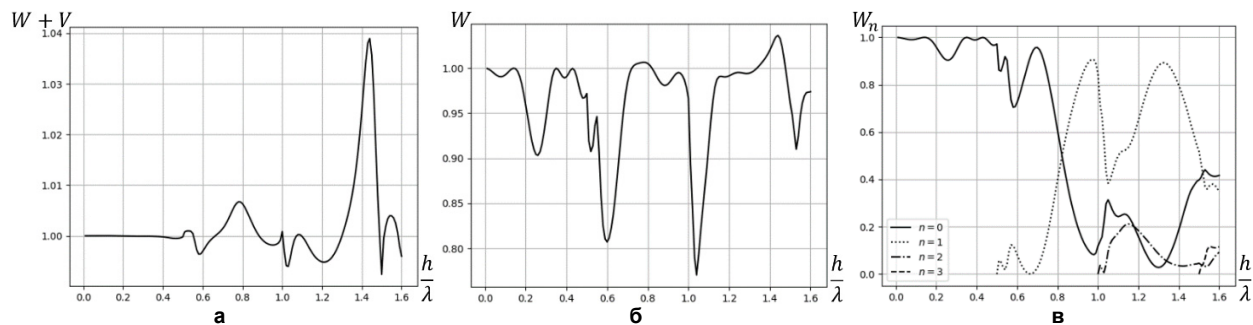


Рис. 3. Залежність енергетичних коефіцієнтів від хвильових розмірів за $\alpha = 135^\circ, \beta = 135^\circ$: а – $W + V$; б – W ; в – W_n

Дискусія і висновки

Отримані результати узгоджуються з даними попередніх досліджень. У роботі (Трунов, 2014), наприклад, показано, що за певних комбінацій параметрів хвильоводу з вигином спостерігаємо домінування однієї з мод, і цей ефект посилюється зі зменшенням відношення радіуса заокруглення до ширини хвильоводу. У роботі (Вовк, Маципура, & Трунов, 2014) встановлено, що за домінування нульової моди коефіцієнт проникнення є високим, тоді як у разі переважання вищих мод він знижується.

Отже, у представленій роботі розв'язано задачу поширення звукових хвиль у плоскому хвильоводі з акустично жорсткими стінками та геометричними неоднорідностями у вигляді двох округлих вигинів із використанням методу часткових областей. Проведено верифікацію закону збереження енергії. Показано, що зародження вищих однорідних мод супроводжується зменшенням коефіцієнта проникнення. Крім цього, можливе повне запирання потоку енергії хвилі, що пов'язане з появою першої однорідної моди. Встановлено, що за кутів вигинів 45° нульова мода домінує практично в усьому діапазоні хвильових розмірів. Натомість за кутів 135° виявлено додаткові області зниження коефіцієнта проникнення, зумовлені резонансними властивостями геометрії хвильоводу. У цьому випадку у нижній частині діапазону хвильових розмірів переважає нульова мода, тоді як у верхній – перша. У майбутніх дослідженнях плануємо проаналізувати залежність характеристик акустичного поля у хвильоводі від радіуса заокруглення та кутів вигинів, а також розглянути конфігурацію хвильоводу із протилежним напрямком другого вигину.

Отримані результати мають практичну цінність для розвитку теорії та застосування хвильоводів складної форми. Кількісні характеристики трансформації мод можуть бути використані, наприклад, під час проектування повітряних звукопроводів. Подальше вивчення механізмів перетворення енергії між модами відкриває перспективи створення ефективних шумопоглинальних систем і частотно-селективних пристроїв.

Внесок авторів: Володимир Кузьменко – формальний аналіз, програмне забезпечення, написання (оригінальна чернетка); Ярослав Троценко – аналіз джерел, концептуалізація, методологія, написання (перегляд і редагування).

Джерела фінансування. Фінансування здійснюється власним коштом авторів та частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Вовк, І. В., Маципура, В. Т., & Трунов, О. О. (2014). Поширення хвилі у криволінійному хвильоводі з різкою зміною ширини вигину. *Вісник Київського національного університету імені Тараса Шевченка. Серія: фізико-математичні науки*, 3, 28–31.
- Грінченко, В. Т., & Маципура, В. Т. (2025). Дискретно-аналітичний метод розв'язання задач математичної фізики. *Вісник Національного технічного університету "ХПІ". Серія: Математичне моделювання в техніці та технологіях*, 1, 55–60. [https://doi.org/10.20998/2222-0631.2025.01\(8\).06](https://doi.org/10.20998/2222-0631.2025.01(8).06)
- Маципура, В. Т., Вовк, І. В., & Трунов, О. О. (2013). Поширення хвиль в криволінійному хвильоводі. *Вісник Київського національного університету імені Тараса Шевченка. Серія: фізико-математичні науки*, 3, 192–195.
- Трунов, О. О. (2014). Хвильове поле у хвильоводі з вигином. *Вісник Київського національного університету імені Тараса Шевченка. Серія: фізико-математичні науки*, 2, 83–86.
- Grinchenko, V. T., Vovk, I. V., & Matsypura, V. T. (2018). *Acoustic wave problems*. Begell House, Inc. <https://doi.org/10.1615/978-1-56700-459-5.0>
- Naida, S. A., Korzhyk, O. V., Naida, N. S., Korzhyk, M. O., Naida, A. S., & Popovych, P. V. (2023). Application of the Partial Domain Method to the Determination of the Directional Properties of a Finite-Length Cone Horn for a Broadband Acoustic Ear Echo Spectrometer. *Journal of nano- and electronic physics*, 15(6), 06012. [https://doi.org/10.21272/jnep.15\(6\).06012](https://doi.org/10.21272/jnep.15(6).06012)

References

- Grinchenko, V. T., & Matsypura, V. T. (2025). Discrete-analytical method of solving problems in mathematical physics. *Bulletin of the National Technical University "KhPI". Series: Mathematical modeling in engineering and technologies*, 1, 55–60 [in Ukrainian]. [https://doi.org/10.20998/2222-0631.2025.01\(8\).06](https://doi.org/10.20998/2222-0631.2025.01(8).06)
- Grinchenko, V. T., Vovk, I. V., & Matsypura, V. T. (2018). *Acoustic wave problems*. Begell House, Inc. <https://doi.org/10.1615/978-1-56700-459-5.0>
- Matsypura, V. T., Vovk, I. V., & Trunov, O. O. (2013). Wave propagation in the bent waveguide. *Bulletin of the Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*, 3, 192–195 [in Ukrainian].
- Naida, S. A., Korzhyk, O. V., Naida, N. S., Korzhyk, M. O., Naida, A. S., & Popovych, P. V. (2023). Application of the Partial Domain Method to the Determination of the Directional Properties of a Finite-Length Cone Horn for a Broadband Acoustic Ear Echo Spectrometer. *Journal of nano- and electronic physics*, 15(6), 06012. [https://doi.org/10.21272/jnep.15\(6\).06012](https://doi.org/10.21272/jnep.15(6).06012)
- Trunov, O. O. (2014). A wave field in a waveguide with a bend. *Bulletin of the Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*, 2, 83–86 [in Ukrainian].
- Vovk, I. V., & Matsypura, V. T. & Trunov, O. O. (2014). The wave propagation in bent waveguide with thickened bend. *Bulletin of the Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*, 3, 28–31 [in Ukrainian].

Отримано редакцією журналу / Received: 27.08.25

Прорецензовано / Revised: 26.09.25

Схвалено до друку / Accepted: 10.10.25

Volodymyr KUZMENKO, Student
ORCID ID: 0009-0005-7922-430X
e-mail: vovakuzmenko0200@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Yaroslav TROTSENKO, PhD, Assist.
ORCID ID: 0009-0000-8387-9118
e-mail: yaroslav.trotsenko@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

MODELING OF SOUND PROPAGATION IN A PLANAR WAVEGUIDE WITH ROUNDED BENDS

The relevance of this study is determined by the practical significance of waveguides with complex geometries and the strong dependence of their wave propagation characteristics on the physical and geometric parameters of the system. This work aims to investigate the regularities of sound wave propagation in a planar waveguide filled with an ideal compressible fluid and containing two rounded bends. A numerical-analytical model based on the partial domains method has been developed, enabling the analysis of acoustic wave behavior in the waveguide depending on the selected parameters. The features of wave transmission through geometric inhomogeneities are analyzed for two bending angle configurations across a wide range of normalized wavelengths. The energy transmission coefficients of the sound wave through the bends, as well as the propagating mode excitation coefficients, have been computed over an extensive set of parameters. The obtained results have been verified by checking the fulfillment of the energy conservation law. It is shown that the computational error increases with larger normalized wavelengths and bending angles, but can be reduced by accounting for a greater number of inhomogeneous modes. The onset of higher-order propagating modes reduces the transmission coefficient, while the emergence of the first mode can lead to a complete suppression of the wave energy flux. For bending angles of 45°, the zero mode dominates across almost the entire range of normalized wavelengths. For bending angles of 135°, additional regions of reduced transmission are observed due to the resonant properties of the waveguide geometry. In this case, the zero mode dominates in the lower half of the considered wavelength range, while the first mode becomes dominant in the upper half. The obtained results can be applied in the design of air ducts, frequency-selective acoustic devices, and noise reduction systems. A promising direction for future research is the detailed study of intermodal energy conversion mechanisms, which provides opportunities for controlling the spatial structure of the acoustic field.

Keywords: acoustic wave, waveguide, sound field, homogeneous modes, inhomogeneous modes, energy coefficients, partial domains method.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

UDC 519.925.51

DOI: <https://doi.org/10.17721/1812-5409.2025/2.17>

Victor KULIAN, PhD (Engin.), Assoc. Prof.

ORCID ID: 0000-0002-6320-4719

e-mail: vkulian@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Olena YUNKOVA, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0002-8233-5087

e-mail: yunkova.olena@kneu.edu.ua

Kyiv National Economic University named after Vadym Getman

Maryna KOROBOVA, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0003-4286-7828

e-mail: marynakorobova@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

PARAMETRIC IDENTIFICATION OF ASSET DYNAMIC MODELS

The article describes the mathematical formulation of the problem of constructing the optimal structure of a securities portfolio on the stock market. The most important mathematical problems in decision-making by investors are the problems of constructing a forecast for a single share and optimal diversification of the portfolio structure. In applied mathematics, such a problem is called a mathematical problem of optimal control. In this work, mathematical models are used that describe the dynamics of the formation of the market value of a single share and a portfolio. The corresponding models are written in the class of ordinary differential equations with parameters. The procedure for constructing a dynamic model of the formation of the market value of a single share is based on the application of the market model of W. Sharpe and the fundamental theory of H. Markowitz.

The paper formulates a problem of mathematical models identifying the parameters of dynamic processes that can be described by ordinary differential equations and systems. Using the example of mathematical models of dynamic formation of the market value of a single share and a portfolio of shares, algorithms for constructing optimal values of parameters of such models have been developed. Parametric identification and optimization algorithms are based on iterative procedures that allow, at each step, to form the "best" values of model parameters from the point of view of selected quality criteria. An algorithm for constructing guaranteed parameter estimates in the class of ellipsoidal sets has been developed. The mathematical models and algorithms presented in this study make it possible, together with the methods of technical analysis, to develop effective tactics and strategies for optimal investment in securities. The results obtained should be considered as one of the alternative approaches to modeling the dynamics of market assets and their portfolios.

Keywords: *mathematical modeling, asset dynamics models, parameter identification, stock market.*

AMS 2020 classification: 65K10, 93B30.

Introduction

This article is devoted to the construction of algorithms for determining the parameters of mathematical models of the dynamics of risky assets of the stock market formation. The relevance of this scientific direction is confirmed by the fact that a large number of scientific studies have been devoted to the development of such models and algorithms, which were carried out by H. Markowitz, W. Sharpe, K. Ito, F. Black, M. Scholes, and R. Merton. The object of the study is parametrically specified mathematical models of the stock market assets.

Modern applied mathematical modeling operates with various approaches to calculating parameters with a certain structure of mathematical models. The most well-known and widespread methods for identifying parameters of discrete and continuous models include: methods of covariance-dispersion analysis; methods of simulation modeling; approaches based on the principles of stability theory; approaches based on methods of sensitivity analysis of solutions; methods of minimax identification.

However, due to the complexity of the model structure, the nonlinear nature of the processes being modeled, as well as other factors, it is often necessary to turn to computer modeling, which makes it possible to build computer images of models by searching through a large number of candidate models using self-organization or simulation modeling methods (Xin-Yu Guo, & Sheng-En Fang, 2023). The complexity of a parametrically specified model will be understood as the dimensionality of the parameter vector, and the main principle of simulation modeling will be: "the more complex the model, the more accurate it is", in contrast to the main principle of model self-organization about "the existence of a of optimal complexity model" (Elder, Zehetner, & Kunze, 2016). Such approaches, which include methods such as neural networks or group consideration of arguments, are often used in solving problems of modeling complex and interconnected economic processes, including mathematical problems of the stock market and the banking sector of the economy (Rios, Efimov, & Moreno, 2017).

The approaches to building mathematical models of the formation of price dynamics of a single share and a portfolio of shares are based on the principles of portfolio theory, formulated by H. Markowitz and W. Sharpe. H. Markowitz proposed approaches to measuring risk and basic principles of portfolio investment (Markowitz, 1952). This laid the theoretical foundation for research on a portfolio of risky investments. Sharpe's capital asset pricing model (Sharpe, 1964) is also actively used by investors in solving problems of modeling the dynamics and building short-term forecasts of market prices of assets.

Among other important approaches to solving applied portfolio theory problems, the following can be noted:

- Evolutionary and Memetic Computing for Project Portfolio Selection (Kolish, & Fliedner, 2022; Yue, Jing, & Liu, 2025);
- An exploration of meta-heuristic approaches for the project portfolio selection (Harrison et al., 2020);
- A simheuristic method for project portfolio selection under uncertainty (Panadero et al., 2020);
- A bi-objective fuzzy portfolio selection model (Kar et al., 2019);
- A genetic programming for portfolio selection problem solution (Lin, Zhu, & Gao, 2020);

© Kulian Victor, Yunkova Olena, Korobova Maryna, 2025

• An alternative mathematical method for the portfolio selection problem (Mohagheghi, Meysam Mousavi, & Mohagheghi, 2020; Li et al., 2020);

• Project portfolio selection and scheduling optimization based on risk measure (Dixit, & Tiwari, 2020).

The first part of this study is devoted to the consideration and analysis of mathematical models of the market value dynamics of a single share and a portfolio of shares (Kulian, Yunkova, & Korobova, 2022). An important feature of models of the formation of stock market asset dynamics is their parametric representation. In this regard, any applied problem that can be formulated for such models requires preliminary determination of parameter values. Constructing a short-term forecast or optimal diversification of an investment portfolio is a most important applied problem (Kulian, Yunkova, & Korobova, 2021). The second part of the study is devoted to constructing mathematical algorithms that can be used for solving the problem of parametric identification of mathematical models of the formation of asset dynamics. The third part of the study considers an application of the constructed algorithms to solve applied problems of modeling the asset dynamics in the real stock market.

1. Analysis of stock price formation models in the stock market

As shown in (Kulian, Yunkova, & Korobova, 2021), based on the market model of W. Sharpe, generalized deterministic mathematical models of the formation of the market value of a single stock and a portfolio of stocks can be written in the form

$$\dot{r}_i = \frac{dr_i}{dt} = f(r_i, t, \alpha), \quad r_i(t_0) = r_0, \quad t \in [t_0, T], \quad i = \underline{1, n} \quad (1)$$

and

$$\dot{r}_p = f_p(r_p, x_i, \dot{x}_i, r_i, \dot{r}_i, t) \quad (2)$$

respectively, and are given parametrically. Here, r_i is the expected market value of the stock; r_p is the expected market value of the investment portfolio; x_i is a share of the i -type asset in the portfolio; t is time.

An important mathematical problem, the solution of which is used to correctly formulate the problem of constructing the optimal structure of an investment portfolio, is the problem of identifying the optimal values of the parameters α of mathematical models of the formation of the market value of one share (1) and a portfolio of shares (2). The problem of optimizing a portfolio of shares from the point of view of its profitability in the static case has the form

$$r_p = \sum_{i=1}^n x_i r_i \rightarrow \max_x.$$

In practical investing, it can be considered in the following formulation

$$r_p(T) = \sum_{i=1}^n x_i(T) r_i(T) \rightarrow \max_{x \in X(t)}, \quad r_p(t_0) = r_{p0}, \quad t \in [t_0, T]$$

and consists in constructing an investment portfolio with optimal profitability at the final point T . Here $X(t)$ is a bounded closed set of stocks. The right-hand sides of the equations of the system (1) are linear and depend on the parameters. To ensure the possibility of its integration

$$\frac{dr_i}{dt} = (\alpha_1 SM_{ind}(t) + \alpha_2 I(t)) r_i(t) + \alpha_3 (r_i(t), \rho_{ij}), \quad i, j = \underline{1, n}$$

we will build procedures for their determination. Here, $SM_{ind}(t)$ – Stock Market Index; $I(t)$ – Inflation index; ρ_{ij} – correlation coefficient between i and j shares.

In different formulations of the problem, the vectors $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ or $\alpha = (\alpha_1, \alpha_2)$ can be considered as the parameter vector. The first representation of the parameter vector will be used in the case when the information about the market value of shares is "noisy". In the case when the market value of shares adequately reflects the market dynamics, the second representation is used.

2. Algorithms for constructing optimal parameters of the models

Algorithm 1. Another approach to solving the parametric identification problem is to use sensitivity functions. Consider the problem of constructing sensitivity functions for pricing models of a single stock and a portfolio of stocks.

For the convenience of transformations, we denote the right-hand side by $f(t, r, \alpha)$, where $\alpha = (\alpha_1, \alpha_2)$. We will also assume that the initial conditions $r(t_0) = r_0 = r_0(\alpha)$, $t_0 = t_0(\alpha)$ and that α is a scalar and $\alpha = \alpha_1$, since in short-term modeling of the dynamics of stock price formation α_2 can be considered constant. In such a statement, ρ_{ij} are the correlation coefficients between stocks, which can be calculated based on known statistical information. According to (Kulian, Yunkova, Korobova, 2021), provided that the right-hand side of (1) is continuous in its arguments and continuously differentiable in r, α and also under the continuous differentiability of the functions r_0, t_0 in α_1 , there is a continuous derivative

$$u(t, \alpha_1) = \frac{\partial r(t, \alpha_1)}{\partial \alpha_1}.$$

This derivative satisfies the differential equation

$$\frac{du}{dt} = \frac{\partial f(r, t, \alpha_1)}{\partial r} u + \frac{\partial f(r, t, \alpha_1)}{\partial \alpha_1}$$

under the initial condition

$$u(t_0) = \frac{dr_0(\alpha_1)}{d\alpha_1} - f(r_0(\alpha_1), t_0(\alpha_1), \alpha_1) \frac{dt_0(\alpha_1)}{d\alpha_1}. \quad (3)$$

Equation (3) is obtained by direct differentiation with respect to α_1 , and the initial condition is obtained from the differentiation of the integral equation

$$r(t, \alpha_1) = \int_{t_0(\alpha_1)}^t f(r(\tau, \alpha_1), \tau, \alpha_1) d\tau + r_0(t_0, \alpha_1)$$

and looks like

$$u(t, \alpha_1) = \frac{\partial r(t, \alpha_1)}{\partial \alpha_1} = \int_{t_0(\alpha_1)}^t \left[\frac{\partial f(r(\tau, \alpha_1), \tau, \alpha_1)}{\partial r} u(\tau, \alpha_1) + \frac{\partial f(r(\tau, \alpha_1), \tau, \alpha_1)}{\partial \alpha_1} \right] d\tau + \frac{dr_0(\alpha_1)}{d\alpha_1} - f(r(t_0(\alpha_1), \alpha_1), t_0(\alpha_1), \alpha_1) \frac{dt_0(\alpha_1)}{d\alpha_1}.$$

The initial condition for the sensitivity function $u(t_0)$ is obtained by putting $t = t_0$ in (3). Therefore, for (1), we have defined the initial condition and the sensitivity function, which describes the dependence of the stock price on the parameter α_1 . Let us apply the sensitivity function to identify the parameters α_1, α_2 of the mathematical model (1). The mathematical problem has the following form

$$\min_{\alpha_1, \alpha_2} (r(t, \alpha_1, \alpha_2) - z(t))^2, \quad t \in [t_0, T].$$

Here $z(t)$ – observation of the share price behavior in the market.

Algorithm 2. The vector of model parameters is written in the form $\alpha = (\alpha_1, \alpha_2)$ provided $r(t_0) = r_0$ at $t \in [t_0, t_1]$.

Consider the point α_0 in the parameter space of the mathematical model

$$\alpha_0 = (\alpha_{01}, \alpha_{02}).$$

And let this value of parameters be the solution of the problem of parametric identification of the model (1). Around the point α_0 we describe a circle of unit radius

$$(\alpha_1 - \alpha_{01})^2 + (\alpha_2 - \alpha_{02})^2 = 1$$

with the center in p. $(\alpha_{01}, \alpha_{02})$. The length of the circle will then be $l = 2\pi$.

Let us divide this line into n equal parts, and the value of n is chosen depending on the accuracy of the obtained solution to the problem of constructing a guaranteed multiple-parameter estimate. Consider an arbitrary point $\alpha_k = (\alpha_{k1}, \alpha_{k2})$.

Let's draw a tangent to the circle at this point. Its equation will be $F'_{\alpha_1}(\alpha_{k1}, \alpha_{k2})(\alpha_1 - \alpha_{k1}) = F'_{\alpha_2}(\alpha_{k1}, \alpha_{k2})(\alpha_2 - \alpha_{k2})$.

The equation of the normal to the tangent at this point will be

$$F'_{\alpha_1}(\alpha_{k1}, \alpha_{k2})(\alpha_2 - \alpha_{k2}) = F'_{\alpha_2}(\alpha_{k1}, \alpha_{k2})(\alpha_1 - \alpha_{k1}).$$

We will search for new values of the parameters α that satisfy the conditions for the software functioning of the system on the normal. From the equation of the normal we determine α_2

$$\alpha_2 = \frac{F'_{\alpha_2}(\alpha_{k1}, \alpha_{k2})(\alpha_1 - \alpha_{k1})}{F'_{\alpha_1}(\alpha_{k1}, \alpha_{k2})} + \alpha_{k2}.$$

The new position of the coordinate α_2 is determined by changing the position of the coordinate α_1

$$\alpha_{N1} = \alpha_1 + \delta\alpha_1, \quad \delta > 0,$$

$$\alpha_{N2} = \frac{F'_{\alpha_2}(\alpha_{k1}, \alpha_{k2})(\alpha_1 + \delta\alpha_1 - \alpha_{k1})}{F'_{\alpha_1}(\alpha_{k1}, \alpha_{k2})} + \alpha_{k2}.$$

Thus, a new position of the parameter value $\alpha_N \in A$, is obtained, where A is a bounded closed set of parameters of the mathematical model

$$\alpha_N = (\alpha_{N1}, \alpha_{N2}).$$

The above procedure implements the movement of the parameter α along the normal until one of the selected criteria above is met. By performing such steps for each of the previously defined points, we will obtain new values of the coordinates of the vector α in the parameter space of the mathematical model.

Table 1 below presents the results of a computational experiment for the practical application of Algorithm 2 for model (1).

Table 1

Results of modeling the dynamics of two stocks

Characteristic	Day								
	1	2	3	4	5	6	7	8	9
$r_1, \$$	42.22	43.74	43.60	44.13	43.11	42.87	42.59	42.57	42.08
$r_2, \$$	3046.01	3106.89	3133.15	3190.00	3140.02	3125.71	3115.51	3088.99	3087.67
SM_{ind}	5133.07	5170.97	5196.52	5204.08	5162.25	5082.77	5056.90	5068.05	5080.94
Δ_1	0.047	0.054	0.034	0.035	0.081	0.062	0.054	0.043	0.040
Δ_2	0.031	0.035	0.047	0.034	0.096	0.071	0.059	0.037	0.032

In the table r_1 denotes the market value of the first company's stock, and r_2 denotes the market value of the second company's stock. Through Δ_1 indicates the deviation between actual and calculated trajectories of the first stock, and through Δ_2 indicates the deviation between actual and calculated trajectories of the second stock.

To calculate deviation Δ_1 of trajectories were used $\alpha_1 = 2.348$, $\alpha_2 = 0.017$, $\alpha_3 = -0.76$. To calculate deviation Δ_2 of trajectories were used $\alpha_1 = 1.981$, $\alpha_2 = 0.032$, $\alpha_3 = -0.489$. $l(t) = \text{const}$. Correlation dependencies ρ_{ij} and model (1) parameters $\alpha_1, \alpha_2, \alpha_3$ were calculated based on known statistical information of the S&P 500 market using Algorithm 2. Analysis of the results in Tab. 1 shows the correlation between the dynamics of the market index and the dynamics of market prices of shares of two companies. Analysis of computational experiment results shows that the application of Algorithm 2 in areas of a stable positive or negative trend allows obtaining a 2–5 % deviation of the real and calculated trajectories, while in areas of a change in the direction of dynamics, an increase in the deviation of the trajectories to 8–9 % is observed.

Algorithm 3. When solving forecasting or management problems in the above application areas, it is necessary to develop algorithms and methods to obtain guaranteed financial results. An example of such an approach can be the development of the Value at Risk methodology, and confirmation of its relevance is its general recognition as a reliable mathematical tool when making investment decisions. Let us turn to the problem of constructing a guaranteed multiple-parameter estimate for a mathematical model of general form (1). Let us formulate a procedure for constructing an optimal ellipsoidal parameter estimate of the mathematical model (1).

After constructing and analyzing solutions to the problem of identifying the parameters of the mathematical model of the formation of the market value of a stock (or a portfolio of stocks), we will determine the point that is farthest from the point $D(\alpha_1^c, \alpha_2^c)$, the coordinates of which will be determined using the formulas

$$\alpha_1^c = \frac{1}{k} \sum_{j=0}^k \alpha_1^j, \quad \alpha_2^c = \frac{1}{k} \sum_{j=0}^k \alpha_2^j.$$

To do this, we will solve the one-dimensional optimization problem

$$R = \max_i L(\alpha_i, \alpha^c),$$

where $L(p, q)$ is a function of the distance between points p and q .

On the boundary of the circle $S(1, D)$ of radius 1, with the center at the point $D(\alpha_1^c, \alpha_2^c)$ we form an arbitrary δ -grid $\{y_i\}_{i \in K_\delta}$ with nodes y_i . At each node y_i we construct the external normal vector $\bar{n}_{e_i}$ to the circle according to the rules described above, and we solve the auxiliary problem of constructing new values of the parameters α analogously to the approach described in Algorithms 1, 2.

Having performed the above procedure for each of the vectors $n_i, i \in K_n$, we will obtain a new set of points $\alpha_{e_n} \in A$. Let us describe the ellipsoid of the smallest volume (the ellipse of the smallest area in the plane) around the points thus constructed, which we will call the "guaranteed ellipsoidal estimate" of the parameters of the model (1), constructed on the defined criteria

$$\min_{\alpha \in A} \int_{t_0}^T (r^0(t_0, t_1, r_0, \alpha) - r_{exp}(t)) dt, \\ \min_{\alpha \in A} \max_{t \in [t_0, T]} \|(r^0(t_0, t_1, r_0, \alpha) - r_{exp}(t))\|.$$

Here $r^0(t_0, t_1, r_0, \alpha)$ is the solution of the Cauchy problem for model (1) under the condition $r^0(t_0, t_1, r_0, \alpha)$, r_{exp} – the experimental or program trajectory.

Based on the solutions of the parametric identification problem of the mathematical model (1) and known observations, we will construct a guaranteed multiple estimate of the parameters in the class of ellipsoidal sets

$$Q(B, d): \{(B(\alpha - d), \alpha - d) \leq 1\}, \quad (4)$$

where B is a symmetric positive-definite matrix that specifies the geometry of the multiple estimate in the parameter space; d is geometric center of the ellipsoid, $\alpha, d \in R^n$. The set (4) will be called a "guaranteed multiple estimate" if each value of the parameter α obtained for arbitrary observations belongs to $Q(B, d)$.

The iterative procedure of the method for constructing a guaranteed multiple estimate implements the refinement of the elements of the matrix B and the vector d at each step. The initial matrix B is chosen to be the unit $B^0 = E$, and the initial position of the center d is defined as the center of mass

$$d^0 = \frac{1}{k} \sum_{j=1}^k \alpha_j,$$

where k is the number of points.

At the $k + 1$ - step of the procedure, the formulas for determining B^{k+1} and d^{k+1} have the form

$$B^{k+1} = B^k + \lambda_1^{k+1} \nabla_B \{(B^k(\alpha - d^k), \alpha - d^k)\}, \\ d^{k+1} = d^k + \lambda_2^{k+1} \nabla_d \{(B^k(\alpha - d^k), \alpha - d^k)\},$$

where λ_1^{k+1} and λ_2^{k+1} are such that

$$(B^{k+1}(\lambda_1^{k+1})(\alpha_i - d^{k+1}(\lambda_2^{k+1}), \alpha_i - d^{k+1}(\lambda_2^{k+1})) < \\ < (B^k(\alpha_i - d^k), \alpha_i - d^k), \text{ where } \underline{i} = \arg \arg \max_i (B(\alpha_i - d), \alpha_i - d). \\ (B^j(\alpha - d^j), \alpha - d^j) < 1.$$

On the boundary of the sphere $S(1, d^j)$ with the center at the point d^j (where the j -number of the procedure step for which the inequality $(B^j(p - d^j), p - d^j) < 1$ is satisfied), we form an arbitrary δ -grid $\{y_i\}_{i \in K_\delta}$ with nodes y_i .

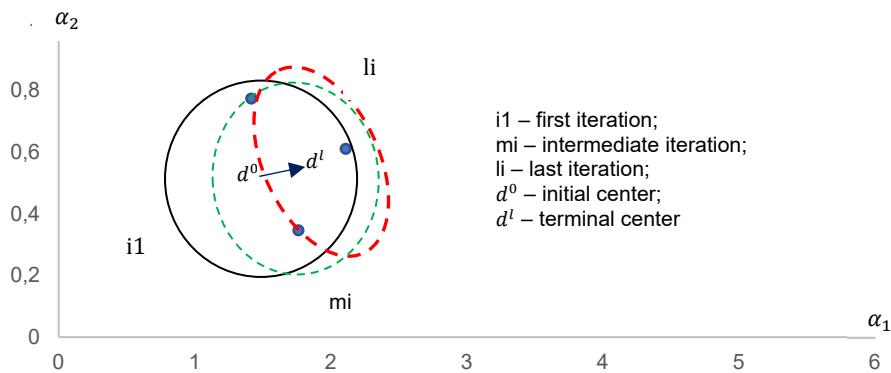


Fig. 1. Construction of multiple evaluation

As shown in Fig.1, at each node, using the described method of constructing the ellipsoid of the smallest volume around the given points, we construct a guaranteed multiple parameter estimate, which, as determined by Algorithm 3, ensures the fulfillment of the selected quality criteria. The set constructed in this way allows us to form an admissible set of parameters, and with further optimization of this multiple estimate, the logical next step is to formulate and solve the problem of constructing an "effective" by H. Markowitz multiple estimate.

Discussion and conclusions

The presented mathematical procedures make it possible for dynamic mathematical models of a single stock and a portfolio of stocks to solve the problems of parametric and guaranteed multiple identification of parameters and, based on the developed algorithms for determining the admissible set of parameters, to find solutions to the problems of constructing a market value trajectory for mathematical model (1) and optimal diversification of an investment portfolio of risky securities based on model (2).

Authors' contribution: Victor Kulian – conceptualization; methodology; Olena Yunkova – empirical data collection and validation; Maryna Korobova – preparation of literature review or theoretical foundations of research.

Sources of funding. The research was partially funded by Taras Shevchenko National University of Kyiv.

References

- Dixit, V., Tiwari, M. K. (2020). Project portfolio selection and scheduling optimization based on risk measure: a conditional value at risk approach. *Annals of Operations Research*, 285(1–2), 9–33. <https://doi.org/10.1007/s10479-019-03214-1>
- Elder, R., Zehetner, C., Kunze, W. (2016). Comparison of parameter identification techniques. The 3-rd International Conference for Manufacturing and Industrial Technologies. *MATEC Web Conferences* 70:09007 (pp. 112–118). <https://doi.org/10.1051/mateconf/20167009007>
- Harrison, K. R., Elsayed, S., Weir, T., Garanovich, I. L., Taylor, R., Sarker, R. (2020). An exploration of meta-heuristic approaches for the project portfolio selection and scheduling problem in a defence context. *Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, Canberra*, 1395–1402. <https://doi.org/10.1109/ssci47803.2020.9308608>
- Kar, M. B., Kar, S., Guo, S., Li, X., Majumder, S. (2019). A new bi-objective fuzzy portfolio selection model and its solution through evolutionary Algorithms. *Soft Computing*, 23(12), 4367–4381. <https://doi.org/10.1007/s00500-018-3094-0>
- Kolish, R., Friedner, T. (2022). *Evolutionary and Memetic Computing for Project Portfolio Selection and Scheduling*. Springer. <https://link.springer.com/book/10.1007/978-3-030-88315-7>
- Kulian, V., Yunkova, O., Korobova, M. (2021). Mathematical problem of banking assets diversification. *Bulletin of Taras Shevchenko National University of Kyiv*, 1, 85–88. <https://doi.org/10.17721/1812-5409.2021/1.11>
- Kulian, V., Yunkova, O., Korobova, M. (2022). Solutions sensitivity when modeling of investment dynamics. *Bulletin of Taras Shevchenko National University of Kyiv*, 4, 51–54. <https://doi.org/10.17721/1812-5409.2022/4.6>
- Li, X., Huang, Y.-H., Fang, S.-C., Zhang, Y. (2020). An alternative efficient representation for the project portfolio selection problem. *European Journal of Operational Research*, 281(1), 100–113. <https://doi.org/10.1016/j.ejor.2019.08.022>
- Lin, J., Zhu, L., Gao, K. (2020). A genetic programming hyper-heuristic approach for the multi-skill resource constrained project scheduling problem. *Expert Systems with Applications*, 140, art. ID 112915. <https://doi.org/10.1016/j.eswa.2019.112915>
- Markowitz, H. (1952). Portfolio selection. *The Journal of finance*, 7(1), 77–91. <http://links.jstor.org/sici?sici=0022-1082%28195203%297%3A1%3C77%3AAPS%3E2.0.CO%3B2-1>
- Mohagheghi, V., Meysam Mousavi, S., Mojtahedi, M. (2020). Project portfolio selection problems. Two decades review from 1999 to 2019. *Journal of Intelligent and Fuzzy Systems*, 38, 1675–1689. <https://doi.org/10.3233/jifs-182847>
- Panadero, J., Doering, J., Kizys, R., Juan, A. A., Fito, A. (2020). A variable neighborhood search simheuristic for project portfolio selection under uncertainty. *Journal of Heuristics*, 26, 1–23. <https://doi.org/10.1007/s10732-018-9367-z>
- Rios, H., Efimov, D., Moreno, J. A. (2017). Time-Varying parameter identification algorithms: Finite and fixed time. Convergence. *IEEE Transactions on automatic control*, 62(7), 3671–3678. <https://doi.org/10.1109/TAC.2017.2673413>
- Sharpe, W. (1964). Capital assets prices: A theory of market equilibrium under conditions of risk. *The Journal of finance*, 19(3), 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Xin-Yu Guo, Sheng-En Fang (2023). Structural parameter identification using physics-informed neural networks. *Measurement*, 220, 78–85. <https://doi.org/10.1016/j.measurement.2023.113334>
- Yue, X., Jing, X., Liu, X. (2025). Parameter identification of dynamical systems based on short-term prediction by the generalized cell mapping method with deep learning. *Nonlinear Dynamics*, 113, 4031–4044. <https://doi.org/10.1007/s11071-024-09943-8>

Отримано редакцією журналу / Received: 06.01.25

Прорецензовано / Revised: 24.09.25

Схвалено до друку / Accepted: 10.10.25

Віктор КУЛЯН, канд. тех. наук, доц.

ORCID ID: 0000-0002-6320-4719

e-mail: vkulian@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Олена ЮНЬКОВА, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0002-8233-5087

e-mail: yunkova.olena@kneu.edu.ua

Київський національний економічний університет імені Вадима Гетьмана, Київ, Україна

Марина КОРОБОВА, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0003-4286-7828

e-mail: marynakorobova@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

ІДЕНТИФІКАЦІЯ ПАРАМЕТРІВ МОДЕЛЕЙ ДИНАМІКИ АКТИВІВ

Описано математичну постановку задачі побудови оптимальної структури портфеля цінних паперів. Найважливішими прикладними проблемами в ухваленні рішень про інвестування є математичні задачі про побудову короткострокового прогнозу для окремої акції та задача про оптимальну диверсифікацію структури портфеля. Під час формулювання та розв'язання задач моделювання динаміки активів фондового ринку розглянуто математичні моделі формування ринкових цін однієї акції та портфеля акцій. Відповідні моделі сформульовано у класі звичайних диференціальних рівнянь з параметрами. Параметрами таких моделей є коефіцієнти чутливості ринкової вартості до зміни факторів, що на неї впливають. Процедура побудови динамічної моделі формування ринкової вартості окремої акції ґрунтується на застосуванні ринкової моделі В. Шарпа й елементів фундаментальної теорії Г. Марковиця. У дослідженні сформульовано загальну задачу ідентифікації параметрів динамічних процесів, що можуть бути описані звичайними диференціальними рівняннями та системами. На прикладі математичних моделей формування динаміки ринкової вартості однієї акції та портфеля акцій розроблено алгоритм визначення оптимальних значень параметрів таких моделей. Алгоритм побудовано на основі ітераційної процедури, яка дає змогу на кожному кроці формувати "оптимальні", з позиції обраних критеріїв якості, значення параметрів моделі. Розроблений алгоритм дозволяє визначати такі оцінки параметрів у класі еліпсоїдальних множин. На прикладі математичних задач побудови оптимальної структури портфеля ризикованих активів розроблено формальний підхід до формування "гарантованих" фінансових показників інвестиційної діяльності. Математичні моделі й алгоритми, представлені у пропонуваному дослідженні, дають змогу разом із методами технічного аналізу розробити ефективні тактику та стратегію оптимального інвестування у цінні папери. Результати дослідження варто розглядати як один з альтернативних підходів фундаментального аналізу до моделювання динаміки ринкових активів.

Ключові слова: математичне моделювання, динаміка активів, ідентифікація параметрів, фондовий ринок.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 531.7

DOI: <https://doi.org/10.17721/1812-5409.2025/2.18>Микола МЕЛЬНИЧЕНКО¹, канд. фіз.-мат. наук, ст. наук. співроб.

ORCID ID: 0000-0002-3473-0701

e-mail: mykola.melnichenko@knu.ua

Леонід ЯРОВОЙ¹, канд. техн. наук.

ORCID ID: 0000-0001-6374-4160

e-mail: yarovoi@univ.kiev.ua

Наталія ПУЧКО¹, провід. інж.

e-mail: natalia.puchko@knu.ua

Михайло ВОДОТОВКА¹, електрослюсар

e-mail: nanocrystallab@univ.kiev.ua

¹Київський національний університет імені Тараса Шевченка, Київ, Україна

МЕТОДИКА ЕКСПЕРИМЕНТАЛЬНОГО МОДЕЛЮВАННЯ ПРОНИКНЕННЯ ВИСОКОШВИДКІСНИХ УДАРНИХ ЕЛЕМЕНТІВ У БРОНЕМАТЕРІАЛАХ

Розроблено та створено експериментальний комплекс настільного розгінного стенду, у якому куля діаметром 8 мм зі сталі ШХ15 прискорюється енергією газів монтажного патрону під час руху каналом завдовжки 570 мм з керуваною швидкістю від 200 до 1500 м/с та різними кутами атаки від 45 до 90°. Проведено апробацію розгінного стенду. Встановлено, що після обстрілу зі швидкістю 200 м/с та 600 м/с для різних зразків бронематеріалів з'являється пластична вм'ятина різного вигляду та різної глибини. Водночас у залишковій вм'ятині виявлено залишки процесу плавлення та витискування бронематеріалу із вм'ятини. У разі обстрілу багатошарового бронепакету кулею зі швидкістю 600 м/с не відбувається деформації кулі. Для оцінки фізико-механічних характеристик досліджуваного багатошарового бронематеріалу виконано серію вимірювань мікротвердості біля місць взаємодії на поверхні зразка. Тестування на мікротвердість проведено статичним методом Віккерса (ПМТ-3). Встановлено, що значення мікротвердості біля залишкової пластичної вм'ятини зазнає істотних змін. Спочатку значення мікротвердості зменшується на $\approx 0,4$ ГПа і, залежно від того, уздовж яких взаємно перпендикулярних ліній відбувалося профілювання поверхні, дорівнює 0,026 ГПа та 0,07 ГПа. Потім значення мікротвердості збільшується до практично середніх значень 0,35 ГПа. Отже, зміна значення мікротвердості під час профілювання поверхні вказує на зміни фізико-механічних властивостей матеріалу в невеликій зоні поблизу місця зіткнення. Проте в цих зонах твердість матеріалу суттєво зменшується і за повторного обстрілу можливе повне пробиття багатошарового бронепакету.

Ключові слова: бронематеріал, ударно-хвильове навантаження, бронепробиття, мікротвердість, картографування мікротвердості.

Класифікація відповідно до AMS 2020: 74E30, 74E15.

Вступ

Для виготовлення сучасних багатокомпонентних бронежилетів використовуються різноманітні комбінації спеціальних захисних матеріалів, таких як Kevlar®, Twaron®, Artec®, Nomex®, Kermel®, Teijinconex®, Kynol®, а також надвискомолекулярний поліетилен (Dyneema®, Spectra®, Zylon®, M5® та ін.), виготовлені у формі волокон і тканин (Abtew et al., 2019). Для виробництва бронепанелей завдяки високій жорсткості і низькій щільності застосовують композитні та різні види керамічних матеріалів, таких як оксид, карбід і нітрид бору. Водночас пластини із броневих сталей і сплавів титану також не втрачають актуальності для виготовлення броні. Нині традиційно використовують високолеговані броньові сталі з такими високими характеристиками, як твердість, міцність, відносне видовження, модуль пружності, температура плавлення. Крім того, розробники броні пропонують дисперсійно зміцнені сплави на основі алюмінію, міді, більш дорогих і легких металів, таких як берилій і титан, або виготовлених на їхній основі композиційних систем. Отже, натеper вченими розроблено низку методів одержання бронестійких матеріалів і композицій (Crouch, 2019). Проте оскільки бронежилет має забезпечувати ефективний балістичний захист життєво важливих органів на рівні не менше 90 %, то для забезпечення такого високого рівня захисту бронежилети мають бути ретельно перевірені й оцінені за відповідними вимогами та критеріями (Yaneva, 2020; Tahenti, Coghe, & Nasri, 2020).

Сучасне застосування бронематеріалів вимагає підвищеної стійкості щодо дії високошвидкісного ударного навантаження, зниження їхньої питомої ваги, собівартості та спрощення технології виготовлення. Це можливо за умови науково-обґрунтованого підходу до їхньої розробки та виготовлення, що передбачає з'ясування механізмів руйнування бронематеріалів у разі дії на них високошвидкісного навантаження й оптимізації структурних особливостей відповідно до одержаних результатів.

Варто зауважити, що випробування зразків одержаної броні на балістичну стійкість проходять переважно лише на полігонах і зазвичай за неконтрольованих швидкостей і кутів атаки вражаючих елементів (Yaneva, 2020; Yun-ho, Jung-hoon, & Jong-Hwan, 2018). Цей шлях є досить витратним і малоінформативним. У зв'язку із цим виникає необхідність у розвитку нових підходів до змісту самих досліджень, що ґрунтуються на систематичних експериментальних дослідженнях динаміки процесів проникнення високошвидкісних ударних елементів у перепони з різних балістичностійких матеріалів.

У пропонованій роботі представлено методику експериментального моделювання проникнення високошвидкісних ударних елементів у бронематеріалах. Для ударних випробувань передбачається використовувати розроблений настільний розгінний стенд, що дає можливість формувати ударні хвилі в широкому діапазоні тисків, які змінюють структуру матеріалу перешкоди. Використання стенду для ударних випробувань і моделювання динаміки процесів проникнення

© Мельниченко Микола, Яровой Леонід, Пучко Наталія, Водотовка Михайло, 2025

високошвидкісних ударних елементів у перепони з різних бронестійких матеріалів допоможе розробити методичні особливості визначення динамічної стійкості бронематеріалів у великому діапазоні швидкостей і кутів атаки ударного навантаження. На основі отриманих параметрів стану бронематеріалів у зоні перебігу процесів руйнування будуть можливими атестація та розроблення рекомендацій щодо доцільності їхнього використання.

1. Основні означення та результати

Для постановки методики експериментального моделювання проникнення високошвидкісних ударних елементів у бронематеріалах у роботі було розроблено конструкцію та створено лабораторний експериментальний комплекс настільного розгінного стенду, у якому куля діаметром 8 мм зі сталі ШХ15 прискорюється енергією газів монтажного патрону під час руху каналом завдовжки 570 мм з керованою швидкістю від 200 до 1500 м/с та різними кутами атаки – від 45 до 90° (рис. 1).

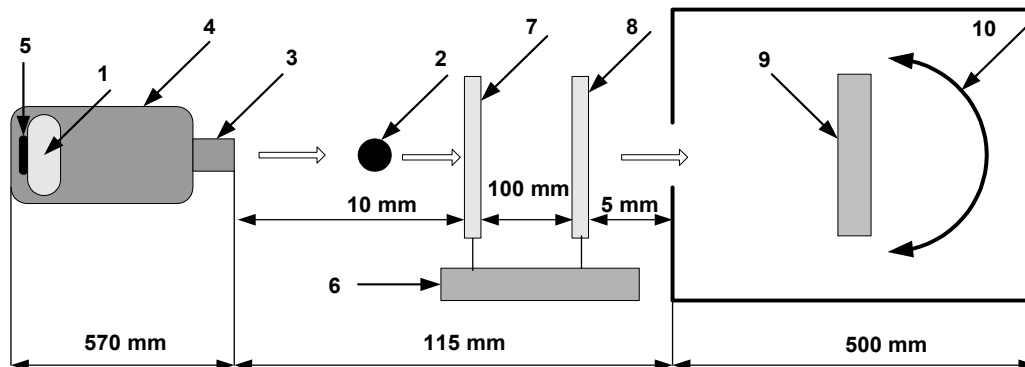


Рис. 1. Схематичне зображення настільного розгінного стенду

На рис. 1 наведено схематичне зображення настільного розгінного стенду. У розглянутій конструкції запропоновано використовувати енергію газів змінного монтажного набоя (1). Це дає можливість для розгону сталевих сферичних ударників діаметром 8 мм та масою 2,2 г (2) від 200 м/с до 1500 м/с. Фіксована довжина ствола (570 мм) забезпечує компактність розгінного блоку і всього стенду загалом. Ствол (3) міститься в захисній сталевій оболонці (4). Монтажний набій має електричний підпал (5), що дає змогу використовувати дистанційний пуск із сусіднього приміщення. Вимірювання початкової швидкості кулі здійснюється хронометром (6) за допомогою стартового (7) та кінцевого (8) тригерів із похибкою не більше 10 м/с. Мішень із бронематеріалу (9) встановлюємо поблизу хронографа на відстані 250–400 мм під різними кутами атаки (10). За експериментального моделювання проникнення кулі всі параметри стенду залишаються сталими. Змінним параметром є лише швидкість кулі. Для подальшого використання настільного розгінного стенду можна застосовувати стволи різного діаметру й ударні елементи не тільки у вигляді куль, а й також у вигляді стрижнів з різними наконечниками.

Варто зауважити, що втрата стійкості прямолінійного руху ударних елементів без стабілізуючого обертання навколо осі відбувається на великих відстанях від дульного зрізу, тому в запропонованому розгінному стенді вони мінімальні. Невеликі розміри стенду дають змогу встановлювати його в будь-яких приміщеннях і лабораторіях. Розроблений настільний розгінний стенд має можливість встановлення як у горизонтальному, так і у вертикальному положенні.

Для апробації розгінного стенду в роботі використовували зразки бронематеріалів зі сталі та багатшаровий бронематеріал із заліза та міді. Після обстрілу зі швидкістю 200 м/с та 600 м/с різних зразків бронематеріалів не спостерігаємо деформації кулі, а на поверхні з'являється пластична вм'ятини різного вигляду та різної глибини (рис. 2). Під час пострілу в один із досліджуваних бронематеріалів зі сталі (рис. 2 б) в залишковій вм'ятині спостерігаємо залишки процесу плавлення та витискування бронематеріалу із вм'ятини. Для іншого зразка бронематеріалу (рис. 2 в) такі процеси не виникають. Необхідно підкреслити, що після обстрілу багатшарового бронепакету також не спостерігаємо процесу плавлення (рис. 2 г). На рис. 2 д наведено зображення залишкової вм'ятини за невеликого оптичного збільшення у п'ять разів. При цьому відбувається багаторазове послідовне витискування бронематеріалу.

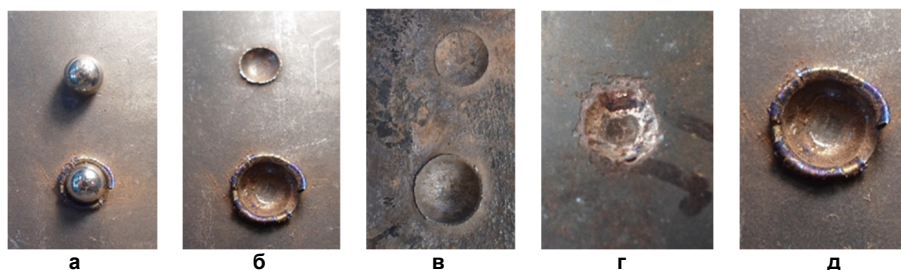


Рис. 2. Загальний вигляд поверхні після взаємодії кулі зі сталевими бронематеріалами (а, б, в, д) та сталевим-мідним багатшаровим бронематеріалом (г)

У роботі було проведено оцінку фізико-механічних характеристик досліджуваного багатшарового бронематеріалу. Для картографування механічних властивостей після взаємодії кулі з поверхнею виконано серію вимірювань мікротвердості статичним методом Віккерса (ПМТ-3).

Графіки зміни величини мікротвердості за зміни місця індентування побудовані на основі середніх значень мікротвердості, отриманих із 10 послідовних вдавнень з інтервалами 50 мкм за навантаження 200 г. На рис. 3 представлено

зміну значення мікротвердості вздовж двох взаємно перпендикулярних ліній не враженої поверхні та біля залишкової пластичної вмятини від кулі з початковою швидкістю 600 м/с. Значення мікротвердості на не враженій поверхні суттєво не змінюється (рис. 3 а).

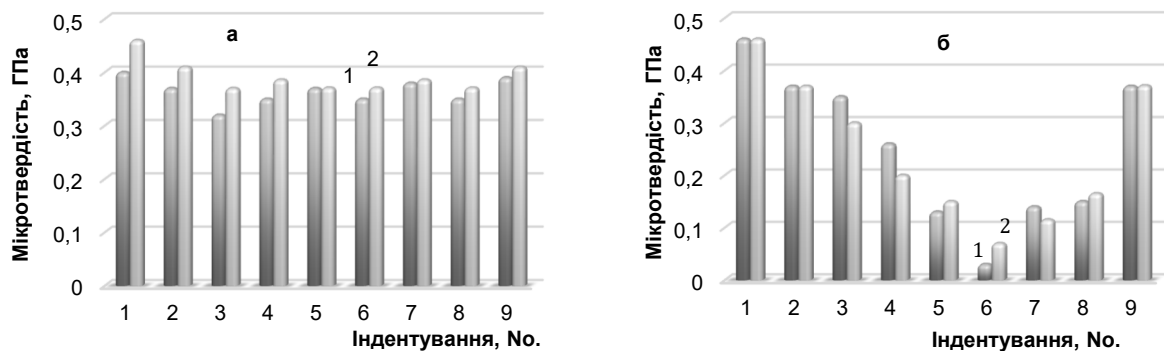


Рис. 3. Зміна значення мікротвердості вздовж двох взаємно перпендикулярних ліній (1, 2) на поверхні (а) та біля залишкової пластичної вмятини (б) багатошарового бронематеріалу

Із рис. 3 б видно, що значення мікротвердості біля залишкової пластичної вмятини зазнає значних змін. Спочатку значення мікротвердості зменшується на $\approx 0,4$ ГПа і, залежно від уздовж яких заданих ліній відбувалося профілювання поверхні, дорівнює 0,026 ГПа (рис. 3 б, 1) та 0,07 ГПа (рис. 3 б, 2). Потім значення мікротвердості збільшується до практично середніх значень 0,35 ГПа однаково за профілювання поверхні вздовж заданих ліній. Отже, зміна значення мікротвердості при профілюванні поверхні вказує на зміни фізико-механічних властивостей матеріалу в невеликій зоні поблизу місця зіткнення. Проте в цих зонах твердість матеріалу суттєво зменшується і за повторного обстрілу можливе повне пробиття багатошарового бронепакету.

Дискусія і висновки

Розроблено конструкцію та створено експериментальний комплекс настільного розгінного стенду, у якому куля діаметром 8 мм із сталі ШХ15 прискорюється енергією газів монтажного патрону під час руху каналом завдовжки 570 мм з керованою швидкістю від 200 до 1500 м/с та різними кутами атаки від 45 до 90°. Встановлено, що після обстрілу зі швидкістю 200 м/с та 600 м/с для різних зразків бронематеріалів з'являється пластична вмятина різного вигляду та різної глибини. Водночас у залишковій вмятині спостерігаємо залишки процесу плавлення та витискування бронематеріалу із вмятини. У разі обстрілу багатошарового бронепакету кулею зі швидкістю 600 м/с не відбувається деформації кулі. У роботі було проведено оцінку фізико-механічних характеристик досліджуваного багатошарового бронематеріалу. Для картографування механічних властивостей після високошвидкісної взаємодії кулі з поверхнею багатошарового бронематеріалу було виконано серію вимірювань мікротвердості біля місць взаємодії на поверхні зразка. Тестування на мікротвердість проводили статичним методом Віккерса (ПМТ-3). Графіки залежності величини мікротвердості у разі зміни місця індентування побудовані на основі середніх значень мікротвердості, отриманих із 10 послідовних вдавнень з інтервалами 50 мкм за навантаження 200 г. Встановлено, що значення мікротвердості біля залишкової пластичної вмятини зазнає істотних змін. Спочатку значення мікротвердості зменшується на $\approx 0,4$ ГПа і, залежно від уздовж яких взаємно перпендикулярних ліній відбувалося профілювання поверхні, дорівнює 0,026 ГПа та 0,07 ГПа. Потім значення мікротвердості збільшується до практично середніх значень 0,35 ГПа. Отже, зміна значення мікротвердості за профілювання поверхні вказує на зміни фізико-механічних властивостей матеріалу в невеликій зоні поблизу місця зіткнення. Проте в цих зонах твердість матеріалу суттєво зменшується і за повторного обстрілу можливе повне пробиття багатошарового бронепакету.

Внесок авторів: Микола Мельниченко – концептуалізація, методика, емпіричне дослідження; Леонід Яровой – збір і перевірка емпіричних даних; Наталя Пучко – аналіз джерел, підготовка огляду літератури; Михайло Водотовка – емпіричне дослідження.

Джерела фінансування. Дослідження було виконано за фінансової підтримки МОН України в межах виконання проекту "Методика дослідження динамічної стійкості та процесів руйнування бронематеріалів під дією ударного навантаження".

Список використаних джерел

- Abtew, M., Boussu, F., Bruniauxab, P., Loghin, C., & Cristian, I. (2019). Engineering of 3D warp interlock p-aramid fabric structure and its energy absorption capabilities against ballistic impact for body armour applications. *Composite Structures*, 225, 111179.
- Crouch, I. (2019). Body armour – New materials, new systems. *Defence Technology*, 15(3), 241–253.
- Tahenti, B., Coghe, F., & Nasri, R. (2020). Ballistic Limit Estimation Approaches for Ballistic Resistance Assessment. *Defence Science Journal*, 70(1), 82–89. <https://doi.org/10.14429/dsj.70.14122>
- Yaneva, S. (2020). Ballistic resistance of bulletproof vests level IIIA. Development of testing methodology. *MATEC Web Conference*, 317, 06003. <https://doi.org/10.1051/mateconf/202031706003>
- Yun-ho, S., Jung-hoon, C., & Jong-Hwan, K. (2018). Test and estimation of ballistic armor performance for recent naval ship structural materials. *International Journal of Naval Architecture and Ocean Engineering*, 10, 762–781.

References

- Abtew, M., Boussu, F., Bruniauxab, P., Loghin, C., & Cristian, I. (2019). Engineering of 3D warp interlock p-aramid fabric structure and its energy absorption capabilities against ballistic impact for body armour applications. *Composite Structures*, 225, 111179.
- Crouch, I. (2019). Body armour – New materials, new systems. *Defence Technology*, 15(3), 241–253.

- Tahenti, B., Coghe, F., & Nasri, R. (2020). Ballistic Limit Estimation Approaches for Ballistic Resistance Assessment. *Defence Science Journal*, 70(1), 82–89. <https://doi.org/10.14429/dsj.70.14122>
- Yaneva, S. (2020). Ballistic resistance of bulletproof vests level IIIA. Development of testing methodology. *MATEC Web Conference*, 317, 06003. <https://doi.org/10.1051/mateconf/202031706003>
- Yun-ho, S., Jung-hoon, C., & Jong-Hwan, K. (2018). Test and estimation of ballistic armor performance for recent naval ship structural materials. *International Journal of Naval Architecture and Ocean Engineering*, 10, 762–781.

Отримано редакцією журналу / Received: 25.06.25
Прорецензовано / Revised: 06.10.25
Схвалено до друку / Accepted: 10.10.25

Mykola MELNICHENKO¹, PhD (Phys. & Math.), Senior Researcher
ORCID ID: 0000-0002-3473-0701
e-mail: mykola.melnychenko@knu.ua

Leonid YAROVOL¹, PhD (Engin.)
ORCID ID: 0000-0001-6374-4160
e-mail: yarovoi@univ.kiev.ua

Natalia PUCHKO¹, Leading Engineer
e-mail: natalia.puchko@knu.ua

Mykhailo VODOTOVKA¹, Electrical Mechanic
e-mail: nanocrystallab@univ.kiev.ua

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

METHOD OF EXPERIMENTAL MODELING OF PENETRATION OF HIGH-VELOCITY IMPACT ELEMENTS IN ARMOR MATERIALS

The work developed and created an experimental complex of a desktop acceleration stand in which a ball with a diameter of 8 mm made of ШХ15 steel is accelerated by the energy of gases from a mounting cartridge while moving along a channel 570 mm long with a controlled speed from 200 to 1500 m/s and at various angles of attack from 45 to 90°. The acceleration stand was tested. It was found that after firing at speeds of 200 m/s and 600 m/s, various types of armor materials exhibited plastic dents of various shapes and depths. At the same time, the residual dent shows traces of the process of melting and extrusion of the armor material from the dent. When firing a multi-layer armor package with a bullet at a speed of 600 m/s, no deformation of the bullet is observed. To evaluate the physical and mechanical characteristics of the multilayer armor material under study, a series of microhardness measurements was performed at the points of interaction on the sample surface. Microhardness testing was performed using the Vickers static method (PMT-3). It was found that the microhardness value near the residual plastic indentation undergoes significant changes. Initially, the microhardness value decreases by ≈ 0.4 GPa and, depending on the mutually perpendicular lines along which the surface profiling took place, is equal to 0.026 GPa and 0.07 GPa. Then the microhardness value increases to practically average values of 0.35 GPa. Therefore, a change in microhardness during surface profiling indicates changes in the physical and mechanical properties of the material in a small area near the point of impact. At the same time, the hardness of the material in these areas is significantly reduced, and repeated firing may result in complete penetration of the multilayer armor package.

Keywords: armor material, shock wave loading, armor penetration, microhardness, microhardness mapping.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 539.3

DOI: <https://doi.org/10.17721/1812-5409.2025/2.19>

Володимир ОНИШКЕВИЧ, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0002-4657-5462

e-mail: onyshkevych@ntu.edu.ua

Національний лісотехнічний університет України, Львів, Україна

Галина БАРАБАШ, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0003-3258-8518

e-mail: galynabarabash71@gmail.com

Львівський національний університет імені Івана Франка, Львів, Україна

КВАЗІСТАТИЧНА КОНТАКТНА ВЗАЄМОДІЯ СКІНЧЕННОГО ЖОРСТКОГО ЦИЛІНДРА, ЩО ОБЕРТАЄТЬСЯ, ТА ПРУЖНОГО ПІВПРОСТОРУ ЗА ФРИКЦІЙНОГО ТЕПЛОУТВОРЕННЯ

Існує багато математичних моделей, що покликані описати процес контактування тіл, проте розв'язати більшість теоретичних і практичних трибологічних проблем неможливо без розгляду геометричних характеристик контактної взаємодії. Однак у контактних задачах термопружності фактори скінченності розмірів штампів та криволінійності його основи зазвичай не враховують.

На прикладі осесиметричної задачі про втискування в пружний півпростір скінченного жорсткого кругового циліндра, що обертається навколо своєї осі з постійною кутовою швидкістю, у пропонованій роботі розглянуто методику, яка дає можливість змодельовувати і дослідити широкий спектр явищ, що можуть виникати за взаємодії тіл: відокремлення контактних поверхонь, затинання, одночасне існування зон зчеплення і ковзання. На ділянці контакту внаслідок дії сил тертя утворюється тепло, і неідеальний тепловий контакт між циліндром і півпростором математично моделюється за допомогою "третього тіла" – тонких приповерхневих і проміжкових шарів контактних тіл, фізико-механічні властивості яких відрізняються від властивостей тіл контактної пари. Для кожного з контактних тіл розв'язано нестационарне рівняння теплопровідності з початковою умовою на температуру, у решті вихідних співвідношень часова змінна фігурує як параметр. Використовуючи інтегральне перетворення Лапласа за часом, застосовуючи інтегральне перетворення Ханкеля за радіальною координатою для півпростору та метод прямих для циліндра, отримано розподіл температури та теплових полів у циліндрі й півпросторі, контактних напружень і переміщень півпростору. Обернення трансформант шуканих функцій здійснювали числово. Виявлено якісно відмінні етапи загальної картини квазістатичної взаємодії для поставленої задачі. З'ясовано, що перебіг відповідних процесів визначається конкретними навантаженнями і характеристиками пари тертя. Відзначено, що квазістатичні контактні задачі доцільно розглядати тоді, коли початкові дані та характер навантаження спричиняють перевищення максимальних значень досліджуваних величин відповідних характеристик у стаціонарному режимі.

Ключові слова: контакт, тертя, розігрів, теплопровідність, термопружність.

Класифікація відповідно до AMS 2020: 74F05.

Вступ

Визначення температурних полів у контактних тілах під час їхнього відносного ковзання є однією з актуальних задач трибології. Різноманітність чинників впливу на тепловий режим пари тертя зумовлює складність опису та математичного моделювання такої практичної науково-технічної задачі. Уперше контактну взаємодію в термопружності розглядали в роботі (George, & Sneddon, 1962), де було досліджено розподіл напружень і переміщень у пружному півпросторі під час втискування в нього жорсткого штампів з певною температурою. Однак у цій та більшості інших публікацій не брали до уваги випинання основи і, як наслідок, цілком можливе відставання деяких ділянок штампів від основи. За спрощених граничних умов задачу неповного контакту плоскої основи штампів з пружним півпростором було розглянуто в роботі (Barber, 1967), дослідження ґрунтувалося на тому, що тепло, яке утворюється внаслідок тертя, розподіляється між двома контактними поверхнями. Математично моделювати режими тертя, зношування та теплоутворення за допомогою розгляду так званого умовного "третього тіла" – тонких приповерхневих і проміжкових шарів контактних тіл, фізико-механічні властивості яких відрізняються від властивостей тіл контактної пари, та мікрогеометрію поверхонь тіл у контактній зоні, було запропоновано у роботі (Levyts'kyi, & Onyshkevych, 2002). Для опису контактної взаємодії було використано узагальнені умови теплового контакту, що враховують узагальнену характеристику – коефіцієнт теплопроникності контакту. Математичному моделюванню та розробленню аналітичних і числових методів дослідження контактної взаємодії тіл різної геометричної форми і фізичної структури з урахуванням шорсткості їхніх меж, проковзування і розшарування спряжених поверхонь присвячено монографію (Мартиняк, & Маланчук та ін., 2022). Однак універсальних аналітико-числових методів розв'язування контактних задач у квазістационарній постановці немає, наприклад, у роботі (Онишкевич, & Барабаш, 2023) запропоновано метод послідовно-дискретного розділення граничних умов з використанням рухомої лінії розділення граничних умов, що може бути використане для розв'язування нестационарної задачі шляхом зведення її до послідовності задач із незмінними ділянками прикладання видозмінених граничних умов.

Об'єктом дослідження є контактна взаємодія жорсткого кругового циліндра скінченної довжини, що обертається навколо своєї осі з постійною кутовою швидкістю і втискується у пружний півпростір за наявності теплоутворення на ділянці контакту від дії сили тертя.

Мета та завдання дослідження – математична постановка осесиметричної квазістационарної задачі термопружності для цієї трибологічної пари з урахуванням теплоутворення в зоні контакту; моделювання неідеального теплового

контакту за допомогою "третього тіла"; розроблення методики розв'язування поставленої задачі з урахуванням можливого відокремлення контактних поверхонь, затинання, одночасного існування зон зчеплення і ковзання.

1. Постановка задачі

Розглянуто циліндричний штамп із радіусом R і висотою H , що втискується вертикальною силою P у пружний півпростір й обертається навколо своєї осі з постійною кутовою швидкістю ω . Вважаємо, що в початковий момент часу температура циліндра і півпростору є нульовими. Між верхньою поверхнею штампа, бічними поверхнями штампа і ненавантаженою поверхнею півпростору з одного боку та зовнішнім середовищем з іншого здійснюється теплообмін за законом Ньютона з різними коефіцієнтами теплообміну. Тепловий контакт між штампом і півпростором є неідеальним, його моделюємо за допомогою так званого "третього тіла" (Levyts'kyi, & Onyshkevych, 2002) – тонких приповерхневих і проміжкових шарів контактних тіл, фізико-механічні властивості яких відрізняються від властивостей тіл контактної пари, та мікрогеометрією поверхонь тіл у контактній зоні, з використанням узагальненої характеристики – коефіцієнта теплопроникності контакту h . Від дії сили тертя на ділянці контакту відбувається теплоутворення, що пропорційне інтенсивності сили тертя. Температуру зовнішнього середовища вважаємо нульовою. Розв'язування задачі полягає в тому, щоб задовольнити:

1) рівняння теплопружності для півпростору

$$\Delta u_r - \frac{u_r}{r^2} + k \frac{\partial e}{\partial r} = \frac{\beta}{\mu} \frac{\partial t}{\partial r}, \Delta u_z + k \frac{\partial e}{\partial z} = \frac{\beta}{\mu} \frac{\partial t}{\partial z}, \Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}, e = \frac{\partial u_r}{\partial r} + \frac{u_r}{r} + \frac{\partial u_z}{\partial z}, \quad (1)$$

2) рівняння теплопровідності для циліндра і півпростору

$$\Delta T = X_1 \frac{\partial T}{\partial \tau}, \Delta t = X_2 \frac{\partial t}{\partial \tau}; \quad (2)$$

3) температурні граничні умови

$$z = 0: \frac{\partial T}{\partial z} = \gamma_0 T (0 \leq r \leq R), r = R: \frac{\partial T}{\partial r} = -\gamma_a T (0 \leq z \leq l), \quad (3)$$

$$z = H: \lambda_2 \frac{\partial t}{\partial z} - \lambda_1 \frac{\partial T}{\partial z} = f_T \omega r \sigma_{zz}, \lambda_2 \frac{\partial t}{\partial z} + \lambda_1 \frac{\partial T}{\partial z} = h(t - T) (0 \leq r \leq R), \frac{\partial t}{\partial z} = \gamma_H t (r > R) \quad (4)$$

4) силові граничні умови

$$z = H: u_z = f (0 \leq r \leq R), \sigma_{zz} = 0 (r \geq R), \tau_{rz} = 0 (r < \infty); \quad (5)$$

5) початкові умови

$$\tau = 0: t = T = 0, f = \partial f / \partial \tau = 0. \quad (6)$$

Тут $k = (\lambda + \mu) / \mu = 1 / (\lambda - 2\nu)$, ν – коефіцієнт Пуассона, $\beta = (3\lambda + 2\mu)\alpha_T$, λ та μ – коефіцієнти Ламе, α_T – температурний коефіцієнт лінійного розширення, $f(\tau)$ – величина осідання штампа, λ_1 та λ_2 – коефіцієнти теплопровідності відповідно для штампа і півпростору, $\gamma_0 = \gamma_a = \gamma_H$ – коефіцієнти теплообміну із зовнішнім середовищем відповідно для торця циліндра, його бічної поверхні і невантаженої поверхні півпростору, T – температура циліндричного штампа, t – температура півпростору, $X_i = 1/a_i (i = 1, 2)$, a_i – коефіцієнти теплопровідності ($i = 1$ відповідає штампі, $i = 2$ відповідає півпростору), h – коефіцієнт теплопроникності контакту, $\tau_{z\theta} = f_T \sigma_{zz}$, де f_T – коефіцієнт тертя. Функції $t, T, u_r, u_z, \sigma_{zz}, \tau_{rz}$ є функціями від трьох дійсних змінних r, z, τ . Вважаємо, що штамп має закруглені краї і в області контакту є обмежений поверхнею обертання, що складається із двох ділянок: за $0 \leq r < r^*$ – площа, а за $r^* \leq r \leq R$ – параболоїд з вершиною в точці r^* :

$$f(r, \tau) = f_1(\tau) - \frac{(r-r^*)^2}{2R_p} H(r-r^*), f_1(\tau) = \text{const}, \quad (7)$$

де величина r^* близька до R , R_p – радіус кривини країв штампа, $H(r-r^*)$ – функція Хевісайда. Під час розв'язування задачі будемо використовувати умову динамічної рівноваги штампа

$$m \frac{\partial^2 f_1(\tau)}{\partial \tau^2} = P(\tau) + 2\pi \int_0^R r \sigma_{zz}(r, H, \tau) dr, \quad (8)$$

де m – маса штампа, P – притискна вертикальна сила.

2. Метод розв'язування задачі

Застосуємо інтегральне перетворення Лапласа за часом до рівнянь (2), урахувавши початкові умови (6). Використовуючи метод прямих і задовольняючи другу умову (3), отримане рівняння у безрозмірних змінних $\rho = r/R, \zeta = z/H$ можна записати в кожній з N точок розбиття $\rho_i = \Delta\rho(i-1) = (i-1)/(N-1), (i = 1, \dots, N)$ та звести задачу теплопровідності для циліндра до системи лінійних диференціальних рівнянь виду $d\vec{w}/d\zeta = B\vec{w}$. Її розв'язок має вигляд

$$\vec{w}^T(\zeta, s) = \left(T^L(\rho_1, \zeta, s), \dots, T^L(\rho_N, \zeta, s), \frac{dT^L(\rho_1, \zeta, s)}{d\zeta}, \dots, \frac{dT^L(\rho_N, \zeta, s)}{d\zeta} \right) \quad (9)$$

і будується за допомогою матричної експоненти

$$\vec{w}(\zeta, s) = \exp(B(s)\zeta) \vec{d}(s) = \sum_{n=0}^{\infty} \frac{B^n(s)\zeta^n}{n!} \vec{d}(s), \quad (10)$$

де $B(s)$ – відома матриця розмірності $2N \times 2N$, а вектор $\vec{d}(s)$ визначається з граничних умов (3).

Застосовуючи інтегральне перетворення Ханкеля нульового порядку по радіальній координаті до рівняння для півпростору в трансформантах Лапласа, отримано вирази для подвійних трансформант шуканих напружень, переміщень, температури та теплових потоків. Переходячи у формулах для $t^{LH}(\xi, H, s), q^{LH}(\xi, H, s), u_z^{LH}(\xi, H, s), \sigma_{zz}^{LH}(\xi, H, s)$ до безрозмірного параметра $\eta = \xi R$, застосовуючи обернене перетворення Ханкеля, задовольняючи граничні умови (4) – (5) та продовжуючи отримані вирази на всю вісь ρ за допомогою функції Хевісайда, подамо невідомі функції у вигляді розкладів у ряди Фур'є – Бесселя

$$\phi(\rho, s) = b_0(s) + \sum_{n=1}^N b_n(s) J_0(\mu_n \rho), \sigma_{zz}^L(\rho, 1, s) = \sum_{n=1}^N a_n(s) J_0(\mu_n \rho), \quad (11)$$

де μ_n – нулі функції Бесселя нульового порядку $J_0(x)$. Виразивши функції через невідомі коефіцієнти, задовольняючи (8) та першу граничну умову (3) в кожній точці рівномірного розбиття ρ_i , прийдемо до системи рівнянь з $4N$ невідомими: $a_n(s), b_n(s) (n = 1, \dots, N)$ та $d_n(s) (n = 1, \dots, 2N)$. Обчислення оригіналів функцій $t^L(\rho, H, s), q^L(\rho, H, s), u_z^L(\rho, H, s)$ здійснювали числовим методом у вигляді розвинення в ряд Фур'є за функціями $\sin(2k+1)\theta$, які для дійсних значень $s_n = (2n+1)\sigma (n = 0, 1, \dots, N^*)$ утворюють повну ортогональну систему.

3. Числові результати

Під час обчислень вважали, що притискна вертикальна сила P задається рівнянням $P(\tau) = F(1 - \exp(-a\tau))$, де a та F є деякими константами. Розглядали, що матеріал циліндра – сталь, а матеріал півпростору – алюміній.

На рис. 1 зображено розподіл контактних напружень, а на рис. 2 – температура на границі півпростору в різні моменти часу.

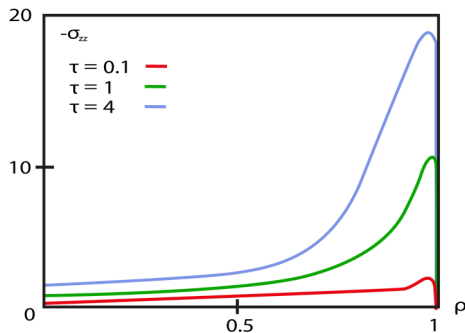


Рис. 1. Розподіл контактних напружень (МПа)

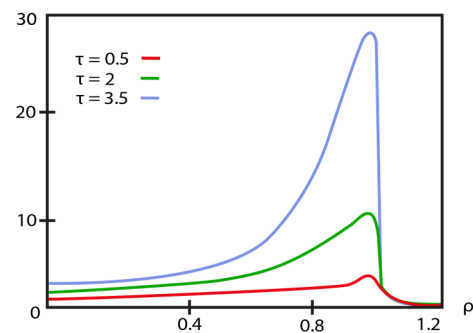


Рис. 2. Температура на границі півпростору (°C)

Дискусія і висновки

На основі проведених числових обчислень й аналізу ділянок зміни знаку контактних напружень з'ясовано, що загальну картину квазістатичної взаємодії у задачі можна описати такими якісно відмінними етапами:

- повне зчеплення основи штампа із суміжною ділянкою півпростору; теплоутворення не відбувається;
- часткове зчеплення основи штампа з контактною ділянкою півпростору; на зовнішньому кільці контакту спостерігаємо проковзування, а відтак і теплоутворення;
- на всій ділянці контакту (за винятком малого околу осевої точки) діють сили тертя і відбувається фрикційний нагрів; контактні напруження перерозподіляються, але радіус ділянки контакту не змінюється;
- відокремлення частини основи циліндра від поверхні півпростору як наслідок дії термічних факторів;
- стаціонарна стадія; жодна із шуканих функцій з часом не змінюється.

Перебіг описаних процесів визначається конкретними навантаженнями і характеристиками пари тертя. У випадку, коли притискна сила і крутий момент, поступово зростаючи, з часом стабілізуються, можна отримати узгодження з відомими класичними результатами роботи (George, & Sneddon, 1962). Квазістатичні контактні задачі доцільно розглядати тоді, коли початкові дані і характер навантаження спричиняються до того, що максимальні значення досліджуваних величин перевищують відповідні характеристики стаціонарного режиму. Математичне моделювання виявлених явищ недосконалого контакту планується продовжити в подальших дослідженнях.

Внесок авторів: Володимир Онишкевич – постановка задачі, методика розв'язування; числові обчислення; Галина Барабаш – аналіз джерел, огляд літератури, збір і перевірка емпіричних даних.

Джерела фінансування. Фінансування здійснюється власним коштом авторів і частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Мартиняк, Р. М., & Маланчук, Н. І. (2022). *Контактна механіка. Шорсткість, розшарування і зношування поверхонь* : кол. монограф. Видавець Вікторія Кундельська. <https://www.researchgate.net/publication/366177313>
- Онишкевич, В. М., & Барабаш, Г. М. (2023). Фрикційний розігрів системи штамп-пружна півплощина при ковзанні вздовж твірної. *Вісник КНУ ім. Т. Шевченка. Серія: Фізико-математичні науки*, 2, 140–143. <https://doi.org/10.17721/1812-5409.2023/2.23>
- Barber, J. R. (1967). Distribution of heat between sliding surfaces. *Journal of Mechanical Engineering Science*, 9(5), 351–354.
- George, D. L., & Sneddon, I. N. (1962). The axisymmetric Boussinesq problem for a heat punch. *Journal of Mathematics and Mechanics*, 11(5), 665–689.
- Levyts'kyi, V. P., & Onyshkevych, V. M. (2002). Investigation of the influence of properties of a third body on heat generation due to friction. *Journal of Mathematical Sciences*, 109(1), 1251–1256. <https://doi.org/10.1023/A:1013752929389>

References

- Barber, J. R. (1967). Distribution of heat between sliding surfaces. *Journal of Mechanical Engineering Science*, 9(5), 351–354.
- George, D. L., & Sneddon, I. N. (1962). The axisymmetric Boussinesq problem for a heat punch. *Journal of Mathematics and Mechanics*, 11(5), 665–689.
- Levyts'kyi, V. P., & Onyshkevych, V. M. (2002). Investigation of the influence of properties of a third body on heat generation due to friction. *Journal of Mathematical Sciences*, 109(1), 1251–1256. <https://doi.org/10.1023/A:1013752929389>
- Martyniak, R. M., & Malanchuk, N. I. (2022). *Contact mechanics. Roughness, layering and surface wear: coll. Monograph*. Publisher Viktoria Kundelska [in Ukrainian]. <https://www.researchgate.net/publication/366177313>
- Onyshkevych, V. M., & Barabash, G. M. (2023). Frictional heating of the system punch-elastic half plane when sliding along the creative line. *Bulletin of Taras Shevchenko National University of Kyiv. Series Physics & Mathematics*, 2, 140–143 [in Ukrainian]. <https://doi.org/10.17721/1812-5409.2023/2.23>

Отримано редакцією журналу / Received: 15.08.25
Прорецензовано / Revised: 02.10.25
Схвалено до друку / Accepted: 10.10.25

Volodymyr ONYSHKEVYCH, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-4657-5462
e-mail: onyshkevych@nltu.edu.ua
Ukrainian National Forestry University, Lviv, Ukraine

Galyna BARABASH, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0003-3258-8518
e-mail: galynabarabash71@gmail.com
Ivan Franko National University of Lviv, Lviv, Ukraine

THE UNSTEADY CONTACT INTERACTION OF A ROTATING STIFF FINITE CYLINDER AND ELASTIC HALF-SPACE WITH FRICTIONAL HEAT GENERATION

There are several mathematical models that help describe the contacting process for bodies, but solving the majority of theoretical and practical problems is not impossible without considering the geometric characteristics of contact interaction. However, in the contact problems of thermoelasticity factors of finiteness of stiff dimensions and curvilinearity of its basis are not taken into account mostly.

In the example of an axisymmetric problem about pressing a finite rigid circular cylinder, rotating around its axis with constant angular velocity, into an elastic half-space, we consider the method that allows us to model and investigate a wide range of phenomena arising at the interaction of bodies: separation of contact surfaces, wedging, and simultaneous existence of adhesion and sliding zones. On the contact area, heat is generated by friction forces, and non-ideal heat contact between cylinder and half-space is mathematically modeled using "third body" – thin near-surfaces and intermediate layers of interacting bodies, the physical and mechanical properties of which differ from those of the contact pair, and by the microgeometry of their surfaces in the contact zone. For every contact body, the non-stationary equation of thermal conductivity with initial condition on temperature is solved, and the time variable appears as a parameter in the rest initial ratios. A Laplace integral transformation with respect to the time coordinate and a Hankel integral transformation with respect to the radial coordinate for the half-space and the method of straight lines cylinder are used to solve the problem. The temperature and thermal flux fields in the cylinder and in the half-space, the contact stresses, and the displacements of the half-space are determined. The inversion of the transforms of the desired functions was carried out numerically. Qualitatively distinct stages of the overall picture of unsteady interaction for the formulated problem have been identified. It was found that the course of the relevant processes is determined by specific loads and friction pair characteristics. It is noted that quasi-stationary contact problems should be considered when the initial data and the nature of the load cause the maximum values of the studied quantities of the corresponding characteristics to be exceeded in the stationary mode.

Key words: contact, friction, heating, thermal conductivity, thermoelasticity.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

ФУНДАМЕНТАЛЬНИЙ РОЗВ'ЯЗОК ЗАДАЧІ КОШІ ДЛЯ ДИСИПАТИВНОГО УЛЬТРАПАРАБОЛІЧНОГО РІВНЯННЯ КОЛМОГОРОВА З НЕЗАЛЕЖНИМИ ВІД ЗМІННИХ ВИРОДЖЕННЯ КОЕФІЦІЄНТАМИ

Для ультрапараболічного рівняння Колмогорова з коефіцієнтами, що не залежать від змінних виродження і є обмеженими за просторовою змінною, побудовано фундаментальний розв'язок задачі Коші. Зростання коефіцієнтів рівняння залежить від зростаючої функції, яку називаємо характеристикою дисипації. Коефіцієнти рівняння задовольняють спеціальну умову Гельдера, що містить характеристику дисипації. Диференційовність коефіцієнтів не вимагається, тому на характеристику накладено додаткову умову. Ультрапараболічні рівняння Колмогорова зі зростаючими коефіцієнтами виникають у задачах теорії випадкових процесів, у статистичній радіотехніці.

Фундаментальний розв'язок задачі Коші побудовано методом Леві. За параметрикс узято фундаментальний розв'язок задачі Коші розглядуваного рівняння з коефіцієнтами, залежними від часової та параметричної просторової змінної. Це ускладнює дослідження, оскільки у випадку обмежених коефіцієнтів за параметрикс береться фундаментальний розв'язок задачі Коші рівняння із замороженими коефіцієнтами тільки групи старших похідних.

Отримані оцінки похідних фундаментального розв'язку задачі Коші можуть бути використані під час вивчення питань розв'язності задачі Коші й інтегрального зображення розв'язків у спеціальних вагових просторах, а також для побудови класичного фундаментального розв'язку задачі Коші для виродженого рівняння типу Колмогорова зі зростаючими за $|x| \rightarrow \infty$ коефіцієнтами, що залежать і від змінних виродження, і систем таких рівнянь.

Ключові слова: ультрапараболічне рівняння Колмогорова, зростаючі коефіцієнти, фундаментальний розв'язок задачі Коші, метод Леві.

Класифікація відповідно до AMS 2020: 35K70, 35A08.

Вступ

Поняття фундаментального розв'язку задачі Коші є основним у дослідженні розв'язності задачі Коші. Найповніші результати про фундаментальний розв'язок задачі Коші одержано для параболічних за І. Петровським та за С. Ейдельманом рівнянь і систем рівнянь з обмеженими і зростаючими за $|x| \rightarrow \infty$ коефіцієнтами. Вони викладені в монографії (Eidelman, Ivasyshen, & Kochubei, 2004) та цитованих у ній публікаціях. Побудові і дослідженню фундаментального розв'язку задачі Коші для ультрапараболічних рівнянь присвячені праці С. Ейдельмана, Г. Малицької, Л. Тичинської, С. Івасишена, Л. Андросової, В. Дроня. Переважно ці результати проаналізовані та викладені у згаданій монографії, де побудовано фундаментальний розв'язок задачі Коші для виродженого рівняння типу Колмогорова з коефіцієнтами, залежними від змінної t . У працях італійських математиків вивчаються L -розв'язки задачі Коші ультрапараболічних рівнянь із блочною структурою (Di Francesco, & Pascucci, 2005). Зважаючи на різноманітні застосування, дослідження ультрапараболічних рівнянь здійснюється також іншими методами без побудови фундаментального розв'язку задачі Коші (Процах, & Пташник, 2017).

Результати про фундаментальні розв'язки задачі Коші для ультрапараболічних рівнянь з обмеженими коефіцієнтами досить точні, однак такими вони не є у випадку зростаючих коефіцієнтів. Побудові фундаментального розв'язку задачі Коші для рівняння типу Колмогорова з обмеженими коефіцієнтами, залежними від усіх змінних, присвячено ряд праць С. Івасишена та І. Медінського. У статті (Ivasyshen, & Medyn's'ky, 2018) реалізовується поетапне застосування методу Леві, причому коефіцієнти рівняння задовольняють спеціальну умову Гельдера. Для виродженого параболічного рівняння другого порядку зі сталими у групі старших і зростаючими у групі молодших похідних із двома групами змінних виродження в праці (Івасишен, & Пасічник, 2014) одержано в явному вигляді формулу фундаментального розв'язку задачі Коші.

У пропонованій статті розглянуто ультрапараболічне рівняння типу Колмогорова із двома групами змінних і коефіцієнтами, що не залежать від змінних виродження, а їхнє зростання обумовлене деякою зростаючою функцією. Як і у випадку невироджених рівнянь, для коефіцієнтів розглядуваного рівняння вводиться дві серії умов. Тут розглянуто випадок гелдерових коефіцієнтів із додатковою умовою на функцію, від якої залежить ріст коефіцієнтів. Для побудови фундаментального розв'язку задачі Коші використано метод Леві. З огляду на це значна увага приділена вивченню властивостей головного члена формули фундаментального розв'язку задачі Коші.

1. Основні означення та результати

У шарі $\Pi_{(0,T]} := (0, T] \times \mathbb{R}^n$ скінченної товщини $T > 0$ розглядаємо рівняння

$$Lu(t, x_1, \partial_t, \partial_x) := \left(S - \sum_{j,s=1}^{n_1} a_{js}(t, x_1) \partial_{x_{1j}} \partial_{x_{1s}} - \sum_{j=1}^{n_1} a_j(t, x_1) \partial_{x_{1j}} - a_0(t, x_1) \right) u(t, x) = 0. \quad (1)$$

Тут n_1, n_2 – задані натуральні числа, причому $n_2 \leq n_1$, $n = n_1 + n_2$, змінна $x = (x_1, x_2) \in \mathbb{R}^n$ складається із двох груп, де $x_1 = (x_{11}, \dots, x_{1n_1}) \in \mathbb{R}^{n_1}$, $l \in \{1, 2\}$, $S := \partial_t - \sum_{j=1}^{n_2} x_{1j} \partial_{x_{2j}}$. Рівняння (1) є виродженим рівнянням типу Колмогорова другого порядку (ультрапараболічним рівнянням типу Колмогорова), його коефіцієнти $a_{js}, \{j, s\} \subset \{1, \dots, n_1\}$, $a_j, j \in \{1, \dots, n_1\}$ і a_0 не залежать від змінних виродження $x_{2j}, j \in \{1, \dots, n_2\}$.

Означення. Рівняння (1) називатимемо дисипативним ультрапараболічним рівнянням типу Колмогорова в $\Pi_{(0,T]}$, якщо існує неперервна функція $D: R^{n_1} \rightarrow [1, \infty)$, яка задовольняє такі умови:

- 1) $D(x_1) \rightarrow \infty$ при $|x_1| \rightarrow \infty$;
- 2) функції $b_{js}(t, x_1) = a_{js}(t, x_1)$, $b_j(t, x_1) = a_j(t, x_1)D(x_1)^{-1}$, $\{j, s\} \subset \{1, \dots, n_1\}$, $b_0(t, x_1) = a_0(t, x_1)D(x_1)^{-2}$, $x_1 \in R^{n_1}$, $0 \leq t \leq T$, обмежені;
- 3) для рівняння

$$\left(S - \sum_{j,s=1}^{n_1} b_{js}(t, x_1) \partial_{x_{1j}} \partial_{x_{1s}} - \sum_{j=1}^{n_1} b_j(t, x_1) \partial_{x_{1j}} (-i \partial_{x_{n+1}}) - b_0(t, x_1) (-i \partial_{x_{n+1}})^2 \right) v(t, x) = 0$$

з обмеженими коефіцієнтами і додатковою просторовою змінною x_{n+1} виконується умова параболічності

$$\exists \delta > 0 \quad \forall \sigma_1 \in R^{n_1} \quad \forall \mu \in R: \operatorname{Re} \left(- \sum_{j,s=1}^{n_1} b_{js}(t, x_1) \sigma_{x_{1j}} \sigma_{x_{1s}} - \sum_{j=1}^{n_1} b_j(t, x_1) \sigma_{x_{1j}} \mu - b_0(t, x_1) \mu^2 \right) \geq \delta (|\sigma_1|^2 + \mu^2).$$

Функція D за такої умови називається характеристикою дисипації рівняння (1).

Припустимо, що виконуються такі умови на коефіцієнти рівняння (1).

B_1 . Рівняння (1) є дисипативним ультрапараболічним рівнянням типу Колмогорова в $\Pi_{(0,T]}$, із характеристикою дисипації D .

B_2 . $\exists C > 0 \quad \exists \lambda \in (0,1] \quad \forall \{x_1, y_1\} \subset R^{n_1} \quad \forall t \in [0, T]: |\Delta_{x_1}^{y_1} a_{js}(t, x_1)| \leq C |x_1 - y_1|^\lambda$,

$$|\Delta_{x_1}^{y_1} a_j(t, x_1)| \leq C |x_1 - y_1|^\lambda (D(x_1) + D(y_1)), \quad |\Delta_{x_1}^{y_1} a_j(t, x_1)| \leq C |x_1 - y_1|^\lambda \left((D(x_1)) + (D(y_1))^2 \right);$$

функції b_{js} , b_j , $\{j, s\} \subset \{1, \dots, n_1\}$, b_0 , як функції t , є неперервними рівномірно щодо $x_1 \in R^{n_1}$.

B_3 . Характеристика дисипації D задовольняє такі умови:

- 1) $\exists C > 0 \quad \forall \{x_1, y_1\} \subset R^{n_1}, |x_1 - y_1| \leq 1: D(x_1) \leq CD(y_1)$;
- 2) $\exists C > 0 \quad \forall \{x_1, y_1\} \subset R^{n_1}, |x_1 - y_1| > 1: D(x_1) \leq C \exp \{ \varepsilon |x_1 - y_1| D(y_1) \}$,

де ε – досить мале додатне число, вибором якого в конкретній ситуації будемо оперувати.

Нехай $g: R^{n_1} \rightarrow [1, \infty)$ – функція, що пов'язана із характеристикою дисипації D умовою

B_4 . $g(x_1) \rightarrow \infty$ за $|x_1| \rightarrow \infty$, існують похідні $\partial_{x_1}^{k_1} g$, $0 < |k_1| \leq 2$, для яких справджуються нерівності

$$|\partial_{x_1}^{k_1} g(x_1)| \leq C \eta (D(x_1))^{|k_1|}, \quad |\Delta_{x_1}^{y_1} \partial_{x_1}^{k_1} g(x_1)| \leq C \eta |x_1 - y_1|^\lambda \left((D(x_1))^{|k_1|} + (D(y_1))^{|k_1|} \right), \quad \{x_1, y_1\} \subset R^{n_1}, \quad 0 < |k_1| \leq 2,$$

де $C > 0$, $\lambda \in (0,1]$ з умови B_2 , η – досить мале додатне число, вибором якого в конкретній ситуації будемо розпоряджатись.

Наведемо основний результат про фундаментальний розв'язок задачі Коші для рівняння (1).

Використовуватимемо ще такі позначення: $M := n_1 m_1 + n_2 m_2$, де $m_1 := 1/2$, $m_2 := 3/2$; $x_1 := (x'_1, x''_1)$, де $x'_1 := (x_{11}, \dots, x_{1n_2})$, $x''_1 := (x_{1(n_2+1)}, \dots, x_{1n_1})$; $X_1(t) = x_1$, $X_2(t) = x_2 + tx'_1$;

$$E_c(t, x, \xi) := \exp \{ -c(|X_1(t) - \xi_1|^2 t^{-1} + |X_2(t) - \xi_2|^2 t^{-3}) \} = E_c^1(t, x_1, \xi_1) E_c^2(t, X_2(t), \xi_2);$$

$M_k := |k_1| m_1 + |k_2| m_2$, де мультиіндекс $k = (k_1, k_2) \in Z_+^n$, а якщо $k_2 = 0$, то $M_k := |k_1| m_1$.

Теорема. Нехай для рівняння (1) виконуються умови $B_1 - B_3$. Тоді для нього існує фундаментальний розв'язок задачі Коші $G(t, x; \tau, \xi)$, $0 \leq \tau < t \leq T$, $\{x, \xi\} \subset R^n$, для якого справджуються оцінки

$$|\partial_{x_1}^{k_1} G(t, x; \tau, \xi)| \leq C C_l \left((t - \tau)^{-M - M_{k_1}} \exp \{ -c(t - \tau)(D(\xi_1))^2 \} + (D(\xi_1))^{-l} \right) E_c(t - \tau, x, \xi), \quad |k_1| < 2, \quad (2)$$

$$|\partial_{x_1}^{k_1} G(t, x; \tau, \xi)| \leq C_l \left((t - \tau)^{-M - M_{k_1}} \exp \{ -c(t - \tau)(D(\xi_1))^2 \} + (D(\xi_1))^{-l} \right) E_c(t - \tau, x, \xi) \times \\ \times \left(1 + (D(x_1))^2 + \exp \{ \varepsilon (t - \tau)(D(x_1))^2 \} \right), \quad |k_1| = 2, \quad (3)$$

$$0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n,$$

де l – довільно фіксоване додатне число $C_l > 0$, $c > 0$.

Крім того, правильними є оцінки

$$|\partial_{x_1}^{k_1} G(t, x; \tau, \xi)| \leq C \sum_{j=0}^{|k_1|} (t - \tau)^{-M - M_{k_1} + j m_1} (D(x_1))^j E_c(t - \tau, x, \xi) \exp \{ g(x_1) - g(\xi_1) \}, \quad |k_1| < 2,$$

$$|\partial_{x_1}^{k_1} G(t, x; \tau, \xi)| \leq C_l \sum_{j=0}^{|k_1|} (t - \tau)^{-M - M_{k_1} + j m_1} (D(x_1))^j \left(1 + (D(x_1))^2 + \exp \{ \varepsilon (t - \tau)(D(x_1))^2 \} \right), \quad |k_1| = 2,$$

$$0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n,$$

де $C > 0$, $c > 0$, g – будь-яка функція, яка задовольняє умову B_4 .

Фундаментальний розв'язок задачі Коші шукаємо методом Леві, тому опишемо спочатку головний член формули для фундаментального розв'язку задачі Коші.

2. Головний член формули для фундаментального розв'язку задачі Коші

Розглянемо допоміжне рівняння

$$\left(\partial_t - \sum_{j=1}^{n_2} x_{1j} \partial_{x_{2j}} - \sum_{j,s=1}^{n_1} b_{js}(t, y_1) \partial_{x_{1j}} \partial_{x_{1s}} - \sum_{j=1}^{n_1} b_j(t, y_1) \partial_{x_{1j}} (-i \partial_{x_{n+1}}) - b_0(t, y_1) (-i \partial_{x_{n+1}})^2 \right) v(t, x, x_{n+1}) = 0,$$

$$(t, x, x_{n+1}) \in \Pi_{(0,T]} := (0, T] \times R,$$

де y_1 – фіксована точка простору R^{n_1} . Шукаючи розв'язок задачі Коші для цього рівняння у вигляді

$$v(t, x, x_{n+1}) = F_{\sigma \rightarrow x, \eta \rightarrow x_{n+1}}^{-1} [\tilde{v}(t, \sigma, \eta)], \quad t > 0, \quad (x, x_{n+1}) \in R^{n+1},$$

і, використавши властивості оберненого перетворення Фур'є, одержимо для невідомої функції $\tilde{v}$ задачу Коші для лінійного неоднорідного рівняння із частинними похідними першого порядку, яку розв'язуємо методом характеристик. У підсумку приходимо до формули

$$v(t, x, x_{n+1}) = \int_{R^{n+1}} G_0(t, x, \tau, \xi, x_{n+1} - \xi_{n+1}, y_1) \varphi(\xi) d\xi d\xi_{n+1}, \quad t > 0, \quad (x, x_{n+1}) \in R^{n+1},$$

у якій

$$G_0(t, x, \tau, \xi, x_{n+1} - \xi_{n+1}, y_1) = F_{\mu \rightarrow x(t-\tau)-\xi}^{-1} v \rightarrow x_{n+1}-\xi_{n+1} [V(t, \tau, \mu, v, y_1)](t, \tau, x, \xi, x_{n+1} - \xi_{n+1}, y_1), \quad (5)$$

а функція $V(t, \tau, \mu, v, y_1)$ є розв'язком задачі

$$\begin{aligned} & (\partial_t - \sum_{j,s=1}^{n_1} b_{js}(t, y_1) ([\mu_{1j} + (\theta - \tau)\mu_{2j}] \zeta'_j + \mu_{1j} \zeta''_j) ([\mu_{1s} + (\theta - \tau)\mu_{2s}] \zeta'_s + \mu_{1s} \zeta''_s) - \\ & - i \sum_{j=1}^{n_1} b_j(t, y_1) ([\mu_{1j} + (\theta - \tau)\mu_{2j}] \zeta'_j + \mu_{1j} \zeta''_j) v - b_0(t, y_1) v^2) V(t, \tau, \mu, v, y_1) = 0, \quad V|_{t=\tau} = 1, \end{aligned} \quad (6)$$

де $\zeta'_j := \{1, j \in \{1, \dots, n_2\}, 0, j \in \{n_2 + 1, \dots, n_1\}, \zeta''_j := \{0, j \in \{1, \dots, n_2\}, 1, j \in \{n_2 + 1, \dots, n_1\}\}$.

Як для залежних від t в (Eidelman, Ivasyshen, & Kochubei, 2004) коефіцієнтів, одержуємо оцінку

$$|V(t, \tau, \sigma + i\gamma, \eta, y_1)| \leq C \exp\{(-\delta_1 |\sigma_1|^2 + c_1 |\gamma_1|^2)(t - \tau) + (-\delta_1 |\sigma_2|^2 + c_1 |\gamma_2|^2)(t - \tau)^3 - \delta |\eta|^2(t - \tau)\},$$

$$0 \leq \tau < t \leq T, \quad \{\sigma, \gamma\} \subset R^n, \quad \eta \in R, \quad y_1 \in R^{n_1},$$

у якій $C > 0, \delta_1 > 0, c_1 > 0, \delta > 0$, а на підставі цієї оцінки, (5) й умови B_1 одержуємо

$$\begin{aligned} & |\partial_x^k \partial_{x_{n+1}}^{k_{n+1}} G_0(t, x, \tau, \xi, x_{n+1} - \xi_{n+1}, y_1)| \leq C_{kk_{n+1}} (t - \tau)^{-M-M_k-k_{n+1}/2} E_c(t - \tau, x, \xi) \times \\ & \times \{-c|x_{n+1} - \xi_{n+1}|^2(t - \tau)^{-1}\}, \quad 0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n, \quad \{x_{n+1}, \xi_{n+1}\} \subset R, \quad y_1 \in R^{n_1}, \end{aligned}$$

де $k \in Z_+^n, k_{n+1} \in Z_+, C_{kk_{n+1}} > 0, c > 0$ – деякі сталі. Нехай $\hat{G}_0(t, x; \tau, \xi; v; y_1) = F_{z \rightarrow v}[G_0(t, x; \tau, \xi; z; y_1)]$ і

$\hat{G}(t, x; \tau, \xi; y_1) = \hat{G}_0(t, x; \tau, \xi; D(y_1); y_1)$. Використовуючи властивості перетворення Фур'є, одержуємо

$$\begin{aligned} & |\partial_x^k \hat{G}(t, x; \tau, \xi; y_1)| \leq C_k (t - \tau)^{-M-M_k} E_c(t - \tau, x, \xi) \exp\{-c(t - \tau)(D(y_1))^2\}, \\ & 0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n, \quad y_1 \in R^{n_1}, \end{aligned} \quad (7)$$

причому $\hat{G}(t, x; \tau, \xi; y_1) = F_{\sigma \rightarrow x(t-\tau)-\xi}^{-1} [V_0(t, \tau, \sigma, y_1)](t, \tau, x, \xi, y_1)$ є фундаментальним розв'язком задачі Коші рівняння

$$L(t, y_1, \partial_t, \partial_x) v(t, x) = 0, \quad (t, x) \in \Pi_{(0,T]}, \quad (8)$$

а $V_0(t, \tau, \sigma, y_1) = V(t, \tau, \sigma, D(y_1), y_1)$ згідно з (6) є розв'язком рівняння

$$\begin{aligned} & (\partial_t - \sum_{j,s=1}^{n_1} a_{js}(t, y_1) ([\sigma_{1j} + (t - \tau)\sigma_{2j}] \zeta'_j + \sigma_{1j} \zeta''_j) ([\sigma_{1s} + (t - \tau)\sigma_{2s}] \zeta'_s + \sigma_{1s} \zeta''_s) - \\ & - i \sum_{j=1}^{n_1} a_j(t, y_1) ([\sigma_{1j} + (t - \tau)\sigma_{2j}] \zeta'_j + \sigma_{1j} \zeta''_j) - a_0(t, y_1)) V_0(t, \tau, \sigma, y_1) = 0, \end{aligned} \quad (9)$$

де $\zeta'_j := \begin{cases} 1, j \in \{1, \dots, n_2\}, \\ 0, j \in \{n_2 + 1, \dots, n_1\}; \end{cases} \zeta''_j := \begin{cases} 0, j \in \{1, \dots, n_2\}, \\ 1, j \in \{n_2 + 1, \dots, n_1\}. \end{cases}$ При цьому

$$|V_0(t, \tau, \sigma + i\gamma, y_1)| \leq C \exp\{(-\delta_1 |\sigma_1|^2 + c_1 |\gamma_1|^2)(t - \tau) + (-\delta_1 |\sigma_2|^2 + c_1 |\gamma_2|^2)(t - \tau)^3 - \delta (D(y_1))^2(t - \tau)\},$$

$$0 \leq \tau < t \leq T, \quad \{\sigma, \gamma\} \subset R^n, \quad \eta \in R, \quad y_1 \in R^{n_1}, \quad (10)$$

де $C > 0, \delta_1 > 0, c_1 > 0, \delta > 0$.

Оцінка (10) лежить в основі отримання повного аналітичного опису $\hat{G}$. Наведемо деякі властивості фундаментального розв'язку рівняння (8).

Твердження 1. Нехай коефіцієнти рівняння (8) задовольняють умови B_1, B_2 . Тоді справджуються оцінки

$$\begin{aligned} & \Delta_{y_1}^{z_1} \partial_x^k \hat{G}(t, x; \tau, \xi; y_1) \leq C |y_1 - z_1|^\lambda (t - \tau)^{-M-M_k} (\exp\{-c(t - \tau)(D(y_1))^2\} + \exp\{-c(t - \tau)(D(z_1))^2\}) \times \\ & \times ((D(y_1))^2 + (D(z_1))^2) E_c(t - \tau, x, \xi), \quad 0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n, \quad \{y_1, z_1\} \subset R^{n_1}, \quad k \in Z_+^n. \end{aligned} \quad (11)$$

Доведення твердження 1. На підставі зазначеного, для рівняння для отримання оцінки (11) досить одержати оцінку різниці $\Delta_{y_1}^{z_1} V_0(t, \tau, \varrho, y_1)$, $0 \leq \tau < t \leq T, \{y_1, z_1\} \subset R^{n_1}, \varrho \in C^n$. На підставі (9) запишемо

$$\begin{aligned} & \Delta_{y_1}^{z_1} V_0(t, \tau, \varrho, y_1) = \int_\tau^t V_0(t, \theta, \varrho, y_1) \left(\sum_{j,s=1}^{n_1} \Delta_{y_1}^{z_1} a_{js}(\theta, y_1) ([\sigma_{1j} + (\theta - \tau)\sigma_{2j}] \zeta'_j + \sigma_{1j} \zeta''_j) ([\sigma_{1s} + (\theta - \tau)\sigma_{2s}] \zeta'_s + \right. \\ & \left. + \sigma_{1s} \zeta''_s) - i \sum_{j=1}^{n_1} a_j(\theta, y_1) ([\sigma_{1j} + (\theta - \tau)\sigma_{2j}] \zeta'_j + \sigma_{1j} \zeta''_j) - a_0(\theta, y_1) \right) V_0(\theta, \tau, \sigma, y_1) d\theta = 0. \end{aligned}$$

Скориставшись умовою B_2 й оцінкою (10), матимемо

$$\begin{aligned} & \Delta_{y_1}^{z_1} V_0(t, \tau, \varrho, y_1) \leq C |y_1 - z_1|^\lambda \exp\{(-\delta_1 |\sigma_1|^2 + c_1 |\gamma_1|^2)(t - \tau) + (-\delta_2 |\sigma_2|^2 + c_1 |\gamma_2|^2)(t - \tau)^3\} \times \\ & \times ((D(y_1))^2 + (D(z_1))^2) \int_\tau^t \{-\delta (D(y_1))^2(t - \theta) - \delta (D(z_1))^2(\theta - \tau)\} d\theta, \\ & 0 \leq \tau < t \leq T, \quad \{y_1, z_1\} \subset R^{n_1}, \quad \varrho = \sigma + i\gamma \in C^n, \quad c_1 > 0, \quad \delta_1 > 0, \quad \delta_2 = \delta_1/4. \end{aligned}$$

Оскільки

$$\int_\tau^t \{-\delta (D(y_1))^2(t - \theta) - \delta (D(z_1))^2(\theta - \tau)\} d\theta \leq \frac{t-\tau}{2} \left(\exp\left\{-\frac{\delta}{2}(t - \tau)(D(y_1))^2\right\} + \exp\left\{-\frac{\delta}{2}(t - \tau)(D(z_1))^2\right\} \right),$$

то

$$\Delta_{y_1}^{z_1} V_0(t, \tau, \varrho, y_1) \leq C |y_1 - z_1|^\lambda \exp\{(-\delta_1 |\sigma_1|^2 + c_1 |\gamma_1|^2)(t - \tau) + (-\delta_2 |\sigma_2|^2 + c_1 |\gamma_2|^2)(t - \tau)^3\} \times$$

$$\times \left(\exp \left\{ -\frac{\delta}{2} (t - \tau) (D(y_1))^2 \right\} + \exp \left\{ -\frac{\delta}{2} (t - \tau) (D(z_1))^2 \right\} \right) \left((D(y_1))^2 + (D(z_1))^2 \right).$$

$$0 \leq \tau < t \leq T, \quad \{y_1, z_1\} \subset R^{n_1}, \quad q = \sigma + i\gamma \in C^n$$

Твердження 2. Правильна рівність

$$\int_{R^n} \hat{G}(t, x; \tau, \xi; y_1) d\xi = \left\{ \int_{\tau}^t a_0(\theta, y_1) d\theta \right\}, \quad 0 \leq \tau < t \leq T, \quad x \in R^n, \quad y_1 \in R^{n_1}.$$

Доведення твердження 2. Потрібну формулу отримуємо з

$$\int_{R^n} \hat{G}(t, x; \tau, \xi; y_1) d\xi = \int_{R^n} F_{\sigma \rightarrow X(t-\tau)-\xi}^{-1} [V_0(t, \tau, \sigma, y_1)](t, \tau, x, \xi, y_1) d\xi$$

після заміни змінних інтегрування ξ за формулою $X(t - \tau) - \xi = \eta$.

З твердження 2 випливає, що для довільних $k \in Z_+^n$, $0 \leq \tau < t \leq T$, $x \in R^n$ і $y_1 \in R^{n_1}$ правильна рівність

$$\partial_x^k \int_{R^n} \hat{G}(t, x; \tau, \xi; y_1) d\xi = 0, \quad 0 \leq \tau < t \leq T, \quad x \in R^n, \quad y_1 \in R^{n_1}. \quad (12)$$

Поряд із функцією E_c використовуватимемо ще оцінну функцію

$$F_c(t, x, \xi) := \exp \left\{ -c \left(|x_1 - \xi_1|^2 (4t)^{-1} + 3|x_2 + (2^{-1}t(x'_1 + \xi'_1) - \xi_2)|^2 t^{-3} \right) \right\}, \quad t > 0, \quad \{x, \xi\} \subset R^n,$$

де $c > 0$, яка з точністю до сталого множника визначає фундаментальний розв'язок задачі Коші для модельного рівняння (1) зі сталими коефіцієнтами. Оцінні функції E_c та F_c згідно з (Івасишен, & Мединський, 2014) мають такі властивості :

$$E_c(t, x, \xi) \leq F_{c_1}(t, x, \xi) \leq E_{c_2}(t, x, \xi), \quad t > 0, \quad \{x, \xi\} \subset R^n, \quad 0 < c_2 < c_1 < c, \quad (13)$$

$$\int_{R^n} E_c(t, x, \xi) t^{-M} d\xi = C, \quad (14)$$

$$\forall \delta \in (0, 1) \exists C_\delta > 0 \quad \forall \{t, \theta, \tau\} \subset R, 0 \leq \tau < \theta < t \leq T \quad \forall \{x, \xi\} \subset R^n:$$

$$\int_{R^n} E_c(t - \theta, x, y) E_c(\theta - \tau, y, \xi) ((t - \theta)(\theta - \tau))^{-M} dy \leq C_\delta (t - \tau)^{-M} E_{c(1-\delta)/4}(t - \tau, x, \xi), \quad (15)$$

$$\int_{R^n} F_c(t - \theta, x, y) F_c(\theta - \tau, y, \xi) ((t - \theta)(\theta - \tau))^{-M} dy = C (t - \tau)^{-M} E_c(t - \tau, x, \xi), \quad c > 0. \quad (16)$$

Зазначимо, що

$$E_c(t - \tau, X(t - \theta), \xi) \leq E_c(t - \tau, x, \xi), \quad 0 \leq \tau < \theta < t, \quad \{x, \xi\} \subset R^n, \quad (17)$$

$$E_c(\theta - \tau, (y_1, X_2(t - \theta)), \xi) \leq E_{c/2}(t - \tau, x, \xi) E_{-2c}(\theta - \tau, y_1 - x_1), \quad 0 \leq \tau < \theta < t, \quad \{x, \xi\} \subset R^n, \quad (18)$$

а за $\theta \in [(t + \tau)/2, t]$ одержуємо

$$E_c(\theta - \tau, (y_1, X_2(t - \theta)), \xi) \leq E_{c/2}(t - \tau, x, \xi) E_{-2c}(t - \theta, y_1 - x_1). \quad (19)$$

Далі вистовуватимемо прирости $\Delta_{x_p}^{z(p)} f(\cdot, x, \cdot) := \Delta_x^{z(p)} f(\cdot, x, \cdot)$, $p \in \{1, 2\}$, де $z^{(1)} := (z_1, x_2)$, $z^{(2)} := (x_1, z_2)$.

Твердження 3. Нехай коефіцієнти рівняння (8) задовольняють умови $B_1 - B_3$, а неперервна функція $\varphi(t, x; \tau, \xi)$, $0 \leq \tau < t \leq T$, $\{x, \xi\} \subset R^n$ задовольняє умови

$$|\varphi(t, x; \tau, \xi)| \leq C_l \left((t - \tau)^{-M-1+\lambda/2} \exp \left\{ -c(t - \tau) (D(\xi_1))^2 \right\} + (t - \tau)^{\lambda/2} (D(\xi_1))^{-l} \right) E_c(t - \tau, x, \xi), \quad (20)$$

$$\begin{aligned} \left| \Delta_{x_p}^{z(p)} \varphi(t, x; \tau, \xi) \right| &\leq C_l |x_p - z_p|^{\lambda_1^p} \left((t - \tau)^{-M-1+\lambda_1^p m_p} \exp \left\{ -c(t - \tau) (D(\xi_1))^2 \right\} + (t - \tau)^{\lambda_2^p m_p} (D(\xi_1))^{-l} \right) \times \\ &\times (F_c(t - \tau, x, \xi) + F_c(t - \tau, z^{(p)}, \xi)), \quad 0 < \lambda_1^1 < \lambda, \quad \lambda_2^1 \equiv \lambda - \lambda_1^1, \quad 0 < \lambda_1^2 < \frac{\lambda}{3}, \quad \lambda_2^2 \equiv \frac{\lambda}{3} - \lambda_1^2, \quad p \in \{1, 2\}, \end{aligned} \quad (21)$$

де l – довільно фіксоване додатне число. Тоді функція

$$w(t, x; \tau, \xi) = \int_{\tau}^t d\theta \int_{R^n} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy, \quad 0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n,$$

має неперервні похідні $\partial_{x_1}^{k_1} w$, $|k_1| \leq 2$, і Sw , для яких правильні формули

$$\partial_{x_{1j}} w(t, x; \tau, \xi) = \int_{\tau}^t d\theta \int_{R^n} \partial_{x_{1j}} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy,$$

$$\partial_{x_{1j}} \partial_{x_{1l}} w(t, x; \tau, \xi) = \int_{\tau}^{t_1} d\theta \int_{R^n} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy +$$

$$+ \int_{t_1}^t d\theta \int_{R^n} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; y_1) \Delta_y^{X(t-\theta)} \varphi(\theta, y; \tau, \xi) +$$

$$+ \int_{t_1}^t \left(\int_{R^n} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; y_1) dy \right) \varphi(\theta, X(t - \theta); \tau, \xi) d\theta,$$

$$Sw(t, x; \tau, \xi) = \varphi(t, x; \tau, \xi) + \int_{\tau}^{t_1} d\theta \int_{R^n} S\hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy +$$

$$+ \int_{t_1}^t d\theta \int_{R^n} S\hat{G}(t, x; \theta, y; y_1) \Delta_y^{X(t-\theta)} \varphi(\theta, y; \tau, \xi) dy + \int_{t_1}^t \left(\int_{R^n} S\hat{G}(t, x; \theta, y; y_1) dy \right) \varphi(\theta, X(t - \theta); \tau, \xi) d\theta,$$

$$0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n, \quad t_1 := (t + \tau)/2.$$

Доведення твердження 3. Для $0 \leq \tau < t \leq T$, $\{x, \xi\} \subset R^n$ та $|k_1| \leq 2$ запишемо

$$\partial_{x_1}^{k_1} \int_{R^n} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy = \int_{R^n} \partial_{x_1}^{k_1} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy \equiv J^{(k_1)}.$$

Інтеграл $J^{(k_1)}$ збігається рівномірно щодо $x_1 \in R^n$ за фіксованих t, τ, θ і ξ . За $|k_1| < 2$ це випливає з оцінок (7), (20) та збіжності інтегралів $\int_{\tau}^t (t-\theta)^{-|k_1|m_1} (\theta-\tau)^{-1+\lambda/2} d\theta$ і $\int_{\tau}^t (t-\theta)^{-|k_1|m_1} (\theta-\tau)^{M+\lambda/2} d\theta$.

Для похідних другого порядку за x_1 і $\theta \in [t_1, t]$ така оцінка не придатна. Її можна поліпшити для $x_1 \in B(0, R/2)$, де R – будь-яке фіксоване додатне число, а $B(y_1, R) := \{x_1 \in R^{n_1}, |x_1 - y_1| \leq R\}$. Запишемо

$$J^{(k_1)} = \int_{R^n} \partial_{x_{1j}} \partial_{x_{1j}} \hat{G}(t, x; \theta, y; y_1) \Delta_y^{X(t-\theta)} \varphi(\theta, y; \tau, \xi) dy + \left(\int_{R^n} \Delta_{\zeta=y_1}^{\zeta=x_1} \partial_{x_{1j}} \partial_{x_{1j}} \hat{G}(t, x; \theta, y; \zeta) dy \right) \varphi(\theta, X(t-\theta); \tau, \xi) \equiv \\ \equiv I_1 + I_2 \varphi(\theta, X(t-\theta); \tau, \xi).$$

Використовуючи $|\Delta_y^{X(t-\theta)} \varphi(\theta, y; \tau, \xi)| \leq |\Delta_y^{(y_1, X_2(t-\theta))} \varphi(\theta, y; \tau, \xi)| + |\Delta_{(y_1, X_2(t-\theta))}^{X(t-\theta)} \varphi(\theta, y; \tau, \xi)|$, (7), (14), (15), (18), (19) із $c = c_0/3$, (21) і те, що $D \geq 1$, одержуємо

$$|I_1| \leq C(t-\tau)^{-M} E_{c_2}(t-\tau, x, \xi) ((\theta-\tau)^{-1+\lambda_1^2 m_1} + (\theta-\tau)^{-1+\lambda_2^2 m_2}),$$

де $C_1 > C$, $0 < c_1 < c_0 < c$, $c_2 = \min\{c_1, c_0/6\}$. За допомогою (11), (14), (15) матимемо

$$|I_2| \leq C(t-\theta)^{\lambda m_1-1} ((D(x_1))^2 + \int_{R^{n_2}} dy_2 \int_{B(x_1, 1)} (t-\theta)^{-M} E_{c_0}(t-\theta, x, y) (D(y_1))^2 dy_1 + \\ + \int_{R^{n_2}} dy_2 \int_{R^{n_1} \setminus B(x_1, 1)} (t-\theta)^{-M} E_{c_0}(t-\theta, x, y) (D(y_1))^2 dy_1).$$

Згідно з умовою B_3 $D(y_1) \leq C D(x_1)$ для $y_1 \in B(x_1, 1)$, а для $y_1 \in R^{n_1} \setminus B(x_1, 1)$ за допомогою нерівності Юнга $(D(y_1))^2 \leq C E_{-\varepsilon}^1(t-\theta, x_1 - y_1) \exp\{\varepsilon(t-\theta)(D(x_1))^2\}$, то, обравши ε , щоб $c_1 \equiv c_0 - \varepsilon > 0$, та використавши (14), одержимо

$$|I_2| \leq C(t-\theta)^{-1+\lambda/2} ((D(x_1))^2 + \exp\{\varepsilon(t-\theta)(D(x_1))^2\}).$$

Отже, враховуючи ще (17), за $|k_1| = 2$

$$|J^{(k_1)}| \leq C(t-\tau)^{-M} E_{c_2}(t-\tau, x, \xi) ((t-\tau)^{-1+\lambda_1^2 m_1} + (t-\tau)^{-1+\lambda_2^2 m_2}) + \\ + C C_1 (t-\theta)^{-1+\lambda/2} ((D(x_1))^2 + \exp\{\varepsilon(t-\tau)(D(x_1))^2\}) ((t-\tau)^{-M-1+\lambda/2} + (t-\tau)^{\lambda/2}) E_c(t-\tau, x, \xi), \\ 0 \leq \tau < t_1 < \theta < t \leq T, \{x, \xi\} \subset R^n.$$

Для доведення правильності формул для S_w оцінимо інтеграл $I^0 \equiv \int_{R^n} S \hat{G}(t, x; \theta, y; y_1) dy$. Щоб оцінити I^0 , скористаємося рівністю

$$S \hat{G}(t, x; \theta, y; y_1) = \left(\sum_{j,s=1}^{n_1} a_{js}(t, y_1) \partial_{x_{1j}} \partial_{x_{1s}} + \sum_{j=1}^{n_1} a_j(t, y_1) \partial_{x_{1j}} + a_0(t, x_1) \right) \hat{G}(t, x; \theta, y; y_1),$$

$$0 \leq \theta < t \leq T, \{x, \xi\} \subset R^n, y_1 \in R^{n_1},$$

рівностями (12), оцінками (11) і (19) й умовами на коефіцієнтами рівняння. Маємо

$$|I^0| \leq \left| \int_{R^n} \left(\sum_{j,s=1}^{n_1} a_{js}(t, y_1) \Delta_{z_1=y_1}^{z_1=x_1} \partial_{x_{1j}} \partial_{x_{1s}} + \sum_{j=1}^{n_1} a_j(t, y_1) \Delta_{z_1=y_1}^{z_1=x_1} \partial_{x_{1j}} \right) \hat{G}(t, x; \theta, y; z_1) dy \right| + \\ + \left| \int_{R^n} \left(\sum_{j,s=1}^{n_1} \Delta_{y_1}^{x_1} a_{js}(t, y_1) \partial_{x_{1j}} \partial_{x_{1s}} + \sum_{j=1}^{n_1} \Delta_{y_1}^{x_1} a_j(t, y_1) \partial_{x_{1j}} \right) \hat{G}(t, x; \theta, y; z_1) |_{z_1=y_1} dy \right| + \\ + \left| \int_{R^n} a_0(t, y_1) \hat{G}(t, x; \theta, y; y_1) dy \right| \leq \\ \leq C(t-\theta)^{-1+\lambda/2} \int_{R^n} (t-\theta)^{-M} E_{c_1}(t-\theta, x, y) ((D(y_1))^3 + D(y_1)(D(x_1))^2) dy.$$

Останній інтеграл оцінюється так, як відповідний інтеграл з I_2 . У підсумку одержуємо таку оцінку:

$$|I^0| \leq C(t-\theta)^{-1+\lambda/2} ((D(x_1))^3 + \exp\{3\varepsilon/2(t-\theta)(D(x_1))^2\}).$$

□

3. Доведення основної теореми

Згідно з методом Леві фундаментальний розв'язок задачі Коші для рівняння (1) шукаємо у вигляді

$$G(t, x; \tau, \xi) = \hat{G}(t, x; \tau, \xi; \xi_1) + \int_{\tau}^t d\theta \int_{R^n} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy, \quad 0 \leq \tau < t \leq T, \{x, \xi\} \subset R^n, \quad (22)$$

де $\varphi(\cdot, \cdot; \tau, \xi): \Pi_{[0, T]} \rightarrow C$ – невідома функція, яку підберемо так, щоб функція $G(\cdot, \cdot; \tau, \xi): \Pi_{[0, T]} \rightarrow C$ була розв'язком рівняння (1) для будь-якої фіксованої точки $(\tau, \xi) \in \Pi_{[0, T]}$.

Нехай шукана функція φ є неперервною і для неї справджуються оцінки (20) і (21). Застосувавши диференціальний вираз L з (1) до функції (22) та використавши припущення щодо φ , на підставі твердження 3 одержимо для φ інтегральне рівняння

$$\varphi(t, x; \tau, \xi) = K(t, x; \tau, \xi) + \int_{\tau}^t d\theta \int_{R^n} K(t, x; \theta, y) \varphi(\theta, y; \tau, \xi) dy, \quad 0 \leq \tau < t \leq T, \{x, \xi\} \subset R^n, \quad (23)$$

де

$$K(t, x; \tau, \xi) = \left(\sum_{j,s=1}^{n_1} \Delta_{x_1}^{\xi_1} a_{js}(t, x_1) \partial_{x_{1j}} \partial_{x_{1s}} + \sum_{j=1}^{n_1} \Delta_{x_1}^{\xi_1} a_j(t, x_1) \partial_{x_{1j}} + \Delta_{x_1}^{\xi_1} a_0(t, x_1) \right) \hat{G}(t, x; \tau, \xi; \xi_1).$$

Оцінимо ядро K . За допомогою оцінки (7), умови B_2 маємо

$$|K(t, x; \tau, \xi)| \leq C(t - \tau)^{-M+\lambda/2} E_{c_0}(t - \tau, x, \xi) \left\{ -c(t - \tau)(D(\xi_1))^2 \right\} \times \\ \times \left((n_1)^2 (t - \tau)^{-1} + n_1(D(x_1) + D(\xi_1))(t - \tau)^{-1/2} + ((D(x_1))^2 + (D(\xi_1))^2) \right), \quad 0 \leq \tau < t \leq T, \quad \{x, \xi\} \subset R^n,$$

де $C > 0$, $0 < c_0 < c$. Якщо $|x_1 - \xi_1| \leq 1$, то з цієї нерівності на підставі умови B_3 одержуємо

$$|K(t, x; \tau, \xi)| \leq C_0(t - \tau)^{-M+\lambda/2-1} E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c_0(t - \tau)(D(\xi_1))^2 \right\}.$$

Якщо $|x_1 - \xi_1| > 1$, то за допомогою умови B_3 , вибравши в ній ε так, щоб $c_1 \equiv c_0 - \varepsilon > 0$, одержимо

$$|K(t, x; \tau, \xi)| \leq C'_0(t - \tau)^{-M+\lambda/2-1} E_{c_1}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\}.$$

Отже,

$$|K(t, x; \tau, \xi)| \leq C'_0(t - \tau)^{-M+\lambda/2-1} E_{c_1}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\}. \quad (24)$$

На підставі оцінки (24) рівняння (23) згідно з результатами з роботи (Пасічник, 2024) розв'язується методом послідовних наближень і правильна оцінка (20).

Доведемо правильність оцінки (21). При $|x_1 - z_1|^{1/m_1} > (t - \tau)/2$ та $|x_2 - z_2|^{1/m_2} > (t - \tau)/2$ відповідно оцінки (21) випливають з (20). Тому досить розглянути випадок, коли $|x_p - z_p|^{1/m_p} \leq (t - \tau)/2$, $p \in \{1, 2\}$.

Скориставшись (7) та нерівністю $E_c(t - \tau, \chi^{(p)}, \xi) \leq C_1 E_{c_0}(t - \tau, x, \xi)$, $0 < c_0 < c$, де $\chi^{(p)}$ – точка, яка належить відріzkу, що сполучає точки x і $z^{(p)}$, $p \in \{1, 2\}$, з (Івасишен, & Мединський, 2014), одержуємо

$$\left| \Delta_{x_1}^{z_1} \partial_{x_{1j}} \partial_{x_{1s}} \hat{G}(t, x; \tau, \xi; \xi_1) \right| \leq \sum_{l=1}^{n_1} |x_{1l} - z_{1l}| \left| \partial_{x_{1l}} \partial_{x_{1j}} \partial_{x_{1s}} \hat{G}(t, x; \tau, \xi; \xi_1) \right| \leq C \sum_{l=1}^{n_1} |x_{1l} - z_{1l}| (t - \tau)^{-M-3/2} \times \\ \times \exp \left\{ -c(t - \tau)(D(\xi_1))^2 \right\} \leq C |x_1 - z_1|^{\alpha_1^0} (t - \tau)^{-M-1-\alpha_1^0 m_1} E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c(t - \tau)(D(\xi_1))^2 \right\}, \\ \left| \Delta_{x_1}^{z_1} \partial_{x_{1j}} \hat{G}(t, x; \tau, \xi; \xi_1) \right| \leq C |x_1 - z_1|^{\alpha_1^0} (t - \tau)^{-M-1/2-\alpha_1^0 m_1} E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c(t - \tau)(D(\xi_1))^2 \right\}, \\ \left| \Delta_{x_1}^{z_1} \hat{G}(t, x; \tau, \xi; \xi_1) \right| \leq C |x_1 - z_1|^{\alpha_1^0} (t - \tau)^{-M-\alpha_1^0 m_1} E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c(t - \tau)(D(\xi_1))^2 \right\},$$

де $\alpha_1^0 \in (0, 1)$, $0 < c_0 < c$. Використовуючи ці оцінки з $\alpha_1^0 = \lambda$ та оцінки (7), припущення щодо коефіцієнтів рівняння (1), нерівність $|z_1 - \xi_1|^\lambda \leq |x_1 - z_1|^\lambda + |x_1 - \xi_1|^\lambda$, прийдемо до

$$\left| \Delta_{x_1}^{z_1} K(t, x; \tau, \xi) \right| = \left| \left(\sum_{j,s=1}^{n_1} \Delta_{x_1}^{z_1} a_{js}(t, x_1) \partial_{x_{1j}} \partial_{x_{1s}} + \sum_{j=1}^{n_1} \Delta_{x_1}^{z_1} a_j(t, x_1) \partial_{x_{1j}} + \Delta_{x_1}^{z_1} a_0(t, x_1) \right) \hat{G}(t, x; \tau, \xi; \xi_1) \right| \leq \\ + \sum_{j=1}^{n_1} \left| \Delta_{x_1}^{z_1} a_j(t, x_1) \partial_{x_{1j}} \hat{G}(t, x; \tau, \xi; \xi_1) + \Delta_{x_1}^{z_1} a_0(t, x_1) \hat{G}(t, x; \tau, \xi; \xi_1) \right| \leq \\ \leq C |x_1 - z_1|^\lambda (t - \tau)^{-M-1} R(t, \tau, x, z^{(1)}, \xi),$$

де

$$R(t, \tau, x, z^{(1)}, \xi) = \left(2 + (D(x_1) + 2D(z_1) + D(\xi_1))(t - \tau)^{\frac{1}{2}} + ((D(x_1))^2 + 2(D(z_1))^2 + (D(\xi_1))^2)(t - \tau) \right) \times \\ \times E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c(t - \tau)(D(\xi_1))^2 \right\},$$

$C > 0$, $0 < c_0 < c$. Якщо $|x_1 - \xi_1| \leq 1$, $|z_1 - \xi_1| \leq 1$, то на підставі умови B_3 маємо

$$R(t, \tau, x, z^{(1)}, \xi) \leq C \left(1 + D(\xi_1)(t - \tau)^{1/2} + (D(\xi_1))^2(t - \tau) \right) E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c(t - \tau)(D(\xi_1))^2 \right\} \leq \\ \leq E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c_0(t - \tau)(D(\xi_1))^2 \right\}, \quad 0 < c_0 < c.$$

Якщо $|x_1 - \xi_1| > 1$, $|z_1 - \xi_1| > 1$, то за допомогою оцінок степенів $D(x_1)$ і $D(z_1)$, як в оцінці $|I_2|$, одержуємо

$$R(t, \tau, x, z^{(1)}, \xi) \leq E_{c_0}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\} (E_{-\varepsilon}^1(t - \tau, x_1 - \xi_1) + E_{-\varepsilon}^1(t - \tau, z_1 - x_1) E_{-\varepsilon}^1(t - \tau, x_1 - \xi_1) + 1),$$

де $c > 0$ таке, як і в оцінці K . Оскільки $|x_1 - z_1|^{1/m_1} \leq (t - \tau)/2$, то $E_{-\varepsilon}^1(t - \tau, z_1 - x_1) \leq C$. Тоді, вибравши ε , можливо, меншим, ніж у (24), і змінивши число c_1 , одержимо $R(t, \tau, x, z^{(1)}, \xi) \leq C E_{c_1}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\}$. Аналогічно розглядаємо випадки, коли $|x_1 - \xi_1| \leq 1$, $|z_1 - \xi_1| > 1$ та $|x_1 - \xi_1| > 1$, $|z_1 - \xi_1| \leq 1$.

У підсумку, за $|x_1 - z_1|^{1/m_1} \leq (t - \tau)/2$ одержуємо

$$\left| \Delta_{x_1}^{z_1} K(t, x; \tau, \xi) \right| \leq C |x_1 - z_1|^\lambda (t - \tau)^{-M-1} E_{c_1}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\}, \quad 0 \leq \tau < t \leq T, \quad \{x, \xi, z^{(1)}\} \subset R^n,$$

звідки, зокрема, за $|x_1 - z_1|^{1/m_1} \leq (t - \tau)/2$ випливає оцінка

$$\left| \Delta_{x_1}^{z_1} K(t, x; \tau, \xi) \right| \leq C |x_1 - z_1|^{\lambda_1^1} (t - \tau)^{-M-1+\lambda_1^1 m_1} E_{c_1}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\}, \quad \lambda_2^1 = \lambda - \lambda_1^1, \quad \lambda_1^1 \in (0, \lambda).$$

Аналогічно за $|x_2 - z_2|^{1/m_2} \leq (t - \tau)/2$ одержуємо

$$\left| \Delta_{x_2}^{z_2} K(t, x; \tau, \xi) \right| \leq C |x_2 - z_2|^{\lambda_2^1} (t - \tau)^{-M-1+\lambda_2^1 m_2} E_{c_1}(t - \tau, x, \xi) \exp \left\{ -c_1(t - \tau)(D(\xi_1))^2 \right\}, \quad \lambda_2^2 = \lambda/3 - \lambda_1^2, \quad \lambda_1^2 \in (0, \lambda/3).$$

Для отримання оцінки (21) за $|x_p - z_p|^{1/m_p} \leq (t - \tau)/2$ згідно із (23) запишемо

$$\Delta_{x_p}^{z_p} \varphi(t, x; \tau, \xi) = \Delta_{x_p}^{z_p} K(t, x; \tau, \xi) + \int_{\tau}^v d\theta \int_{R^n} \Delta_{x_p}^{z_p} K(t, x; \theta, y) \varphi(\theta, y; \tau, \xi) dy + \\ + \int_v^t d\theta \int_{R^n} K(t, x; \theta, y) \varphi(\theta, y; \tau, \xi) dy - \int_v^t d\theta \int_{R^n} K(t, z^{(p)}; \theta, y) \varphi(\theta, y; \tau, \xi) dy, \quad p \in \{1, 2\},$$

де $v = t - |x_p - z_p|^{1/m_p}$, і використовуємо (14), (15), нерівності

$$E_c(t - \theta, x, y) E_c(\theta - \tau, y, \xi) \leq E_{c/4}(t - \tau, x, \xi), \quad 0 \leq \tau < \theta < t \leq T, \quad \{x, \xi, y\} \subset R^n,$$

і

$$\int_v^t (t - \theta)^{-1+\lambda/2} (\theta - \tau)^{-1+\lambda/2} d\theta \leq \frac{2}{\lambda} \left(\frac{t - \tau}{2} \right)^{1+\lambda/2} (|x_p - z_p|^{1/m_p})^{\lambda/2}, \quad |x_p - z_p|^{1/m_p} \leq (t - \tau)/2, \quad p \in \{1, 2\}.$$

Доведемо тепер правильність для G оцінок (2) і (3). Для похідних від першого доданка із (22) справджуються оцінки (7). Тому оцінимо похідні від другого доданка. Розглянемо випадок $|k_1| = 2$. На підставі твердження 3 та (12) запишемо

$$\partial_{x_{1j}} \partial_{x_{1l}} w(t, x; \tau, \xi) = \int_{\tau}^{t_1} d\theta \int_{R^n} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; y_1) \varphi(\theta, y; \tau, \xi) dy + \int_{t_1}^t d\theta \int_{R^n} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; y_1) \times \\ \times \Delta_{y_1}^{x(t-\theta)} \varphi(\theta, y; \tau, \xi) dy + \int_{t_1}^t \left(\int_{R^n} \Delta_{z_1=y_1}^{z_1=x_1} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; z_1) dy \right) \varphi(\theta, X(t - \theta); \tau, \xi) d\theta = \\ = J_1 + J_2 + J_3. \quad (25)$$

За допомогою (7), (14), (15), (20) та $E_c(t - \theta, x, y) E_c(\theta - \tau, y, \xi) \leq E_{c/4}(t - \tau, x, \xi)$, $\tau < \theta < t$, $\{x, \xi, y\} \subset R^n$, маємо

$$|J_1| \leq C(t - \tau)^{-M} E_{c(1-\delta)}(t - \tau, x, \xi) \exp \{-c(t - \tau)(D(\xi_1))^2\} (t - t_1)^{-1} \int_{\tau}^{t_1} (\theta - \tau)^{-1+\lambda/2} d\theta + \\ + C(D(\xi_1))^{-l} E_{c(1-\delta)/4}(t - \tau, x, \xi) (t - t_1)^{-1} \left(\int_{\tau}^{t_1} d\theta + \int_{\tau}^{t_1} (\theta - \tau)^{\lambda/2} d\theta \right) \leq \\ \leq C \left((t - \tau)^{-M-1+\lambda/2} \exp \{-c(t - \tau)(D(\xi_1))^2\} + (D(\xi_1))^{-l} \right) E_{c_1}(t - \tau, x, \xi).$$

При оцінюванні доданка J_2 використовуємо (7), оцінки приростів $\Delta_{y_1}^{x(t-\theta)} \varphi$, потім (17) та (18) з $c = c_0/3$, а далі (14), (15) та умовою B_3 на характеристики дисипації. У підсумку отримуємо оцінку

$$|J_2| \leq C \int_{t_1}^t d\theta \int_{R^n} \left((\theta - \tau)^{-M-1+\lambda_1^2 m_1} \exp \{-c(t - \tau)(D(\xi_1))^2\} + (\theta - \tau)^{\lambda_1^2 m_1} (D(\xi_1))^{-l} \right) \times \\ \times (E_c(t - \tau, x, \xi) + E_{c_0/6}(t - \tau, x, \xi) E_{-2c_0/3}^1(t - \theta, y_1 - x_1)) (t - \theta)^{-M-1+\lambda_1^2 m_1} E_{c_0}(t - \theta, x, y) \exp \{-c(t - \theta)(D(y_1))^2\} dy + \\ + C \int_{t_1}^t d\theta \int_{R^n} (t - \theta)^{-M-1+\lambda_1^2 m_2} \left((\theta - \tau)^{-M-1+\lambda_2^2 m_2} \exp \{-c(\theta - \tau)(D(\xi_1))^2\} + (\theta - \tau)^{\lambda_2^2 m_2} (D(\xi_1))^{-l} \right) \times \\ \times (F_c(\theta - \tau, x, \xi) + E_{c_0/6}(t - \tau, x, \xi) E_{-2c_0/3}^1(t - \theta, y_1 - x_1)) E_{c_0}(t - \theta, x, y) \exp \{-c(t - \theta)(D(y_1))^2\} dy \leq \\ \leq C_1 \left((t - \tau)^{-M-1+\lambda/2} \exp \{-c_1(t - \tau)(D(\xi_1))^2\} + (D(\xi_1))^{-l} \right) E_{c_2}(t - \tau, x, \xi),$$

у якій $C_1 > C$, $0 < c_1 < c_0 < c$, $c_2 = \min \{c_1, c_0/6\}$. Аналогічно оцінюємо J_3 . Із (25) й оцінок $J_1 - J_3$ випливає (3).

Для одержання (4) вводимо нову функцію v за допомогою формули $u(t, x) = e^{g(x_1)} v(t, x)$, $(t, x) \in \Pi_{(0,T)}$, де функція g задовольняє умову B_4 . Водночас для v одержуємо рівняння, яке за досить малого η задовольняє умови $B_1 - B_3$. Тому для нього існує фундаментальний розв'язок задачі Коші $G_1(t, x; \tau, \xi)$, $0 \leq \tau < t \leq T$, $\{x, \xi\} \subset R^n$, для якого справджуються оцінки (2) і (3). Фундаментальні розв'язки задачі Коші G і G_1 зв'язані між собою рівністю

$$G(t, x; \tau, \xi) = \exp \{g(x_1) - g(\xi_1)\} G_1(t, x; \tau, \xi),$$

з якої безпосереднім диференціюванням і використанням оцінок (2) і (3) й умови B_4 одержуємо оцінки (4). □

Дискусія і висновки

Для похідних другого порядку оцінка (2) за такою методикою не отримується, оскільки оцінка інтеграла $\int_{R^n} \partial_{x_{1j}} \partial_{x_{1l}} \hat{G}(t, x; \theta, y; y_1) dy$ має інтегровну особливість за $t = \theta$.

Функція φ із (23) така, що існують похідні $\partial_{x_{2j}} w$ функції w із твердження 3 й існують відповідно похідні $\partial_{x_{2j}} G$ за змінною виродження рівняння. Доведення цього громіздке і здійснюється за методикою отримання відповідних оцінок за змінною x_1 з використанням властивостей оцінних функцій E_c .

Отримані результати можуть бути використані для дослідження розв'язності й інтегрального зображення розв'язків задачі Коші у спеціальних вагових просторах та при дослідженні фундаментального розв'язку задачі Коші для рівняння типу Колмогорова з коефіцієнтами, залежними і від змінних виродження, і систем таких рівнянь.

Джерела фінансування. Це дослідження не отримало жодного гранту від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

Івасишен, С. Д., & Мединський, І. П. (2014). Класичний фундаментальний розв'язок виродженого рівняння Колмогорова, коефіцієнти якого не залежать від змінних виродження. *Буковинський математичний журнал*, 2(2–3), 94–106.

- Івасишен, С. Д., & Пасічник, Г. С. (2014). Фундаментальний розв'язок задачі Коші для одного параболического рівняння зі зростаючими коефіцієнтами групи молодших членів. *Диференціальні рівняння і суміжні питання аналізу: збірник праць Інституту математики НАН України*, 11(2), 126–153.
- Пасічник, Г. С. (2024). Розв'язки деяких класів інтегральних рівнянь другого роду. *Буковинський математичний журнал*, 12(1), 84–93. <https://doi.org/10.31861/bmj2024.01.08>
- Процах, Н. П., & Пташник Б. Й. (2017). *Нелінійні ультрапараболическі рівняння та варіаційні нерівності*. Наукова думка.
- Eidelman, S. D., Ivasyshen, S. D., & Kochubei, A. N. (2004). *Analytic methods in the theory of differential and pseudo-differential equations of parabolic type*. Birkhäuser. <https://doi.org/10.1007/978-3-0348-7844-9>
- Di Francesco, M., & Pascucci, A. (2005). On a class of degenerate parabolic equations of Kolmogorov type. *Applied Mathematics Research eXpress*, 3, 77–116. <https://doi.org/10.1155/AMRX.2005.77>
- Ivasyshen, S. D., & Medyn's'kyi, I. P. (2018). On the classical fundamental solutions of the Cauchy problem for ultraparabolic Kolmogorov-type equations with two groups of spatial variables. *Journal of Mathematical Sciences*, 231(4), 507–526. <https://doi.org/10.1007/s10958-018-3830-0>

References

- Eidelman, S. D., Ivasyshen, S. D., & Kochubei, A. N. (2004). *Analytic methods in the theory of differential and pseudo-differential equations of parabolic type*. Birkhäuser. <https://doi.org/10.1007/978-3-0348-7844-9>
- Di Francesco, M., & Pascucci, A. (2005). On a class of degenerate parabolic equations of Kolmogorov type. *Applied Mathematics Research eXpress*, 3, 77–116. <https://doi.org/10.1155/AMRX.2005.77>
- Ivasyshen, S. D., & Medyn's'kyi, I. P. (2014). The classical fundamental solution of a degenerate Kolmogorov equation with coefficients independent on variables of degeneration. *Bukovinian Mathematical Journal*, 2(2–3), 94–106 [in Ukrainian].
- Ivasyshen, S. D., & Medyn's'kyi, I. P. (2018). On the classical fundamental solutions of the Cauchy problem for ultraparabolic Kolmogorov-type equations with two groups of spatial variables. *Journal of Mathematical Sciences*, 231(4), 507–526. <https://doi.org/10.1007/s10958-018-3830-0>
- Ivasyshen, S. D., & Pasichnyk, H. S. (2014). Fundamental solution of the Cauchy problem for parabolic equation with growing lowest coefficients. *Proceedings of Institute of Mathematics NAS of Ukraine*, 11(2), 126–153 [in Ukrainian].
- Pasichnyk, H. S. (2024). Solutions of some integral equations of the second kind. *Bukovinian Mathematical Journal*, 12(1), 84–93 [in Ukrainian]. <https://doi.org/10.31861/bmj2024.01.08>
- Protsach, N. P., & Ptashnyk, B. Yo. (2017). *Nonlinear ultraparabolic equations and variational inequalities*. Naukova dumka [in Ukrainian].

Отримано редакцією журналу / Received: 07.03.25

Прорецензовано / Revised: 07.08.25

Схвалено до друку / Accepted: 10.10.25

Halyna PASICHNYK, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0002-1759-5042

e-mail: h.pasichnyk@chnu.edu.ua

Yufiy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine

FUNDAMENTAL SOLUTION OF THE CAUCHY PROBLEM FOR THE DISSIPATIVE ULTRAPARABOLIC KOLMOGOROV EQUATION WITH COEFFICIENTS INDEPENDENT ON VARIABLE DEGENERATION

For the ultraparabolic Kolmogorov equation with coefficients that do not depend on the degeneracy variables and are unbounded in the spatial variable, a fundamental solution of the Cauchy problem is constructed. The growth of the coefficients of the equation depends on an increasing function, which is called the characteristic of dissipation. The coefficients of the equation satisfy the special Hölder condition, which contains the characteristic of dissipation. Differentiability of the coefficients is not required; therefore, an additional condition is imposed on the characteristic of dissipation. Ultraparabolic Kolmogorov equations with increasing coefficients arise in problems of the theory of random processes, statistical radio engineering.

The fundamental solution of the Cauchy problem is constructed by the Levy method. In this case, the parametrics are taken as the fundamental solution of the Cauchy problem of the equation under consideration with coefficients depending on the time and parametric spatial variables. This complicates the study, since in the case of bounded coefficients, the parametrics are taken as the fundamental solution of the Cauchy problem of the equation with frozen coefficients only of the group of higher derivatives.

The obtained estimates of the derivatives of the fundamental solution of the Cauchy problem can be used in studying the issues of solvability of the Cauchy problem and the integral representation of solutions in special weight spaces, as well as for constructing a classical fundamental solution of the Cauchy problem for a degenerate equation of the Kolmogorov type with coefficients increasing as $|x| \rightarrow \infty$, which also depend on the degeneracy variables, and systems of such equations.

Key words: ultraparabolic Kolmogorov equation, increasing coefficients, fundamental solution to the Cauchy problem, the Levy method.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 004.93

DOI: <https://doi.org/10.17721/1812-5409.2025/2.21>

Oleh SAMOILENKO, PhD Student

ORCID ID: 0000-0003-4466-4004

e-mail: oleh.samoilenko@imath.kiev.ua

Institute of Mathematics of the NAS of Ukraine, Kyiv, Ukraine

MATHEMATICAL PROBLEMS OF FEATURE MATCHING FOR VISION-GUIDED VEHICLES WITH LIMITED RESOURCES

Vision-based sensing plays a critical role for autonomous vehicles, where image data serve as the primary input for navigation, mapping, state estimation, and motion control. These operate under real-time and resource-constrained conditions, requiring feature detection and matching algorithms that are accurate and computationally efficient. A key component of such pipelines is the identification of keypoints – distinctive image locations $\mathbf{x} \in \mathbb{R}^2$ within a digital image function $I: \mathbb{R}^2 \rightarrow \mathbb{R}^c$, where c is the channels number, and $I(\mathbf{x}) \in \mathbb{R}^c$ denotes the local pixel intensity. Each detected keypoint is associated with a descriptor $\mathbf{d} \in \mathbb{R}^D$ encoding a local visual structure in a form invariant to translation, rotation, and moderate scale changes. Keypoints between two images $\mathbf{x}_i$ and $\mathbf{x}'_i$ are matched by comparing descriptors yielding the correspondences $\mathbf{x}'_i \leftrightarrow \mathbf{x}_i$. Geometric verification is performed by estimating a transformation matrix $\mathbf{H} \in \mathbb{R}^{3 \times 3}$ and accepting $\|\mathbf{x}'_i - \mathbf{H}\mathbf{x}_i\| < \varepsilon$, where ε is a matching tolerance. The proportion of such inlier correspondences referred to as the inlier ratio serves as an accuracy metric. Classical keypoint methods, e.g. the Scale-Invariant Feature Transform (SIFT) versus learned methods, e.g. the SuperPoint (applied to a manually constructed dataset of satellite imagery), are mathematically evaluated. Performance is analyzed in terms of inlier ratio and computational efficiency, reflecting the trade-offs between robustness and resource use. The results pursue design guidelines for integrating lightweight yet accurate vision algorithms into platforms such as UAVs, enabling reliable visual odometry and the Simultaneous Localization and Mapping (SLAM) under constrained hardware conditions.

Key words: *unmanned autonomous vehicles, image matching, feature detection, computer vision, pattern recognition.*

AMS 2020 classification: 68T45, 68U10, 65D19.

Introduction

Vision-based sensing has become a cornerstone of modern autonomous vehicles to perceive and navigate complex environments, where cameras provide rich environmental information for navigation, mapping, and control, with particular importance in the GPS-denied environments (Boyun et al., 2025). Then the feature detection methods can be divided into two categories: classical and learned ones. Classical techniques, e.g. the Scale-Invariant Feature Transform (SIFT) (Lowe, 2004), rely on hand-crafted heuristics such as image gradients and binary intensity tests. They are lightweight and facilitate real-time Simultaneous Localization and Mapping (SLAM) systems (Mur-Artal, Montiel, & Tardos, 2015) and visual odometry computation pipelines (Scaramuzza, & Fraundorfer, 2011). These are efficient on CPU-based embedded systems but could be sensitive to noise, illumination, and geometric distortions. The learned methods, e.g. the SuperPoint (DeTone, Malisiewicz, & Rabinovich, 2018), employ deep neural networks to jointly detect and describe keypoints, as well as learning robust features, directly from the data. Trained on diverse image collections, they achieve higher repeatability and robustness to viewpoint and illumination changes compared to classical handcrafted descriptors, and became widely adopted in remote sensing applications (Rusyn, Lutsyk, & Kosarevych, 2021). Combining with learned matchers like SuperGlue (Sarlin et al., 2020) yields the state-of-the-art results in image matching and relative pose estimation, consistently outperforming the SIFT. Large-scale benchmarks (Huang et al., 2024) consistently report improved reconstruction quality and localization accuracy for learned pipelines. Nevertheless, these improvements come at the expense of computational complexity. The SuperPoint involves a dense convolutional architecture with multiple encoder/decoder stages, requiring on the order of several gigaflops (GFLOPs) per forward pass and significant memory to store intermediate feature maps. This causes longer inference times and practical need for GPU or hardware acceleration to achieve the real-time performance.

We develop a mathematical framework for evaluating and comparing keypoint-based image matching strategies, formalize the matching pipeline and compare classical and learned methods in terms of inlier correspondences and runtime. To evaluate image matching techniques, we construct a custom dataset consisting of a satellite imagery collected by utilizing the Google Maps Static API (Google, 2025), capturing images tile-by-tile over a predefined geographic region. Each image is defined by the GPS coordinates of its center point and a fixed zoom level ensuring a consistent spatial resolution, scale and minimal distortion across the dataset. The sampling grid was designed to ensure a partial overlap between the neighbouring images enabling a robust pairwise matching evaluation.

1. Scale-Invariant Feature Transform (SIFT)

The SIFT is a hand-crafted pipeline for detecting and describing local features. It constructs the scale space representation $L(\mathbf{x}, \sigma)$ by convolving the image with a variable-scale Gaussian kernel $G(\mathbf{x}, \sigma)$,

$$L(\mathbf{x}, \sigma) = G(\mathbf{x}, \sigma) * I(\mathbf{x}),$$

where $\mathbf{x} = (x, y)^\top$ denotes coordinates of the input image I , $\sigma > 0$ is the scale parameter. To detect scale-invariant keypoints, the SIFT computes the Difference of Gaussians (DoG),

$$D(\mathbf{x}, \sigma) = L(\mathbf{x}, k\sigma) - L(\mathbf{x}, \sigma),$$

where $k > 1$ is the multiplicative factor between successive scales. Local extrema of $D(\mathbf{x}, \sigma)$ in both spatial and scale dimensions are selected as candidate keypoints. Each detected keypoint is assigned with a dominant orientation to achieve rotation invariance: $\theta(\mathbf{x}) = \arctan(L_y/L_x)$, where L_x and L_y denote the image derivatives computed from the smoothed image $L(\mathbf{x}, \sigma)$, to introduce a canonical direction for each keypoint based on the predominant edge direction in its local neighborhood so that the descriptor can be rotated relative to this orientation and remain consistent across rotated views.

The SIFT descriptor is formed by sampling gradient magnitudes and orientations in a region around the keypoint, partitioning this region into a grid of spatial cells, and computing orientation histograms in each cell. The bins number in each histogram determines the final descriptor dimension D , which depends on the cells number and the number of orientation bins per each cell. The resulting vector is normalized to the unit length to improve the robustness.

2. The SuperPoint

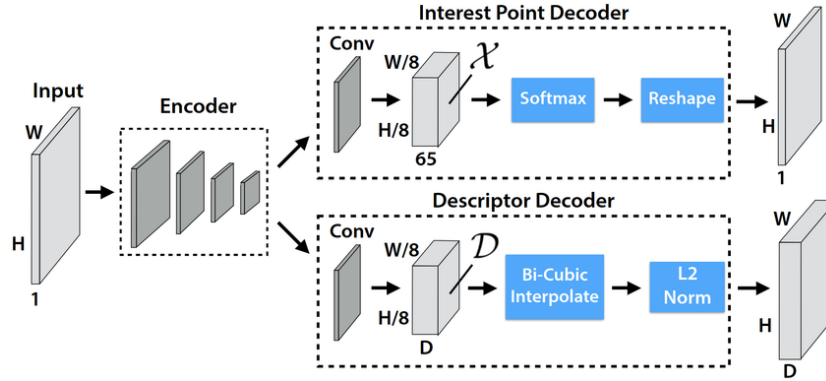


Fig. 1. The SuperPoint (DeTone, Malisiewicz, & Rabinovich, 2018) network architecture. The input image is processed by a shared convolutional encoder into a low-resolution feature map which is then split into two decoder “heads” for the interest point detection and descriptor extraction

The SuperPoint backbone is a fully-convolutional neural network consisting of convolutional layers with the ReLU activations $f(x) = \max(0, x)$ interleaved with three 2×2 max-pooling layers which reduce the spatial resolution by a factor of 8 to $H_c = H/8$ and $W_c = W/8$. The backbone feeds into two decoder heads. The “detector” head, or interest point decoder (Fig. 1, upper branch) outputs a score tensor $\mathbf{z} \in \mathbb{R}^{H_c \times W_c \times 65}$, where the first 64 channels represent positions within an 8×8 grid cell and the 65-th channel (“dustbin”) indicates the absence of a keypoint. A channel-wise softmax,

$$\mathbf{p}_i = \text{softmax}(\mathbf{z}_i), \quad p_{i,k} = \frac{e^{z_{i,k}}}{\sum_{u=1}^{65} e^{z_{i,u}}},$$

produces a probability vector $\mathbf{p}_i \in \mathbb{R}^{65}$ for each spatial location i . The detector is first trained on a synthetic dataset where ground-truth keypoints are defined at corners and intersections of primitive geometric figures (e.g., lines, circles, triangles, squares). Neural network training minimizes the spatial cross-entropy in the general definition (Goodfellow et al., 2016),

$$\mathcal{L}_{\text{detector}} = -\frac{1}{H_c W_c} \sum_{i=1}^{H_c W_c} \sum_{k=1}^{65} y_{i,k} \log p_{i,k}, \quad (1)$$

where $y_{i,k} \in \{0, 1\}$ encodes the ground-truth channel for the spatial location i in the $H_c \times W_c$ output grid. Then, in the homographic adaptation phase, the trained detector is applied to multiple random homographic warps of real, unlabeled images, resulting much richer set of interest points than traditional corner detectors – not only corners and intersections but also textured patches, edges and other repeatable patterns. Detections are warped back to form a pseudo ground-truth, and the network is fine-tuned using the same loss (1).

The descriptor head (Fig. 1, lower branch) outputs $\mathbf{f} \in \mathbb{R}^{H_c \times W_c \times D}$, which is bicubically upsampled to $H \times W \times D$ and is L_2 -normalized for every pixel. The descriptor training uses the same real images and random homographic warps as in the homographic adaptation stage. Keypoints identified by the detector from the previous training stage are extracted from both the original and warped images, and the known homography is used to find corresponding positions between them. Let $d_{ij} = \|\mathbf{f}_i - \mathbf{f}_j\|_2$ be the Euclidean distance between descriptors and $s_{ij} \in \{0, 1\}$ indicate the matches. The descriptor loss is the contrastive loss (Goodfellow et al., 2016):

$$\mathcal{L}_{\text{descriptor}} = \sum_{i=1}^N \sum_{j=1}^M \lambda s_{ij} \max(0, m_p - d_{ij}) + (1 - s_{ij}) \max(0, d_{ij} - m_n),$$

where m_p and m_n are the positive and negative margins, respectively, and λ balances the contributions of positive and negative terms. The first term penalizes matching pairs ($s_{ij} = 1$) whose distance exceeds the positive margin m_p , encouraging descriptors for corresponding points to be close in feature space. The second term penalizes non-matching pairs ($s_{ij} = 0$) whose distance is less than the negative margin m_n , pushing unrelated descriptors apart. The network parameters for both detector and descriptor heads are optimized jointly using stochastic gradient descent with backpropagation.

3. Comparative Study

To objectively compare feature-based image matching methods, such as SIFT and SuperPoint, we adopt a geometric verification strategy based on the Random Sample Consensus (RANSAC) (Fischler, & Bolles, 1981) inlier statistics. Although learned matchers like SuperGlue (Sarlin et al., 2020) significantly enhance the performance of the SuperPoint, this work isolates the contribution of the feature extractor itself by using a standard RANSAC-based evaluation pipeline, ensuring compatibility with both classical and learned descriptors.

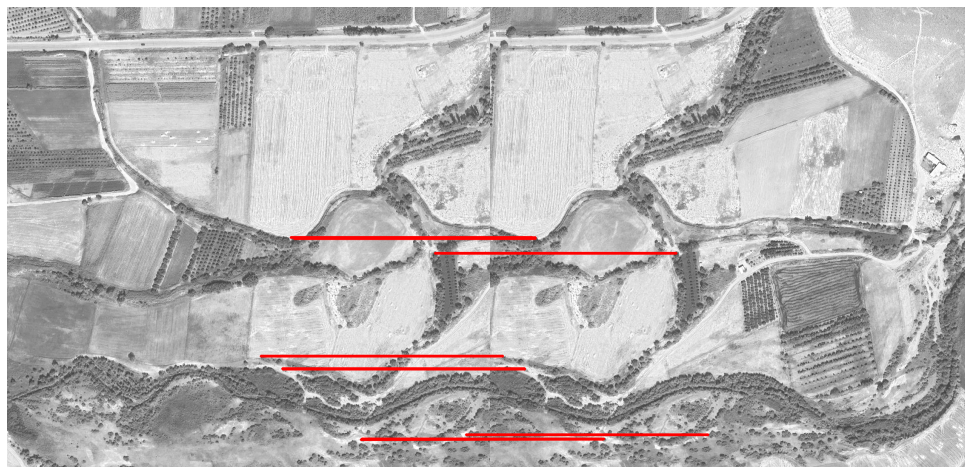


Fig. 2. Keypoint matches between two neighbouring images with an overlapping region. Keypoints are obtained by using the SIFT. The red lines indicate inlier keypoint matches retained by the RANSAC filtering. A collection of the satellite images is retrieved by utilizing the Google Maps Static API (Google, 2025)

Given two images of the same scene under varying viewpoint (Fig. 2), a local feature extractor (e.g., SIFT or SuperPoint) identifies keypoints and their descriptors. The descriptor matching yields the candidate correspondences which are typically noisy and include incorrect matches due to descriptor ambiguity or repeated structures. To filter out incorrect matches, the RANSAC geometric model estimation is applied, estimating a homography matrix $\mathbf{H}$ by iteratively selecting random minimal samples and maximizing the number of inliers, defined as matches that satisfy the geometric model within a tolerance ε . A match (x_i, x_i') is counted as an inlier if $\|x_i' - \mathbf{H}x_i\| < \varepsilon$. This leads to a set of inlier correspondences $I \subset M$, from the full set of candidate matches M . We define the inlier ratio as $r = |I|/|M|$. This ratio quantifies the geometric consistency of matched keypoints. A high inlier ratio indicates that many descriptor matches are geometrically valid and, therefore, reflects higher feature precision and robustness.

Comparison of accuracy and computational efficiency. The number next to each method indicates the number of keypoints extracted per image. Measurements are taken on the same hardware setup in CPU-only mode. **Table 1**

Algorithm	Inlier ratio r (overlapping images), %	Inlier ratio r (non-overlapping images), %	Detection & description time, ms per image
SIFT 100	32.21	0.53	111.53
SIFT 200	32.80	0.59	112.83
SIFT 500	34.45	0.42	115.00
SuperPoint 100	38.96	0.02	728.84
SuperPoint 200	41.34	0.04	729.47
SuperPoint 500	43.81	0.08	726.19

Discussion and conclusions

We compared the classical SIFT detector-descriptor pipeline with the learned SuperPoint method in the context of image matching for a vision-guided system. Both methods are evaluated using a custom-constructed dataset of satellite images, with the RANSAC-based geometric verification employed to filter correspondences and compute the inlier ratio as the primary accuracy metric. While the SIFT offers lower computational cost and CPU-friendly performance, the SuperPoint achieves insignificantly higher inlier ratios and improved robustness to viewpoint and illumination changes, at the expense of multiply increased inference time and hardware requirements (Tab. 1). In line with (Bojanić et al., 2019; Prystavka et al., 2025) comparative studies and related evaluations, our findings confirm that classical algorithms retain a practical advantage for real-time navigation tasks and deliver competitive performance, often matching learned approaches. These findings are relevant to navigational systems (Se, Lowe, & Little, 2005; Mur-Artal, Montiel, & Tardos, 2015; Zhu, Liu, & Cai, 2015) and vision-guided vehicles (Hoffmann et al., 2006; Kawamura et al., 2023), where computational resources, energy consumption, and real-time constraints must be balanced.

Future work will extend this analysis to the challenging task of cross-view matching between aerial/satellite imagery, where large scale variations, seasonal changes, and lighting differences present additional difficulties. We also plan to investigate descriptor aggregation methods, e.g. VLAD, to enable efficient large-scale image retrieval, which can benefit both navigation and mapping pipelines.

Sources of funding. This work was supported by the Simons Foundation grant (SFI-PD-Ukraine-00014586, O.S.) and the project 0125U000299 of the National Academy of Sciences of Ukraine.

References

- Bojanić, D., Bartol, K., Pribanić, T., Petković, T., Donoso, Y. D., & Mas, J. S. (2019). On the comparison of classic and deep keypoint detector and descriptor methods. In *IEEE 11th symposium on image and signal processing and analysis* (pp. 64–69). <https://doi.org/10.1109/ISPA.2019.8868792>
- Boyun, V., Kasim, A., Voznenko, L., & Matvienko, O. (2025). Three models for creating stereo images from UAV sensor data. *International scientific conference Information technologies and computer modelling*, 111–114 [in Ukrainian]. [Бойун, В., Касім, А., Возненко, Л., & Матвієнко, О. (2025). Три моделі створення стереозображень за сенсорними даними БПЛА. *Міжнародна науково-практична конференція Інформаційні технології та комп'ютерне моделювання*, 111–114]. <https://journal.comp-sc.if.ua/test/index.php/ITCM/article/view/716>
- DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In *IEEE conference on computer vision and pattern recognition workshops* (pp. 224–236). <https://doi.org/10.1109/CVPRW.2018.00060>
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395. <https://doi.org/10.1145/358669.358692>
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep Learning*. MIT Press.
- Google. (2025). *Maps static API* (Accessed: 14.08.2025). <https://developers.google.com/maps/documentation/maps-static/>
- Hoffmann, F., Nierobisch, T., Seyffarth, T., & Rudolph, G. (2006). Visual servoing with moments of SIFT features. In *IEEE conference on systems, man and cybernetics* (p. 4262–4267). <https://doi.org/10.1109/ICSMC.2006.384804>
- Huang, Q., Guo, X., Wang, Y., Sun, H., & Yang, L. (2024). A survey of feature matching methods. *IET Image Processing*, 18(6), 1385–1410. <https://doi.org/10.1049/ipr2.13032>
- Kawamura, E., Dolph, C., Kannan, K., Brown, N., Lombaerts, T., & Ippolito, C. A. (2023). VSLAM and vision-based approach and landing for advanced air mobility. In *AIAA SciTech 2023 Forum* (p. 2196). <https://doi.org/10.2514/6.2023-2196>
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5), 1147–1163. <https://doi.org/10.1109/TRO.2015.2463671>
- Prystavka, P., Cholyshkina, O., Kovtun, O., & Pryshchepa, D. (2025). Automation of UAV navigation support based on SIFT-like methods. In *International Workshop on Computational Intelligence (IWSCI)* (pp. 227–239). <https://ceur-ws.org/Vol-4035/>
- Rusyn, B., Lutsyk, O., & Kosarevych, R. (2021). Evaluating the informativity of a training sample for image classification by deep learning methods. *Cybernetics and Systems Analysis*, 57(6), 853–863. <https://doi.org/10.1007/s10559-021-00411-4>
- Sarlin, P.-E., DeTone, D., Malisiewicz, T., & Rabinovich, A. (2020). Superglue: Learning feature matching with graph neural networks. In *IEEE conference on computer vision and pattern recognition* (pp. 4938–4947). <https://doi.org/10.1109/CVPR42600.2020.00499>
- Scaramuzza, D., & Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE robotics & automation magazine*, 18(4), 80–92. <https://doi.org/10.1109/MRA.2011.943233>
- Se, S., Lowe, D. G., & Little, J. J. (2005). Vision-based global localization and mapping for mobile robots. *IEEE Transactions on robotics*, 21(3), 364–375. <https://doi.org/10.1109/TRO.2004.839228>
- Zhu, Q., Liu, C., & Cai, C. (2015). A novel robot visual homing method based on SIFT features. *Sensors*, 15(10), 26063–26084. <https://doi.org/10.3390/s151026063>

Отримано редакцією журналу / Received: 26.05.25

Прорецензовано / Revised: 21.09.25

Схвалено до друку / Accepted: 10.10.25

Олег САМОЙЛЕНКО, асп.

ORCID ID: 0000-0003-4466-4004

e-mail: oleh.samoilenko@imath.kiev.ua

Інститут математики НАН України, Київ, Україна

МАТЕМАТИЧНІ ПРОБЛЕМИ ЗІСТАВЛЕННЯ ОЗНАК ДЛЯ ОБ'ЄКТІВ З ОБМЕЖЕНИМИ РЕСУРСАМИ, КЕРОВАНИХ НА ОСНОВІ ЗОРУ

Візуальне сприйняття відіграє ключову роль в автономних транспортних засобах, де зображення слугують основним вхідним сигналом для навігації, побудови карт, оцінки стану та керування рухом. Такі системи часто працюють в умовах реального часу та обмежених ресурсів, що вимагає алгоритмів виявлення та зіставлення ознак, які є точними й обчислювально ефективними. Важливою складовою таких обчислювальних послідовностей є визначення ключових точок – характерних місць зображення $x \in \mathbb{R}^2$ у цифровій функції зображення $I: \mathbb{R}^2 \rightarrow \mathbb{R}^c$, де c – кількість каналів, а $I(x) \in \mathbb{R}^c$ – локальна інтенсивність пікселя. Кожна виявлена ключова точка асоціюється з дескриптором $d \in \mathbb{R}^D$, що кодує локальну візуальну структуру у формі, інваріантній до зсуву, обертання та помірної зміни масштабу. Ключові точки між двома зображеннями x_i та x'_i зіставлено шляхом порівняння дескрипторів, утворюючи відповідності $x'_i \leftrightarrow x_i$. Геометричну верифікацію виконано шляхом оцінки матриці перетворення $H \in \mathbb{R}^{3 \times 3}$ і прийняття $|x'_i - Hx_i| < \varepsilon$, де ε – допустима похибка. Частку таких внутрішніх відповідей, що називається коефіцієнтом валідних відповідей, використано як метрику точності. Математично оцінено класичні методи визначення ключових точок, такі як Scale-Invariant Feature Transform (SIFT), у порівнянні з методами на основі машинного навчання, такими як SuperPoint, застосовуючи їх до вручну сформованого набору супутникових зображень. Продуктивність проаналізовано з позиції коефіцієнта валідних відповідей та обчислювальної ефективності, що відображає компроміс між надійністю й використанням ресурсів. Отримані результати покликані надати рекомендації щодо інтеграції легких, але точних алгоритмів комп'ютерного зору в платформи, такі як БПЛА, забезпечуючи надійну візуальну одометрію та одночасну локалізацію і картографування в умовах обмежених апаратних ресурсів.

Ключові слова: безпілотні літальні апарати, зіставлення зображень, виявлення ознак, комп'ютерний зір, розпізнавання образів.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження(у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

УДК 519.9

DOI: <https://doi.org/10.17721/1812-5409.2025/2.22>

Олексій ХАРИТОНОВ, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0002-6529-7736

e-mail: kharytonov@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Олексій КУЦЕНКО, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0001-5089-0409

e-mail: oleksii.kutsenko@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Анастасія КУЦЕНКО, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0003-1808-6773

e-mail: kutsenko@nubip.edu.ua

Національний університет біоресурсів і природокористування України, Київ, Україна

АНАЛІЗ ЕФЕКТИВНОСТІ КОМБІНУВАННЯ УЧАСТІ ДВИГУНІВ ВЕЛИКОЇ ТА МАЛОЇ ТЯГИ ПІД ЧАС ПЕРЕЛЬОТІВ ЗЕМЛЯ – МАРС

Міжпланетні пілотовані перельоти потребують швидкої й економічної реалізації. Задовольнити зазначені вимоги можна шляхом поєднання участі двигунів великої та малої тяги, які в змозі забезпечити належні потужність й енерго-ефективність. У пропонованому дослідженні розглянуто задачу "спринтерського" перельоту Земля – Марс, що полягає в переведенні космічного апарату із заданої кругової орбіти навколо Землі на задану кругову орбіту навколо Марсу за заданий час, менший за час класичного перельоту типу Гомана. Комбінування участі двигунів великої та малої тяги приводить до виникнення задачі оптимізації розподілу сумарної характеристичної швидкості маневру між ділянками великої та малої тяги. При цьому рух на геліоцентричній ділянці траєкторії відбувається не за звичайною Кеплерівською траєкторією, а за близькою до неї дугою, що формується двигуном малої тяги. Двигун великої тяги забезпечує маневри у сфері впливу планет. Завдяки цьому, порівняно із класичним варіантом застосування виключно великої тяги, загальний час маневру не змінюється, а частина палива використовується ефективнішою системою малої тяги. Проте виникає питання: за яких умов маса зекономленого в такий спосіб палива перевершить масу додатково встановленої рушійної системи малої тяги. У пропонованій статті отримано оцінку тривалості перельоту, що обмежує зверху інтервал тривалості перельотів, у якому комбінування двигунів великої та малої тяги є оптимальним.

Ключові слова: міжпланетний переліт Земля – Марс, комбінування великої та малої тяги, ядерний двигун, метод транспортувальної траєкторії.

Класифікація відповідно до AMS 2020: 70F15, 34H05.

Вступ

Створення бази на Марсі з подальшою колонізацією планети – одне з головних завдань сучасної космонавтики, розв'язання якого вимагає застосування нових типів рушійних систем. Сучасна концепція здійснення пілотованих перельотів Земля – Марс, прийнята в NASA (див., напр., InspecNet, 2023; Gosse et al., 2023) передбачає застосування багатофункціональних ядерних джерел енергії, які використовуються для утворення реактивної тяги і рушійними системами великої тяги (ядерні теплові ракетні двигуни, ТРД) і рушійними системами малої тяги (електричні ракетні двигуни, ЕРД). Таке поєднання потужніших (ЯТРД) й економічніших (ЕРД) двигунів дало змогу обґрунтувати можливість здійснення перельоту Земля – Марс усього за 45 днів (Gosse et al., 2023). Щоб реалізувати такий "спринтерський" міжпланетний трансфер, NASA обрало вже достатньо відпрацьовану технічно концепцію дворежимного ядерного ракетного двигуна (Gosse et al., 2023), у якому ядерний реактор використовують як джерело теплової енергії рушійної системи великої тяги на основі ЯТРД типу NERVA і як джерело електричної енергії для живлення ЕРД магнітоплаз-модинамічного типу. Отже, аналіз оптимальних траєкторій космічних апаратів (КА) з ДЯРД є актуальною задачею сучасної космонавтики. У представленій роботі запропоновано простий і водночас ефективний алгоритм оцінки переваг від застосування ДЯРД порівняно із класичним підходом, за якого траєкторія формується виключно двигуном великої тяги. Раніше зазначені ідеї були використані при оптимізації маневру з виведення КА на сферу впливу Марса (Харитонов, Савченко, & Міранда, 2019). У пропонованій роботі отримані результати розповсюджені на випадок дво-імпульсного міжпланетного перельоту Земля – Марс.

1. Постановка задач

Досліджується задача оптимізації міжпланетного перельоту Земля – Марс за фіксованої тривалості маневру T . Оптимізації підлягає розподіл загальної характеристичної швидкості маневру Δv між двома імпульсами великої тяги, що формують планетоцентричні ділянки траєкторії, і геліоцентричною ділянкою малої тяги за критерієм максимуму корисного навантаження.

Використовуючи підхід імпульсної апроксимації активних ділянок великої тяги та модель ідеального двигуна обмеженої потужності для опису руху на геліоцентричній ділянці (Beletsky, 2012), отримуємо аналітичний вираз для маси корисного навантаження:

$$M_{\pi} = \frac{M_0 e^{-\frac{(\Delta v_e + \Delta v_m)}{V}}}{1 + \frac{M_0}{2N} e^{-\frac{\Delta v_e}{V}} \cdot J(\vec{v}_{\infty}^e, \vec{v}_{\infty}^m)} - M_x(N) - M_t, \quad (1)$$

© Харитонов Олексій, Куценко Олексій, Куценко Анастасія, 2025

$$\Delta v_e = \sqrt{2 \frac{\mu_e}{R_e} + v_{\infty}^e{}^2} - \sqrt{\frac{\mu_e}{R_e}}, \quad \Delta v_m = \sqrt{2 \frac{\mu_m}{R_m} + v_{\infty}^m{}^2} - \sqrt{\frac{\mu_m}{R_m}}, \quad (2)$$

де M_0 – початкова маса КА, M_{π} – маса корисного навантаження, M_x – маса рушійної системи, M_t – маса баків, N – електрична потужність живлення двигуна малої тяги, V – ефективна швидкість витікання, μ_e, μ_m – гравітаційні параметри Землі і Марса, R_e, R_m – радіуси планетоцентричних кругових орбіт старту та призначення, $\vec{v}_{\infty}^e, \vec{v}_{\infty}^m$ – гіперболічні надлишки швидкості для відповідних орбіт перельоту між цими орбітами та сферами впливу, $J(\vec{v}_{\infty}^e, \vec{v}_{\infty}^m)$ – функціонал варіаційної задачі оптимізації геліоцентричного маневру з малою тягою.

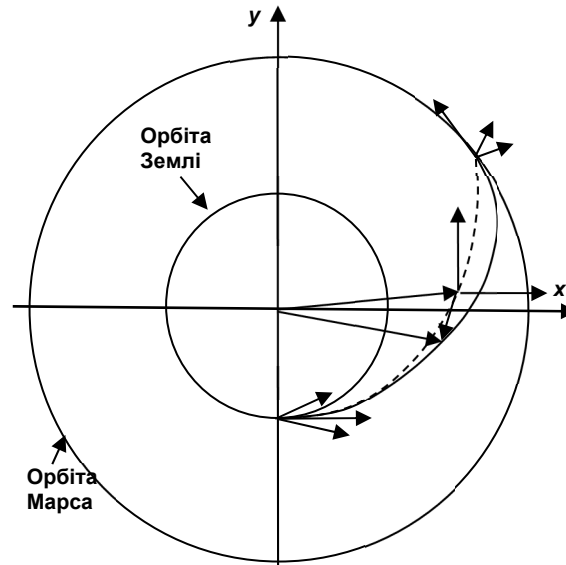


Рис. 1. Схема міжпланетного перельоту й ілюстрація методу транспортувальної траєкторії. Реальна (суцільна лінія) траєкторія КА та транспортувальна траєкторія (штрихова лінія)

У контексті постановки задачі оптимізації геліоцентричного маневру зауважимо, що максимізація M_{π} вимагає мінімізації функціоналу $J(\vec{v}_{\infty}^e, \vec{v}_{\infty}^m)$. З метою отримання аналітичного виразу функціонала ми можемо застосувати метод транспортувальної траєкторії (Beletsky, 2012). Як транспортувальну орбіту використовуємо Кеплерів еліпс, що забезпечує переліт між геліоцентричними круговими орбітами Землі і Марсу за заданий час за мінімальної характеристичної швидкості перельоту (штрихова лінія, рис. 1). На цьому рисунку $\vec{v}_{\infty}^e, \vec{v}_{\infty}^m$ – гіперболічні надлишки швидкості для планетоцентричних орбіт, що відповідають такому маневру. Увімкнення ділянки малої тяги необхідне для зменшення затрат характеристичної швидкості на маневри з великою тягою, так щоб $\Delta v_e + \Delta v_m < \Delta v_{e,t} + \Delta v_{m,t}$, де прирости швидкості визначаються (2). Таке зменшення означає, що частина енергетичних затрат на маневр забезпечується економічнішим двигуном малої тяги. Проте така схема, що передбачає увімкнення двигуна малої тяги, буде виправдана тоді, коли маса зекономленого завдяки його використанню палива перевищить масу додаткової рушійної системи й обладнання, потрібного для генерування малої тяги.

Відповідно до методу транспортувальної траєкторії, подамо радіус-вектор центру мас КА $\vec{r}(t)$ у вигляді:

$$\vec{r}(t) = \vec{r}_t(t) + \vec{\rho}(t), \quad (3)$$

де $\vec{r}_t$ – радіус-вектор початку транспортувальної системи координат, що рухається по Кеплеровому еліпсу, $\vec{\rho}(t) = (\xi(t), \eta(t))$ – радіус-вектор центру мас КА у транспортувальній системі координат $O_t \xi \eta$.

Постановка варіаційної задачі руху з малою тягою у транспортувальній системі координат буде такою:

$$\begin{aligned} \dot{\xi} &= v_{\xi}, \quad \dot{\eta} = v_{\eta}, \quad \dot{v}_{\eta} = a \cos \cos(\theta), \quad \dot{v}_{\xi} = \sin \sin(\theta), \quad \dot{J} = a^2, \\ \xi(0) &= 0, \quad \eta(0) = 0, \quad J(0) = 0, \quad v_{\xi}(0) = v_{\xi}^e, \quad v_{\eta}(0) = v_{\eta}^e, \\ \xi(T) &= 0, \quad \eta(T) = 0, \quad v_{\xi}(T) = v_{\xi}^m, \quad v_{\eta}(T) = v_{\eta}^m, \quad J(T) \rightarrow \min. \end{aligned} \quad (4)$$

У (4) введено такі позначення:

$$v_x^e = v_{\infty,x}^e - v_{\infty,t,x}^e, \quad v_y^e = v_{\infty,y}^e - v_{\infty,t,y}^e, \quad v_x^m = v_{\infty,x}^m - v_{\infty,t,x}^m, \quad v_y^m = v_{\infty,y}^m - v_{\infty,t,y}^m,$$

θ – кут між вектором тяги і віссю Ox (рис. 1), a – реактивне прискорення.

2. Результати та їхній аналіз

Задача (4) має аналітичний розв'язок:

$$J(T) = \frac{4}{T} \{ [v_x^e{}^2 + v_x^e v_x^m + v_x^m{}^2] + [v_y^e{}^2 + v_y^e v_y^m + v_y^m{}^2] \}. \quad (5)$$

Отже, функціонал $J(T)$, і, на основі (1) – (2) основний функціонал оптимізаційної задачі M_{π} , залежать від вибору векторів $\vec{v}_{\infty}^e, \vec{v}_{\infty}^m$, що визначають планетоцентричні гіперболічні орбіти. Однак від напрямку цих векторів залежить лише функціонал (5), а решта виразу (1) – лише від абсолютних величин цих векторів. Звідси, напрями векторів $\vec{v}_{\infty}^e, \vec{v}_{\infty}^m$ маємо визначати з такої задачі на умовний екстремум:

$$\begin{aligned} J(v_{\infty,x}^e, v_{\infty,y}^e, v_{\infty,x}^m, v_{\infty,y}^m) &\rightarrow \min, \\ v_{\infty,x}^e{}^2 + v_{\infty,y}^e{}^2 &= v_{\infty}^e{}^2, \quad v_{\infty,x}^m{}^2 + v_{\infty,y}^m{}^2 = v_{\infty}^m{}^2. \end{aligned} \quad (6)$$

Отже, зважаючи на розв'язок (6), лише дві із чотирьох компонент векторів $\vec{v}_{\infty}^e, \vec{v}_{\infty}^m$ є незалежними. Відповідні співвідношення мають вигляд:

$$\begin{aligned} v_{\infty,y}^e &= \frac{v_{\infty,x}^e \{v_{\infty,t,y}^m + 2v_{\infty,t,y}^m\} + \{v_{\infty,x}^e - v_{\infty,t,x}^e - 2v_{\infty,t,x}^m\} (v_{\infty,t,y}^m + 2v_{\infty,t,y}^e)}{v_{\infty,x}^e v_{\infty,x}^m - \{v_{\infty,x}^e - v_{\infty,t,x}^e - 2v_{\infty,t,x}^m\} \{v_{\infty,x}^m - v_{\infty,t,x}^m - 2v_{\infty,t,x}^e\}}, \\ v_{\infty,y}^m &= \frac{v_{\infty,x}^m \{v_{\infty,t,y}^e + 2v_{\infty,t,y}^e\} + \{v_{\infty,x}^m - v_{\infty,t,x}^m - 2v_{\infty,t,x}^e\} (v_{\infty,t,y}^e + 2v_{\infty,t,y}^m)}{v_{\infty,x}^m v_{\infty,x}^e - \{v_{\infty,x}^m - v_{\infty,t,x}^m - 2v_{\infty,t,x}^e\} \{v_{\infty,x}^e - v_{\infty,t,x}^e - 2v_{\infty,t,x}^m\}}. \end{aligned} \quad (7)$$

Для ілюстрації ефективності запропонованого підходу задачу розглянуто для різних значень тривалості маневру. Припускали, що стартова маса КА становить 160 т, орбіти старту і призначення кругові і мають висоту 400 км. У виразі (1) для розрахунку маси рушійної малої тяги використано емпіричний вираз з роботи (Kharytonov, & Kiforenko, 2011), що подає M_x у вигляді функції електричної потужності:

$$M_x(N) = \alpha(N)N, \quad \alpha(N) = n \left(\frac{A}{BN + C} + F \right), \quad (8)$$

де можна прийняти $A = 72\text{кг}$, $B = 0,01$, $C = 1,05\text{кВт}$, $F = 2,83\text{кг/кВт}$, $n = 0,77$. Коефіцієнт n уведений для врахування можливості подвійного використання ядерного реактора – як джерела теплової енергії і як джерела електричної енергії для живлення двигунів малої тяги. Якщо рушійна система малої тяги має окреме джерело енергії, то $n = 1$.

На рис. 2 а наведені залежності безрозмірної маси корисного навантаження $m_{\pi} = \frac{M_{\pi}}{M_0}$ від електричної потужності N для різних тривалостей перельоту від 240 до 140 днів, менших за тривалість перельоту типу Гомана (259 днів). Для кожної тривалості перельоту і для кожного значення N від 0 до 6МВт, унаслідок мінімізації виразу (1) з урахуванням (6), було визначено оптимальний розподіл паливних витрат між ділянками великої і малої тяги (ці витрати цілком визначаються векторами $\vec{v}_{\infty}^e, \vec{v}_{\infty}^m$). Відповідні оптимальні значення m_{π} наведені на рис. 2 а.

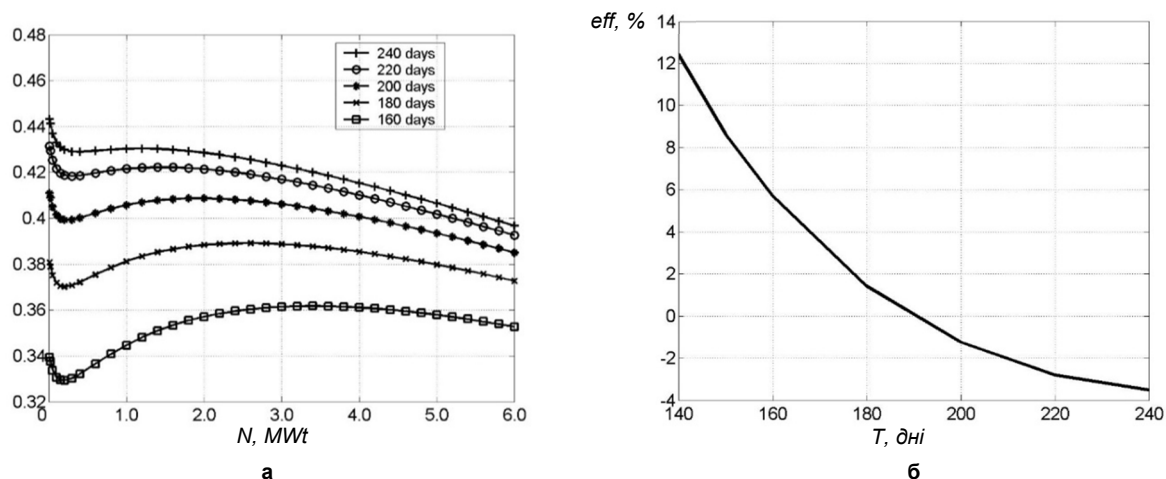


Рис. 2. Ефективність комбінування великої і малої тяги

Наведені на рис. 2 а криві мають локальні мінімуми й максимуми. Наявність мінімумів пояснюємо тим, що, урахувавши фіксовану тривалість перельоту, за малих значень електричної потужності рушійна система малої тяги здатна забезпечити зовсім незначний відсоток енергозатрат на маневр, а отже, її маса фактично стає мертвою. Локальні максимуми визначають оптимальні значення N , для яких маса зекономленого завдяки використанню рушійної системи малої тяги палива найкраще співвідноситься з масою самої рушійної системи. Якщо значення m_{π} в такому максимумі перевищує значення за $N = 0$, то застосування дворежимної рушійної системи стає ефективним, адже маса зекономленого палива перевищує масу рушійної системи малої тяги. Чим менша тривалість маневру, тим вищими є енергозатрати на його виконання, а отже – тим більшими є можливості для економії палива внаслідок застосування системи малої тяги. Тому за зменшення тривалості перельоту максимуми на кривих рис. 2 а зміщуються праворуч.

Ефективність застосування підходу комбінування великої і малої тяги залежно від тривалості перельоту проілюстровано на рис. 2 б. Ефективність обчислювали за формулою

$$eff = \frac{M_{\pi max} - M_{\pi}(0)}{M_{\pi}(0)} \times 100\%,$$

де $M_{\pi max}$ – максимальне значення корисного навантаження за комбінування великої і малої тяги (локальний максимум кривих на рис. 2 а), $M_{\pi}(0)$ – значення корисного навантаження за використання тільки великої тяги.

Як впливає з рис. 2 б, комбінування великої та малої тяги є ефективним, якщо тривалість перельоту не перевищує 190 днів.

Дискусія і висновки

За отриманими результатами можна попередньо оцінити ефективність застосування підходу комбінування двигунів великої та малої тяги для здійснення міжпланетних перельотів. Ефективність зростає за зменшення тривалості перельоту, зважаючи на те, що при цьому збільшується відсоток загальних енерговитрат, який може бути забезпечений рушійною системою малої тяги. Водночас для розрахунку короткотривалих й енергозатратних перельотів потрібно шукати точніші методи отримання аналітичного виразу залежності функціоналу геліоцентричного маневру від планетоцентричних швидкостей на сферах впливу $J(\vec{v}_{\infty}^e, \vec{v}_{\infty}^m)$.

Внесок авторів: Олексій Харитонов – постановка задачі, побудова методики її розв'язання; Олексій Куценко – здійснення розрахунків; Анастасія Куценко – аналіз джерел, підготовка теоретичних основ дослідження.

Джерела фінансування. Фінансування здійснюється власним коштом авторів і частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Харитонов, О. М., Савченко, С. Р., & Міранда, Н. (2019). Оптимізація швидких міжпланетних траєкторій з використанням ядерного дворежимного двигуна. *Вісник Київського національного університету імені Тараса Шевченка. Серія: Фізико-математичні науки*, 1, 74–77.
- Beletsky, V. (2012). *Essays in the Motion of Celestial Bodies*. Birkhäuser Basel. <https://doi.org/10.1007/978-3-0348-8360-3>
- Gosse, R., Osborne, G., Mahmood, H., Burrows, S. K., Roy, S., Watson, J., & Tonks, M. (2024). Nuclear Wave Rotor Bi-Modal Cycle for In-Space Propulsion [Conference proceedings]. *AIAA SCITECH 2024 Forum*. <https://doi.org/10.2514/6.2024-1238>
- InspeNet. (2023). NASA approves nuclear-rocket prototype to travel to Mars in just 45 days, InspeNet. <https://inspenet.com/en/noticias/nasa-approves-nuclear-rocket-prototype-to-travel-to-mars-in-just-45-days>
- Kharytonov, O. M., & Kiforenko, B. M. (2011). Finite-thrust optimization of interplanetary transfers of space vehicle with bimodal nuclear thermal propulsion. *Acta Astronautica*, 69(3), 223–233. <https://doi.org/10.1016/j.actaastro.2011.03.003>

References

- Beletsky, V. (2012). *Essays in the Motion of Celestial Bodies*. Birkhäuser Basel. <https://doi.org/10.1007/978-3-0348-8360-3>
- Gosse, R., Osborne, G., Mahmood, H., Burrows, S. K., Roy, S., Watson, J., & Tonks, M. (2024). Nuclear Wave Rotor Bi-Modal Cycle for In-Space Propulsion [Conference proceedings]. *AIAA SCITECH 2024 Forum*. <https://doi.org/10.2514/6.2024-1238>
- InspeNet. (2023). NASA approves nuclear-rocket prototype to travel to Mars in just 45 days, InspeNet. <https://inspenet.com/en/noticias/nasa-approves-nuclear-rocket-prototype-to-travel-to-mars-in-just-45-days>
- Kharytonov, O. M., & Kiforenko, B. M. (2011). Finite-thrust optimization of interplanetary transfers of space vehicle with bimodal nuclear thermal propulsion. *Acta Astronautica*, 69(3), 223–233. <https://doi.org/10.1016/j.actaastro.2011.03.003>
- Kharytonov, O. M., Savchenko, S. R., & Miranda, N. (2019). Optimization of sprint interplanetary trajectories with nuclear bimodal thermal propulsion. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics & Mathematics*, 1, 74–77.

Отримано редакцією журналу / Received: 26.08.25

Прорецензовано / Revised: 06.10.25

Схвалено до друку / Accepted: 10.10.25

Oleksii KHARYTONOV, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-6529-7736
e-mail: oleksii.kutsenko@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Oleksii KUTSENKO, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0001-5089-0409
e-mail: oleksii.kutsenko@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Anastasiia KUTSENKO, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0003-1808-6773
e-mail: kutsenko@nubip.edu.ua
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

ANALYSIS OF THE EFFECTIVENESS OF COMBINING HIGH-AND LOW-THRUST ENGINES IN EARTH – MARS FLIGHTS

Interplanetary manned flights require rapid and economical implementation. These requirements can be met by combining high-thrust and low-thrust engines, which can provide the necessary power and energy efficiency. The problem of a "sprint" flight from Earth to Mars is considered, which consists in transferring a spacecraft from a given circular geocentric orbit to a given circular areocentric orbit in a given time, less than the time of a classic Hohmann-type transfer. The combination of high-thrust and low-thrust engines leads to the problem of optimizing the distribution of the total Δv budget between the high-thrust and low-thrust maneuvers. At the same time, the heliocentric part is formed by the low-thrust engine and differs from the usual Keplerian arc. The high-thrust engine provides maneuvers in the spheres of influence of the planets. Thanks to this, compared to the classic option of using only high thrust, the total maneuver time does not change, and part of the fuel is used by a more efficient low-thrust system. This raises the question: under what conditions will the mass of fuel saved in this way exceed the mass of the additionally installed low-thrust propulsion system? The paper provides an estimate of flight duration, which assesses the upper limit of the flight duration range in which the combination of high- and low-thrust engines is optimal.

Key words: Earth – Mars interplanetary flight, combination of high and low thrust, nuclear engine, transporting trajectory method.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 519.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.23>

Andrii CHAIKOVSKYI, DSc (Phys. & Math.), Prof.
ORCID ID: 0000-0002-3357-8744
e-mail: andriichaikovskiy@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Oleksandr LIUBIMOV, Student
ORCID ID: 0009-0009-9347-6297
e-mail: liubimov_oleksandr@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Oksana LAHODA, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0003-4515-1753
e-mail: lahoda.oksana@iit.kpi.ua

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

BOUNDEDNESS OF SOLUTIONS OF THE FIRST ORDER LINEAR MULTIDIMENSIONAL DIFFERENCE EQUATIONS

We investigate the boundedness of solutions of the first order linear difference equation of the form $x_{n+1} = Ax_n + y_n$, $n \geq 1$ where A is a square matrix with complex entries, sequence $\{y_n\}_{n \geq 1}$ and initial value x_1 are supposed to be known. Firstly, we discuss the one-dimensional case of this equation $x_{n+1} = ax_n + y_n$, $n \geq 1$ where a is a complex number. In particular, we obtain the sufficient conditions for boundedness or unboundedness of the solutions in case $|a| = 1$ (the critical case) by considering the exponential sums of the forms $\sum y_n e(n\varphi)$ and $\sum e(f(n))$.

Then we proceed to the investigation of the equation in the multidimensional case and reduce our problem to the analysis of the spectrum and Jordan cells of the matrix A . The problem is especially interesting when the spectrum of A contains eigenvalues λ with $|\lambda| = 1$. At the end of the article we obtain a theorem that reveals the connection between equations $x_{n+1} = ax_n + y_n$, $n \geq 1$ with $|a| = 1$ and $x_{n+1} = Jx_n + y_n$, $n \geq 1$ with J being a Jordan cell of an eigenvalue λ , $|\lambda| = 1$.

Key words: Linear difference equation, bounded solution, exponential sums.

AMS 2020 classification: 39A06, 11L03, 11L07.

Introduction

In space $\mathbb{C}^m$ with Euclidean norm we consider the following difference equation

$$x_{n+1} = Ax_n + y_n, \quad n \geq 1 \quad (1)$$

with respect to unknown sequence $\{x_n : n \geq 1\} \subset \mathbb{C}^m$. The first element of this sequence x_1 , sequence $\{y_n : n \geq 1\} \subset \mathbb{C}^m$ and square matrix $A \in \mathcal{M}_m(\mathbb{C})$ of order m are supposed to be known. Various necessary and sufficient conditions for the existence and uniqueness of the bounded solution of this equation were previously investigated by different authors in finite-dimensional and infinite-dimensional cases (see (Gomilko, Gorodnii, & Lagoda, 2003), (Horodnii, & Lahoda, 2001), (Chaikovskiy, & Lagoda, 2022), (Gorodnii, & Kravets', 2020) and cited literature there).

Difference equation (1) has the following well-studied cases:

- 1) If spectrum of A is such that $\sigma(A) \subset \{z \in \mathbb{C} \mid |z| < 1\}$ then for any $x_1 \in \mathbb{C}^m$ the solution is bounded;
- 2) If $\sigma(A) \subset \{z \in \mathbb{C} \mid |z| > 1\}$ then only for $x_1 = -\sum_{k=1}^{\infty} A^{-k-1} y_k$ the solution is bounded;
- 3) If $\sigma(A) \cap \{z \in \mathbb{C} \mid |z| = 1\} = \emptyset$ then it is possible to decompose the space into a cartesian product of two spaces, which satisfy the conditions of cases 1 and 2.

On the other hand, the case $\sigma(A) \cap \{z \in \mathbb{C} \mid |z| = 1\} \neq \emptyset$ always requires some additional investigation because the simple characterization for it is not known. Existence of the bounded solution of the difference equation (1) in this case depends on the input sequence $\{y_n : n \geq 1\}$ in a rather complicated way in comparison with the cases 1)-3). Our aim is to establish some interesting and significant instances of such a dependency.

1. One-dimensional equation

We consider the difference equation of the form

$$x_{n+1} = ax_n + y_n \quad (2)$$

with respect to unknown sequence $\{x_n\}_{n \geq 1}$. Numbers $x_1, a \in \mathbb{C}$ and the sequence $\{y_n\}_{n \geq 1} \subset \mathbb{C}$ are supposed to be known.

By induction, it is not difficult to deduce that the solution of the equation (2) is

$$x_n = a^{n-1} x_1 + \sum_{k=1}^{n-1} a^{n-k-1} y_k, \quad n \geq 1.$$

For the critical case $|a| = 1$ we investigate the conditions for $\{y_n\}_{n \geq 1}$ and x_1 under which the solution $\{x_n\}_{n \geq 1}$ is bounded or unbounded.

We begin this section with the following observation.

Proposition 1. If $|a| = 1$, then the solution of the difference equation (2) is bounded $\iff \exists M > 0 \forall K \geq 0$:

$$\left| \sum_{k=1}^K \frac{y_k}{a^k} \right| \leq M.$$

Proof. The proof is a simple consequence of the solution formula above. $\square$

The question of boundedness of solution $\{x_n\}_{n \geq 1}$ in case $|a| = 1$ is rather difficult. Unlike in other cases, the simple and full characterization is still not known. So, particular results with sufficient conditions of boundedness are interesting.

We write $a = e^{-2\pi i \varphi}$, where $\varphi \in [0, 1)$. Therefore according to the Proposition 1 solution $\{x_n\}_{n \geq 1}$ of the difference equation (2) is bounded if and only if the sequence of sums

$$\left\{ \sum_{n=1}^N y_n e^{2\pi i n \varphi} : N \geq 1 \right\}$$

is bounded.

Thus, we see that our problem is equivalent to the study of the boundedness of the exponential sums written above.

For real number x define $e(x) := e^{2\pi i x}$.

In this subsection, we are solely concerned with the boundedness of exponential sums

$$\left\{ \sum_{n=1}^N e(f(n)) : N \geq 1 \right\}, \quad (3)$$

where $f : \mathbb{N} \rightarrow \mathbb{R}$ is some function.

Definition 1. For the function $f : \mathbb{N} \rightarrow \mathbb{R}$ we define the **finite difference of f** to be

$$\Delta f(n) := f(n+1) - f(n)$$

and we define the **second order finite difference of f** to be

$$\Delta^2 f(n) := \Delta f(n+1) - \Delta f(n) = f(n+2) - 2f(n+1) + f(n).$$

One of the interesting results connected with the question of boundedness of exponential sums is following.

Theorem 1 (Kusmin-Landau Inequality). Suppose function f and real number θ are such that

$$0 < \theta \leq \Delta f(1) \leq \Delta f(2) \leq \dots \leq \Delta f(N-1) < 1 - \theta.$$

Then

$$\left| \sum_{n=1}^N e(f(n)) \right| \leq \cot\left(\frac{\pi\theta}{2}\right).$$

Proof. For the proof and detailed discussion, see the paper of Mordell (Mordell, 1958). $\square$

We would like to remark that if function f has monotone derivative f' on segment $[1, N]$ and $\theta < f'(x) < 1 - \theta$, then from Lagrange's mean value theorem it follows that f satisfies the conditions of Theorem 1.

The proof of Kusmin-Landau Inequality written in the book of Graham and Kolesnik (Graham, & Kolesnik, 1991) presents us an interesting and useful identity.

Proposition 2.

$$\begin{aligned} - \sum_{n=n_0}^N e(f(n)) &= \frac{1}{2} (1 + i \cdot \cot(\pi \Delta f(N+1))) \cdot e(f(N+1)) - \frac{1}{2} (1 + i \cdot \cot(\pi \Delta f(n_0))) \cdot e(f(n_0)) - \\ &\quad - \frac{1}{2} \sum_{n=n_0}^N e(f(n+1)) \cdot i (\cot(\pi \Delta f(n+1)) - \cot(\pi \Delta f(n))). \end{aligned}$$

Proof. We write

$$e(f(n)) = \frac{e(f(n+1)) - e(f(n))}{e(\Delta f(n)) - 1} = -\frac{1}{2} (e(f(n+1)) - e(f(n))) \cdot (1 + i \cdot \cot(\pi \Delta f(n)))$$

and using Abel's summation by parts formula, we obtain the desired identity. $\square$

Boundedness of exponential sums $\sum_{n \geq 1} e(n\alpha)$, $\alpha \in (0, 1)$ and geometric intuition suggest us a natural question:

If $\Delta f(n) \rightarrow \psi \notin \mathbb{Z}$ as $n \rightarrow \infty$, is it true that the sequence of exponential sums (3) is bounded?

And the answer is Yes! if $\Delta f(n)$ converges fast enough. With the aid of Proposition 2, we prove the following Theorem.

Theorem 2. Let f be such that $\sum_{n \geq 1} |\Delta^2 f(n)| < \infty$. Then:

1. $\Delta f(n) \rightarrow \psi \in \mathbb{R}$ as $n \rightarrow \infty$.
2. If $\psi \notin \mathbb{Z}$ then sequence of exponential sums

$$\left\{ \sum_{n=1}^N e(f(n)) : N \geq 1 \right\}$$

is bounded.

Proof. Firstly, since $\sum_{n \geq 1} |\Delta^2 f(n)| < \infty$ series $\sum_{n \geq 1} \Delta^2 f(n)$ is convergent. Therefore, because

$$\Delta f(N) = \Delta f(1) + \sum_{n=1}^{N-1} \Delta^2 f(n),$$

we deduce that $\exists \lim_{n \rightarrow \infty} \Delta f(n) =: \psi \in \mathbb{R}$.

Now suppose that $\psi \notin \mathbb{Z}$. There exist some $m \in \mathbb{Z}$, $n_0 \in \mathbb{N}$ and $\theta \in (0, 1)$ such that for all $n \geq n_0$ we have

$$m + \theta \leq \Delta f(n) \leq m + 1 - \theta.$$

It means that $|\sin(\pi \Delta f(n))| \geq \sin(\pi \theta) =: T$ for all $n \geq n_0$.

Thus for all $n \geq n_0$ we have

$$\begin{aligned} |\cot \pi(\Delta f(n+1)) - \cot(\Delta f(n))| &= \frac{|\sin(\pi \Delta f(n+1) - \pi \Delta f(n))|}{|\sin(\pi \Delta f(n+1))| |\sin(\pi \Delta f(n))|} = \\ &= \frac{|\sin(\pi \Delta^2 f(n))|}{|\sin(\pi \Delta f(n+1))| |\sin(\pi \Delta f(n))|} \leq \frac{\pi |\Delta^2 f(n)|}{T^2}. \end{aligned}$$

We conclude that for all $N \geq n_0$ we have:

$$\begin{aligned} \left| \sum_{n=n_0}^N e(f(n+1)) \cdot i (\cot(\pi \Delta f(n+1)) - \cot(\pi \Delta f(n))) \right| &\leq \\ &\leq \sum_{n=n_0}^N |\cot(\pi \Delta f(n+1)) - \cot(\pi \Delta f(n))| \leq \frac{\pi}{T^2} \sum_{n=n_0}^N |\Delta^2 f(n)| \leq \frac{\pi}{T^2} \sum_{n \geq 1} |\Delta^2 f(n)| < \infty. \end{aligned}$$

Also

$$|(1 + i \cdot \cot(\pi \Delta f(N+1))) \cdot e(f(N+1))| \leq 1 + |\cot(\pi \theta)|.$$

Thus, according to Proposition 2, sequence of exponential sums $\left\{ \sum_{n=1}^N e(f(n)) \right\}_{N \geq 1}$ is bounded. $\square$

With the help of this Theorem, we can, for example, prove boundedness of the following sequences of exponential sums.

Example. 1. For all $\varphi \in \mathbb{R} \setminus \mathbb{Z}$

$$\left\{ \sum_{n=1}^N e(n\varphi + \sqrt{n} \log(n)) : N \geq 1 \right\}.$$

2. For arbitrary $s > 1$ and $\varphi \in (0, 1)$, $\varphi + \sum_{k \geq 2} \frac{\sin(k)}{k \log^s(k)} \notin \mathbb{Z}$

$$\left\{ \sum_{n=2}^N e \left(n\varphi + \sum_{r=2}^n \sum_{k=2}^r \frac{\sin(k)}{k \log^s(k)} \right) : N \geq 2 \right\}.$$

We proceed to another observation, which is also inspired by geometric intuition.

Theorem 3. If $\frac{1}{N} \sum_{n=1}^N |\Delta f(n)| \rightarrow 0$, $N \rightarrow \infty$, then the sequence of exponential sums

$$\left\{ \sum_{n=1}^N e(f(n)) : N \geq 1 \right\}$$

is unbounded.

Proof. Put $S(N) := \sum_{n=1}^N e(f(n))$ for $N \geq 1$. We suppose by contradiction that $\exists M > 0 \forall N \geq 1 : |S(N)| \leq M$.

Using Abel's summation by parts formula, we get:

$$|S(N) - N| = \left| \sum_{n=1}^N (e(f(n)) - 1) \right| = \left| \sum_{n=1}^N e(f(n))(1 - e(-f(n))) \right| =$$

$$\begin{aligned}
 &= \left| (1 - e(-f(N+1)))S(N+1) - (1 - e(-f(1)))S(1) + \sum_{n=1}^N (e(-f(n+1)) - e(-f(n)))S(n+1) \right| \leq \\
 &\leq 4M + \sum_{n=1}^N |e(-f(n+1)) - e(-f(n))| |S(n+1)| \leq 4M + M \sum_{n=1}^N |-f(n+1) + f(n)| = \\
 &= 4M + M \sum_{n=1}^N |\Delta f(n)|.
 \end{aligned}$$

There exists some N_0 such that for all $N \geq N_0$: $\sum_{n=1}^N |\Delta f(n)| \leq \frac{N}{2M}$ and therefore

$$N - |S(N)| \leq |N - |S(N)|| \leq |S(N) - N| \leq 4M + \frac{1}{2}N,$$

which implies that $|S(N)| \geq \frac{1}{2}N - 4M$. Contradiction. $\square$

Corollary 1. If $\Delta f(n) \rightarrow \psi \in \mathbb{Z}$, $n \rightarrow \infty$, then the sequence of exponential sums

$$\left\{ \sum_{n=1}^N e(f(n)) : N \geq 1 \right\}$$

is unbounded.

Proof. Since $e(f(n)) = e(f(n) - n\psi)$ and $\Delta(f(n) - n\psi) = \Delta f(n) - \psi$ without loss of generality, we may suppose that $\Delta f(n) \rightarrow 0$, $n \rightarrow \infty$. Hence, by Stolz-Cesàro Theorem, $\frac{1}{N} \sum_{n=1}^N |\Delta f(n)| \rightarrow 0$, $N \rightarrow \infty$. The corollary follows from the previous Theorem. $\square$

Now we write $a = e^{-2\pi i \cdot \varphi} = e(-\varphi)$, where $\varphi \in [0, 1)$, and study boundedness of the solutions of difference equation:

$$x_{n+1} = e(-\varphi)x_n + y_n, \quad n \geq 1. \quad (4)$$

We would like to remark that Proposition 1 implies that the boundedness of solutions of (4) depends only on φ and $\{y_n\}_{n \geq 1}$ and does not depend on the initial value x_1 .

Proposition 3. Every solution of the difference equation (4) is bounded if and only if the following sequence of sums

$$\left\{ \sum_{n=1}^N y_n \cdot e(n\varphi) : N \geq 1 \right\}$$

is bounded.

Proof. Follows from the Proposition 1. $\square$

Definition 2. By $\mathbb{W}_\varphi$ we denote the set of all sequences $\{y_n\}_{n \geq 1} \subset \mathbb{C}$ for which every solution of difference equation (4) is bounded.

We remark that $\mathbb{W}_\varphi$ is not empty because $\{0\}_{n \geq 1} \in \mathbb{W}_\varphi$. We make the following easy observation.

Proposition 4. For every $\varphi \in [0, 1)$ the following holds:

- If $\{z_n\}_{n \geq 1}, \{w_n\}_{n \geq 1} \in \mathbb{W}_\varphi$ then $\{z_n + w_n\}_{n \geq 1} \in \mathbb{W}_\varphi$.
- If $\alpha \in \mathbb{C}$ and $\{z_n\}_{n \geq 1} \in \mathbb{W}_\varphi$ then $\{\alpha z_n\}_{n \geq 1} \in \mathbb{W}_\varphi$.
- If $\{z_n\}_{n \geq 1} \in \mathbb{W}_\varphi$, then $\{z_n\}_{n \geq 1}$ is a bounded sequence.
- If $\{b_n\}_{n \geq 1} \subset \mathbb{R}$ is monotone bounded sequence and $\{z_n\}_{n \geq 1} \in \mathbb{W}_\varphi$, then $\{b_n z_n\}_{n \geq 1} \in \mathbb{W}_\varphi$.

Corollary 2. For every $\varphi \in [0, 1)$ $\mathbb{W}_\varphi$ is a vector space over $\mathbb{C}$. In addition, it is the vector subspace of the vector space of all bounded sequences in $\mathbb{C}$.

Theorem 4. Let f be such that $\sum_{n \geq 1} |\Delta^2 f(n)| < \infty$. Then $\Delta f(n) \rightarrow \psi \in \mathbb{R}$ as $n \rightarrow \infty$. If $\psi + \varphi \notin \mathbb{Z}$ and $y_n = e(f(n))$, $n \geq 1$, then every solution of (4) is bounded, that is, $\{e(f(n))\}_{n \geq 1} \in \mathbb{W}_\varphi$.

Proof. Put $g(n) := f(n) + n\varphi$. Then $\Delta^2 g(n) = \Delta^2 f(n)$, $n \geq 1$ which means that $\sum_{n \geq 1} |\Delta^2 g(n)| < \infty$. In addition $\Delta g(n) \rightarrow \psi + \varphi$, $n \rightarrow \infty$ and

$$\left\{ \sum_{n=1}^N y_n \cdot e(n\varphi) : N \geq 1 \right\} = \left\{ \sum_{n=1}^N e(g(n)) : N \geq 1 \right\}.$$

By Theorem 2 sequence of exponential sums $\left\{ \sum_{n=1}^N e(g(n)) \right\}_{N \geq 1}$ is bounded. Hence, by Proposition 3, every solution of (4) is bounded. $\square$

Example. For all $\varphi \in (0, 1)$ every solution of the difference equation

$$x_{n+1} = e(\varphi)x_n + e(n^{\frac{1}{2}} - n^{\frac{1}{3}} + \log(n)), n \geq 1$$

is bounded.

Theorem 5. Let f be such that $\Delta f(n) \rightarrow \psi \in \mathbb{R}$ as $n \rightarrow \infty$. If $\frac{1}{N} \sum_{n=1}^N |\Delta f(n) + \varphi| \rightarrow 0$, $N \rightarrow \infty$ and $y_n = e(f(n))$, $n \geq 1$, then every solution of (4) is unbounded, that is, $\{e(f(n))\}_{n \geq 1} \notin \mathbb{W}_\varphi$.

Proof. Put $g(n) := f(n) + n\varphi$. Then $\Delta g(n) = \Delta f(n) + \varphi$, $n \geq 1$. So $\frac{1}{N} \sum_{n=1}^N |\Delta g(n)| \rightarrow 0$, $N \rightarrow \infty$. Therefore, by Theorem 3 sequence of exponential sums $\left\{ \sum_{n=1}^N e(g(n)) \right\}_{N \geq 1}$ is unbounded.

Because $\left\{ \sum_{n=1}^N y_n \cdot e(n\varphi) \right\}_{N \geq 1} = \left\{ \sum_{n=1}^N e(g(n)) \right\}_{N \geq 1}$, Proposition 3 implies that every solution of the difference equation (4) is unbounded. $\square$

Corollary 3. Let f be such that $\Delta f(n) \rightarrow \psi \in \mathbb{R}$ as $n \rightarrow \infty$. If $\psi + \varphi \in \mathbb{Z}$ and $y_n = e(f(n))$, $n \geq 1$, then every solution of (4) is unbounded, that is, $\{e(f(n))\}_{n \geq 1} \notin \mathbb{W}_\varphi$.

Proof. Put $g(n) := f(n) + n\varphi$. Then $\Delta g(n) = \Delta f(n) + \varphi$, $n \geq 1$. So $\Delta g(n)$ converges to an integer. Therefore, by Corollary 1 sequence of exponential sums $\left\{ \sum_{n=1}^N e(g(n)) \right\}_{N \geq 1}$ is unbounded.

Because $\left\{ \sum_{n=1}^N y_n \cdot e(n\varphi) \right\}_{N \geq 1} = \left\{ \sum_{n=1}^N e(g(n)) \right\}_{N \geq 1}$, Proposition 3 implies that every solution of the difference equation (4) is unbounded. $\square$

Example. For all $\alpha \in (0, 1)$ all the solutions of the following difference equation are unbounded

$$x_{n+1} = x_n + e(n^\alpha), n \geq 1.$$

Example. All the solutions of the difference equation

$$x_{n+1} = e^{i\frac{\pi}{2}} x_n + e^{i \sum_{k=1}^n \arctan(k)}, n \geq 1$$

are unbounded.

2. Multidimensional equation

In space $\mathbb{C}^M$ with Euclidean norm we consider following difference equation

$$x(n+1) = Ax(n) + y(n), n \geq 1 \quad (5)$$

with respect to unknown sequence $\{x(n) : n \geq 1\} \subset \mathbb{C}^M$.

The first element of the sequence $x(1)$, a sequence $\{y(n) : n \geq 1\} \subset \mathbb{C}^M$ and $M \times M$ square matrix A with complex entries are supposed to be known.

Our goal is to find sufficient conditions for the sequence $\{y(n) : n \geq 1\} \subset \mathbb{C}^M$ and value $x(1)$ which guarantee boundedness of the solution in the critical case $\sigma(A) \cap \{z \in \mathbb{C} \mid |z| = 1\} \neq \emptyset$.

Remark. We will use Hilbert-Schmidt norm of matrix $A = (a_{ij}) \in \mathcal{M}_M(\mathbb{C})$ which is defined as $\|A\|_2 = \sqrt{\sum_{i,j} |a_{ij}|^2}$.

Note that the Cauchy-Bunyakovskiy-Schwartz inequality implies

$$\forall x \in \mathbb{C}^M \forall A \in \mathcal{M}_M(\mathbb{C}) : \|Ax\| \leq \|A\|_2 \|x\|.$$

Now we introduce the following notation:

$$(x)^{[n]} := \prod_{i=0}^{n-1} (x+i), \quad (x)_{[n]} := \prod_{i=0}^{n-1} (x-i).$$

These polynomials are widely known as *factorial polynomials* or, more precisely, *raising* and *falling factorials* respectively. They satisfy the useful identities listed below.

Proposition 5. $\binom{x}{n} = \frac{(x)^{[n]}}{n!}$, $\binom{x+n-1}{n} = \frac{(x)^{[n]}}{n!}$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$.

Proposition 6. $(-x)_{[n]} = (-1)^n \cdot (-x)^{[n]}$.

Proposition 7. $(x+y)_{[n]} = \sum_{i=0}^n \binom{n}{i} (x)_{[i]} \cdot (y)_{[n-i]}$.

Proof. Induction on n . $\square$

These properties immediately imply the following lemma, which will be of great importance for us later.

Lemma 1.

$$\binom{n-k}{m} = \sum_{j=0}^m (-1)^{m-j} \binom{n}{j} \frac{(k)^{[m-j]}}{(m-j)!}.$$

We proceed to the analysis of the multidimensional equation. We remark that the computations below are well known; however, for completeness, we would like to include them in our article.

Let J be a Jordan Normal Form (JNF) of the matrix A . There exists an invertible matrix T such that $A = TJT^{-1} \iff J = T^{-1}AT$. Multiply the equation (5) at left by matrix T^{-1} and obtain:

$$T^{-1}x(n+1) = T^{-1}Ax(n) + T^{-1}y(n) \iff T^{-1}x(n+1) = T^{-1}AT(T^{-1}x(n)) + T^{-1}y(n) \iff$$

$$\iff T^{-1}x(n+1) = J(T^{-1}x(n)) + T^{-1}y(n).$$

Put $\{x'(n) : n \geq 1\} := \{T^{-1}x(n)\}$, $\{y'(n) : n \geq 1\} := \{T^{-1}y(n)\}$ and obtain the difference equation

$$x'(n+1) = Jx'(n) + y'(n), \quad n \geq 1. \quad (6)$$

The nonsingularity of the matrix T implies that the solution of the difference equation (5) is bounded if and only if the solution of the difference equation (6) is bounded; thus, we will investigate the equation (6).

In this way, our problem is reduced to the study of the difference equation of the form:

$$x(n+1) = J_{\lambda}x(n) + y(n), \quad n \geq 1, \quad (7)$$

where $J_{\lambda} = \begin{pmatrix} \lambda & 1 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}$ is the Jordan cell that corresponds to the eigenvalue λ of the matrix A of the equation (5).

If the order of J_{λ} is 1, then equation (7) is the First Order Linear Difference Equation in $\mathbb{C}$.

If the order of J_{λ} is $M \geq 2$, then we do the following.

Let $x(n) = \begin{pmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_M(n) \end{pmatrix}$, $y(n) = \begin{pmatrix} y_1(n) \\ y_2(n) \\ \vdots \\ y_M(n) \end{pmatrix}$, then equation (7) can be rewritten as

$$\begin{pmatrix} x_1(n+1) \\ x_2(n+1) \\ \vdots \\ x_{M-1}(n+1) \\ x_M(n+1) \end{pmatrix} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_{M-1}(n) \\ x_M(n) \end{pmatrix} + \begin{pmatrix} y_1(n) \\ y_2(n) \\ \vdots \\ y_{M-1}(n) \\ y_M(n) \end{pmatrix} \iff$$

$$\iff \begin{pmatrix} x_1(n+1) \\ x_2(n+1) \\ \vdots \\ x_{M-1}(n+1) \\ x_M(n+1) \end{pmatrix} = \begin{pmatrix} \lambda x_1(n) + x_2(n) + y_1(n) \\ \lambda x_2(n) + x_3(n) + y_2(n) \\ \vdots \\ \lambda x_{M-1}(n) + x_M(n) + y_{M-1}(n) \\ \lambda x_M(n) + y_M(n) \end{pmatrix}, \quad n \geq 1.$$

The solution of the difference equation $x_M(n+1) = \lambda x_M(n) + y_M(n)$ is

$$x_M(n) = \lambda^{n-1}x_M(1) + \sum_{k=1}^{n-1} \lambda^{n-k-1}y_M(k).$$

Then the solution of difference equation $x_{M-1}(n+1) = \lambda x_{M-1}(n) + x_M(n) + y_{M-1}(n)$ is

$$x_{M-1}(n) = \lambda^{n-1}x_{M-1}(1) + \sum_{k=1}^{n-1} \lambda^{n-k-1} (x_M(k) + y_{M-1}(k)).$$

Hence, the solution of equation $x_i(n+1) = \lambda x_i(n) + x_{i+1}(n) + y_i(n)$ is

$$x_i(n) = \lambda^n x_i(1) + \sum_{k=1}^{n-1} \lambda^{n-k-1} (x_{i+1}(k) + y_i(k)).$$

Thus, we have found the solution of the difference equation (7). Now we turn our attention to the investigation of the case $|\lambda| = 1$, and our idea is to find the connection between one-dimensional and multidimensional equations in this case. At the end of this section, we will describe an interesting statement that, in fact, reveals this connection.

By induction, it is not difficult to deduce that the solution of (7) can be written in the following way:

$$x(n) = J_{\lambda}^{n-1} \cdot x(1) + \sum_{k=1}^{n-1} J_{\lambda}^{n-k-1} y(k).$$

Thus, for $1 \leq m \leq M$:

$$x_m(n) = \sum_{i=0}^{M-m} \left[\binom{n-1}{i} \lambda^{n-1-i} \cdot x_{i+m}(1) + \sum_{k=1}^{n-1} \binom{n-k-1}{i} \lambda^{n-k-1-i} \cdot y_{i+m}(k) \right].$$

We apply Lemma 1:

$$\binom{n-k-1}{i} = \sum_{r=0}^i (-1)^{i-r} \binom{n-1}{r} \frac{(k)^{[i-r]}}{(i-r)!}$$

and obtain

$$\begin{aligned} x_m(n) &= \sum_{i=0}^{M-m} \left[\binom{n-1}{i} \lambda^{n-1-i} \cdot x_{i+m}(1) + \sum_{k=1}^{n-1} \left(\sum_{r=0}^i (-1)^{i-r} \binom{n-1}{r} \frac{(k)^{[i-r]}}{(i-r)!} \right) \cdot \lambda^{n-k-1-i} \cdot y_{i+m}(k) \right] = \\ &= \sum_{i=0}^{M-m} \left[\binom{n-1}{i} \lambda^{n-1-i} \cdot x_{i+m}(1) + \sum_{r=0}^i \left\{ (-1)^{i-r} \binom{n-1}{r} \cdot \left(\sum_{k=1}^{n-1} \frac{(k)^{[i-r]}}{(i-r)!} \cdot \lambda^{n-k-1-i} \cdot y_{i+m}(k) \right) \right\} \right]. \end{aligned}$$

We change the order of summation:

$$\begin{aligned} &\sum_{i=0}^{M-m} \sum_{r=0}^i \left\{ (-1)^{i-r} \binom{n-1}{r} \cdot \left(\sum_{k=1}^{n-1} \frac{(k)^{[i-r]}}{(i-r)!} \cdot \lambda^{n-k-1-i} \cdot y_{i+m}(k) \right) \right\} = \\ &= \sum_{r=0}^{M-m} \sum_{i=r}^{M-m} \left\{ (-1)^{i-r} \binom{n-1}{r} \cdot \left(\sum_{k=1}^{n-1} \frac{(k)^{[i-r]}}{(i-r)!} \cdot \lambda^{n-k-1-i} \cdot y_{i+m}(k) \right) \right\}. \end{aligned}$$

For convenience, we swap the indices i and r :

$$\sum_{i=0}^{M-m} \sum_{r=i}^{M-m} \left\{ (-1)^{r-i} \binom{n-1}{i} \cdot \left(\sum_{k=1}^{n-1} \frac{(k)^{[r-i]}}{(r-i)!} \cdot \lambda^{n-k-1-r} \cdot y_{r+m}(k) \right) \right\}.$$

Therefore,

$$\begin{aligned} x_m(n) &= \sum_{i=0}^{M-m} \left[\binom{n-1}{i} \lambda^{n-1-i} \cdot x_{i+m}(1) + \sum_{r=i}^{M-m} \binom{n-1}{i} \left\{ (-1)^{r-i} \cdot \left(\sum_{k=1}^{n-1} \frac{(k)^{[r-i]}}{(r-i)!} \cdot \lambda^{n-k-1-r} \cdot y_{r+m}(k) \right) \right\} \right] = \\ &= \sum_{i=0}^{M-m} \binom{n-1}{i} \left[\lambda^{n-1-i} \cdot x_{i+m}(1) + \sum_{r=i}^{M-m} \left\{ (-1)^{r-i} \cdot \sum_{k=1}^{n-1} \left(\frac{(k)^{[r-i]}}{(r-i)!} \cdot \lambda^{n-k-1-r} \cdot y_{r+m}(k) \right) \right\} \right] = \\ &= \sum_{i=0}^{M-m} \binom{n-1}{i} \left[\lambda^{n-1-i} \cdot x_{i+m}(1) + \sum_{r=0}^{M-m-i} \left\{ (-1)^r \cdot \sum_{k=1}^{n-1} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{n-k-1-r-i} \cdot y_{r+i+m}(k) \right) \right\} \right]. \end{aligned}$$

Proposition 8. The solution of (7) can be written in a following way:

$$x_m(n) = \sum_{i=0}^{M-m} \binom{n-1}{i} \lambda^{n-1-i} \left[x_{i+m}(1) + \sum_{r=0}^{M-(i+m)} \left\{ (-1)^r \cdot \sum_{k=1}^{n-1} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+i+m}(k) \right) \right\} \right], \quad 1 \leq m \leq M.$$

The main theorem of this section follows.

Theorem 6. Let $M \geq 2$, $\{\tilde{y}_m(n) : n \geq 0\} \subset \mathbb{C}$, $1 \leq m \leq M$ and the following conditions hold:

- For all $1 \leq m \leq M$: $y_m(n) = \frac{\tilde{y}_m(n)}{n^{m-1}}$, $n \geq 1$
- For each $1 \leq m \leq M$ there exists some $\alpha_m \in \mathbb{C}$, such that the solution of the difference equation $z_{n+1} = \lambda z_n + \tilde{y}_m(n)$, $n \geq 1$, $z_1 = \alpha_m$ is bounded. Equivalently, the sequence of sums $\left\{ \sum_{n=1}^N \tilde{y}_m(n) \lambda^{-n} \right\}_{N \geq 1}$ is bounded.

Then the solution of the difference equation (7) is bounded if and only if $\forall m \in \{2, \dots, M\}$:

$$x_m(1) = - \sum_{r=0}^{M-m} \left\{ (-1)^r \cdot \sum_{k=1}^{\infty} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+m}(k) \right) \right\}.$$

Proof of The Main Theorem

Firstly, we would like to remark that each series

$$\sum_{k=1}^{\infty} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+m}(k) \right), \quad 2 \leq m \leq M, \quad 0 \leq r \leq M-m$$

is convergent by Dirichlet's test. Indeed,

$$\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+m}(k) = \frac{\lambda^{-r}}{r!} \cdot \frac{\tilde{y}_{r+m}(k) \lambda^{-k}}{\binom{k^{r+m-1}}{(k)^{[r]}}}. \quad (8)$$

The sequence $\left\{ \left(\frac{k^{r+m-1}}{(k)^{[r]}} \right)^{-1} : k \geq 1 \right\}$ is eventually monotone and is convergent to 0 because

$$\frac{k^{r+m-1}}{(k)^{[r]}} \sim \frac{k^{r+m-1}}{k^r} = k^{m-1}, \quad k \rightarrow \infty.$$

Also, by the assumption, the sequence of sums $\left\{ \sum_{k=1}^K \tilde{y}_m(k) \lambda^{-k} \right\}_{N \geq 1}$ is bounded.

Proof of sufficiency

We need the following Lemma:

Lemma 2. Let $\{a_n : n \geq 1\} \subset \mathbb{R}$ be monotone, convergent to 0 sequence. And let $\{b_n : n \geq 1\} \subset \mathbb{C}$ be a sequence, for which there exists a constant $C > 0$, such that $\forall N \geq 1 : \left| \sum_{n=1}^N b_n \right| \leq C$. Then $\left| \sum_{k=n}^{\infty} a_k b_k \right| \leq C |a_n|$ for all $n \geq 1$.

This lemma follows from Abel's summation by parts formula.

Now we use that $\forall m \in \{1, 2, \dots, m\} \exists C_m > 0 \forall N \geq 1 : \left| \sum_{n=1}^N \tilde{y}_m(n) \lambda^{-n} \right| \leq C_m$.

Therefore, by (8), for all $2 \leq m \leq M, 0 \leq r \leq M-m$ we have:

$$\left| \sum_{k=N}^{\infty} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+m}(k) \right) \right| \leq \frac{C_m}{r!} \cdot \frac{(N)^{[r]}}{N^{r+m-1}} \leq \frac{C_m}{r!} \cdot \frac{(N+r-1)^r}{N^{r+m-1}} = O\left(\frac{1}{N^{m-1}}\right), \quad N \rightarrow \infty.$$

Hence, since $\forall m \in \{2, \dots, M\}$:

$$x_m(1) = - \sum_{r=0}^{M-m} \left\{ (-1)^r \cdot \sum_{k=1}^{\infty} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+m}(k) \right) \right\}.$$

We deduce that for all $1 \leq m \leq M$ and for all $0 \leq i \leq M-m$ we have:

$$x_{i+m}(1) + \sum_{r=0}^{M-m-i} \left\{ (-1)^r \cdot \sum_{k=1}^{n-1} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+i+m}(k) \right) \right\} = O\left(\frac{1}{n^{i+m-1}}\right), \quad n \rightarrow \infty.$$

Since $\binom{n-1}{i} = O(n^i)$, $n \rightarrow \infty$ using Proposition 8 we obtain that the solution of (7) is indeed bounded.

Proof of necessity

Suppose that the solution $\{x(n) : n \geq 1\}$ of the equation (7) is bounded. In particular, the sequence $\{x_1(n) : n \geq 1\}$ is bounded.

$$x_1(n) = \sum_{i=0}^{M-1} A_i(n) \cdot \binom{n-1}{i},$$

where

$$A_i(n) := \lambda^{n-1-i} \left[x_{i+1}(1) + \sum_{r=0}^{M-m-i} \left\{ (-1)^r \cdot \sum_{k=1}^{n-1} \left(\frac{(k)^{[r]}}{r!} \cdot \lambda^{-k-r} \cdot y_{r+i+1}(k) \right) \right\} \right], \quad n \geq 1.$$

Then $A_{M-1}(n) \cdot \binom{n-1}{M-1} = O(n^{M-2})$, so $A_{M-1}(n) = O\left(\frac{1}{n}\right) \Rightarrow A_{M-1}(n) \rightarrow 0, n \rightarrow \infty$. Using the same argument from the proof of the sufficiency we deduce that $A_{M-1} = O\left(\frac{1}{n^{(M-1)+1-1}}\right) \Rightarrow A_{M-1}(n) \cdot \binom{n-1}{M-1} = O(1)$.

Thus $A_{M-2}(n) \cdot \binom{n-1}{M-2} = O(n^{M-3}) \Rightarrow A_{M-2}(n) \rightarrow 0, n \rightarrow \infty$ and analogously $A_{M-2}(n) \cdot \binom{n-1}{M-2} = O(1)$. Continuing this argument we show that $A_m(n) \rightarrow 0, n \rightarrow \infty$ for all $2 \leq m \leq M$.

Discussion and conclusions

In this work, we established novel, insightful sufficient conditions for the existence of a bounded solution of the first order linear difference equation in the one-dimensional case and further applied these conditions to the multidimensional case of this equation. The obtained results enrich the existing body of research in the field and complement previous studies.

Authors' contribution: Andrii Chaikovskiy — conceptualization; Oleksandr Liubimov — methodology; Oksana Lahoda — analysis of the results, preparation of literature review.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Chaikovskiy, A., & Lagoda, O. (2022). Bounded solutions of difference equations in a banach space with input data from subspaces. *Ukrainian Mathematical Journal*, 73, 1810–1824. <https://doi.org/10.1007/s11253-022-02031-3>
- Gomilko, A., Gorodnii, M., & Lagoda, O. (2003). On the boundedness of a recurrence sequence in a banach space. *Ukrainian Mathematical Journal*, 55(10), 1699–1708. <https://doi.org/10.1023/B:UKMA.0000022074.96704.7e>
- Gorodnii, M., & Kravets', V. (2020). On bounded solutions of one difference equation of the second order. *Journal of Mathematical Sciences*, 249, 601–608. <https://doi.org/10.1007/s10958-020-04960-5>
- Graham, S., & Kolesnik, G. (1991). *Van der corput's method of exponential sums*. Cambridge University press.
- Horodnii, M., & Lahoda, O. (2001). Bounded solutions for some classes of difference equations with operator coefficients. *Ukrainian Mathematical Journal*, 53(11), 1817–1824. <https://doi.org/10.1023/A:1015298712652>
- Mordell, J. L. (1958). On the kusmin-landau inequality for exponential sums. *Acta Arithmetica*, 4, 3–9.

Отримано редакцією журналу / Received: 05.02.25

Прорецензовано / Revised: 26.03.25

Схвалено до друку / Accepted: 10.10.25

Андрій ЧАЙКОВСЬКИЙ, д-р фіз.-мат. наук, проф.

ORCID ID: 0000-0002-3357-8744

e-mail: andrii chaikovskiy@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Олександр ЛЮБИМОВ, студ.

ORCID ID: 0009-0009-9347-6297

e-mail: liubimov_oleksandr@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Оксана ЛАГОДА, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0003-4515-1753

e-mail: lahoda.oksana@iik.kpi.ua

НТУУ "КПІ ім. Ігоря Сікорського", Київ, Україна

ОБМЕЖЕНІСТЬ РОЗВ'ЯЗКІВ ЛІНІЙНОГО РІЗНИЦЕВОГО РІВНЯННЯ ПЕРШОГО ПОРЯДКУ У БАГАТОВИМІРНОМУ ВИПАДКУ

Досліджено обмеженість розв'язків лінійного різницевого рівняння першого порядку вигляду $x_{n+1} = Ax_n + y_n$, $n \geq 1$, де A – комплексна квадратна матриця, послідовність $\{y_n\}_{n \geq 1}$ та початкове значення x_1 вважаються відомими. Розглянуто одновимірний випадок цього рівняння $x_{n+1} = ax_n + y_n$, $n \geq 1$, де a – комплексне число. Зокрема, отримано достатні умови, за яких розв'язки рівняння є обмеженими або необмеженими у випадку $|a| = 1$ (критичний випадок), розглядаючи експоненціальні суми вигляду $\sum y_n e(n\varphi)$ та $\sum e(f(n))$.

Потім здійснено дослідження рівняння у багатовимірному випадку та зведено нашу задачу до аналізу спектра та клітинок Жордана матриці A . Задача є особливо цікавою, коли спектр матриці A містить власні числа λ , $|\lambda| = 1$. У підсумку отримано теорему, яка показує зв'язок між рівняннями $x_{n+1} = ax_n + y_n$, $n \geq 1$, $|a| = 1$ та $x_{n+1} = Jx_n + y_n$, $n \geq 1$, де J – клітинка Жордана, що відповідає власному числу λ , $|\lambda| = 1$.

Ключові слова: лінійне різницеве рівняння, обмежений розв'язок, експоненціальні суми.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

UDC 004.8:62-5
DOI: <https://doi.org/10.17721/1812-5409.2025/2.24>

Maksym SHAMRAI, PhD Student
ORCID ID: 0000-0003-0730-494X
e-mail: m.shamrai@imath.kiev.ua
Institute of mathematics of NAS of Ukraine, Kyiv, Ukraine

NONASYMPTOTIC BOUNDS ON RETURN DEGRADATION FOR OBD-PRUNED NEURAL CONTROLLERS

Deep reinforcement learning (RL) has delivered striking results across domains ranging from games to robotics, yet the resulting controllers frequently comprise millions of parameters – far beyond the memory, latency, and energy budgets of embedded platforms such as quadrotors, mobile manipulators, and on-board microcontrollers. Pruning offers a practical path to deployment by removing parameters while preserving accuracy, but a fundamental question remains open for control: how much does pruning degrade closed-loop return? A theory is developed that links parameter-space perturbations produced by pruning to return degradation in a discounted MDP, without relying on global curvature of the training loss. The starting point is a tight, policy-level inequality: we show that the return gap $|J(\pi') - J(\pi)|$ is controlled by the statewise total-variation (TV) distance between the original and pruned policies. This TV-based bound follows directly from the performance-difference lemma and a bounded-advantage argument, and admits a KL variant via Pinsker's inequality. To connect this policy shift to the magnitude of pruning, we provide two complementary routes. First, at a locally optimal policy, a second-order Taylor expansion of the policy probabilities yields an OBD-style bound. Second, recognizing that a global Hessian is infeasible for modern models, we invoke a layer-wise robustness theorem for ReLU MLP controllers. Practically, the bound enables pre-pruning budgeting, post-pruning validation, and principled layer allocation. Conceptually, it bridges compression and safe policy improvement: the same TV/KL machinery that underlies trust-region methods now certifies pruning steps in deep RL. Overall, the results provide the first end-to-end, scalable framework to translate pruning actions into behavior-level guarantees for deep RL controllers, enabling reliable compression under tight on-board constraints.

Keywords: deep reinforcement learning, neural policies, optimal brain damage pruning, safety certificates, compression.

AMS 2020 classification: 68T07, 68T05, 93E20, 93B35.

Introduction

From mastering board and video games – Go (Silver et al., 2016), Atari (Mnih et al., 2015), Dota 2 (Berner et al., 2019), and StarCraft II (Vinyals et al., 2019) – to producing agile locomotion in simulation (MuJoCo) (Todorov, Erez, & Tassa, 2012) and dexterous visuomotor manipulation on real robots (Levine et al., 2016; Andrychowicz et al., 2020; Black et al., 2025), recent breakthroughs have been powered by deep reinforcement learning (RL). The policies behind these systems typically contain *hundreds of millions to billions* of parameters, a scale that accelerates learning and robustness but clashes with deployment realities: limited on-board memory, tight real-time latency, and strict energy budgets on platforms such as quadrotors, mobile manipulators, and autonomous vehicles. This mismatch makes model compression – especially parameter pruning – an indispensable step for bringing RL controllers out of the lab and into embedded, safety-critical settings. Second-order pruning methods such as Optimal Brain Damage (OBD) (LeCun, Denker, & Solla, 1989) and modern calibration-based approaches like SparseGPT (Frantar, & Alistarh, 2023) offer effective reductions in model size, yet their impact on *return* remains poorly understood. Motivated by this gap between empirical practice and provable guarantees, we develop bounds that translate parameter perturbations into certified limits on performance, building on the performance-difference framework (Kakade, & Langford, 2002; Schulman et al., 2015) and recent layer-wise robustness analyses (Shamrai, 2025).

The contributions of the paper are following: (i) *a unified performance–difference bound*: we start from the classical Kakade–Langford lemma and derive a tight upper bound on the return drop in terms of the *total-variation (TV)* distance between the original and pruned policies (Theorem 2); (ii) *Second-order link to OBD*: via a Taylor expansion we connect the TV distance to the *diagonal Hessian* scores used by OBD, producing the first explicit *return guarantee* that scales with the standard OBD metric (Corollary 2); (iii) *scalable, Hessian-free certificate*: to avoid a $d \times d$ global Hessian, we incorporate a recent layer-wise robustness theorem and obtain a *layer-local* bound linear in the ℓ_2 norm of the pruned weights (Corollary 3).

Our analysis bridges two previously disconnected areas: *the neural network compression*, which focuses on reducing model size but seldom reasons about downstream control metrics and the *safe RL*, which studies return-sensitive modifications but rarely addresses large-scale pruning (Gu et al., 2024). The resulting framework enables practitioners to prune deep RL policies *with performance certificates*, paving the way for resource-aware yet reliable autonomy.

1. Preliminaries

Fundamental notations are: *discount factor*: $\gamma \in (0, 1)$ controls the importance of future rewards; $\gamma \approx 1$ encourages long-term planning, whereas small γ focuses on immediate gains; *state & action spaces*: $\mathcal{S}$ and $\mathcal{A}$ are *finite* sets of states and actions; *dynamics and rewards*: $P(s' | s, a)$ is the state-transition kernel and the reward is bounded: $0 \leq r(s, a) \leq R_{\max} < \infty$ for all (s, a) ; *norms*: bold symbols (e.g. $\mathbf{x}$) denote vectors, $\|\cdot\|_1$ and $\|\cdot\|_2$ are the usual ℓ_1 (sum) and ℓ_2 (Euclidean) norms; *probability simplices*: $\Delta(\mathcal{A})$ denotes the set of all distributions over $\mathcal{A}$; *we treat probability distributions as column vectors whose entries sum to 1* (both norm and linear algebra notation apply seamlessly).

Definition 1 (Markov Decision Process). A *Markov Decision Process* (MDP) is the tuple $\langle \mathcal{S}, \mathcal{A}, P, r, \gamma, \rho \rangle$, where ρ is the distribution of the initial state s_0 .

Definition 2 (Policy). A *policy* is any measurable mapping $\pi : \mathcal{S} \rightarrow \Delta(\mathcal{A})$, $s \mapsto \pi(\cdot | s)$. We write $a \sim \pi(\cdot | s)$ when sampling an action from π in state s .

A *pruned policy* π' is obtained from a given neural policy π_θ by setting a subset of network parameters to zero using *Optimal Brain Damage* (OBD) (LeCun, Denker, & Solla, 1989). We use $\pi' = \pi_{\theta'}$ to emphasize the new parameters θ' .

Definition 3 (Value, action-value, and advantage). For any policy π and state-action pair (s, a) ,

$$V^\pi(s) := \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s \right], \quad Q^\pi(s, a) := \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, a_0 = a \right],$$

$$A^\pi(s, a) := Q^\pi(s, a) - V^\pi(s).$$

Intuition. $V^\pi(s)$ is the expected *return* starting from s when following π ; $Q^\pi(s, a)$ is the same, assuming we first force action a . The *advantage* A^π measures how much better (or worse) a is compared to the average prescribed by π in s .

Definition 4 (Discounted visitation distribution). The *discounted state visitation* under π is

$$d^\pi(s) := (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t \Pr(s_t = s \mid \pi, \rho),$$

which places more weight on states visited earlier in an episode.

Definition 5 (Expected return (performance)). Given an initial-state distribution ρ , the (discounted) expected return of policy π is

$$J(\pi) := \mathbb{E}_{s_0 \sim \rho} [V^\pi(s_0)] = \frac{1}{1 - \gamma} \mathbb{E}_{s \sim d^\pi, a \sim \pi(\cdot | s)} [r(s, a)],$$

where the second equality is a standard identity obtained by unrolling the definition of d^π in Def. 4. For a pruned policy π' , we write $J(\pi')$ analogously.

Definition 6 (Total variation and KL divergence). For distributions $p, q \in \Delta(\mathcal{A})$,

$$D_{\text{TV}}(p, q) := \frac{1}{2} \sum_{a \in \mathcal{A}} |p(a) - q(a)|, \quad D_{\text{KL}}(p \| q) := \sum_{a \in \mathcal{A}} p(a) \log \frac{p(a)}{q(a)}.$$

Lemma 1 (Performance–difference lemma (Kakade, & Langford, 2002)). For any two policies π, π' ,

$$J(\pi') - J(\pi) = \frac{1}{1 - \gamma} \mathbb{E}_{s \sim d^{\pi'}} \mathbb{E}_{a \sim \pi'} [A^\pi(s, a)].$$

We parameterize the policy by $\theta \in \mathbb{R}^d$ and write π_θ . *Pruning* removes parameters to reduce memory and inference cost. Formally, let $m \in \{0, 1\}^d$ be a binary *mask* and define the pruned parameters $\theta' := m \odot \theta$ (Hadamard product). The parameter perturbation is $\delta := \theta' - \theta = (m - \mathbf{1}) \odot \theta$, so that $\delta_i = -\theta_i$ for pruned indices and 0 otherwise. For layered networks we also use layerwise notation $\{W_\ell, b_\ell\}_{\ell=1}^L$ and masks M_ℓ so that $\widehat{W}_\ell := M_\ell \odot W_\ell$.

Optimal Brain Damage (OBD) selects weights to prune by minimizing a second-order approximation to the post-pruning loss. If $h_i \geq 0$ denotes the i th diagonal entry of the Hessian of the loss (or a suitable layer-local surrogate), the *saliency* of weight θ_i is $s_i := h_i \theta_i^2$. Under a global sparsity budget $\tau \in (0, 1)$ (or a target of k weights), OBD prunes the indices with the *smallest* s_i , yielding θ' that (approximately) minimizes the predicted loss increase subject to the budget. In the sequel we relate this parameter change δ (or layerwise $\|\delta W_k\|_2$) to changes in the *behavior* of the policy, quantified by $D_{\text{TV}}(\pi'(\cdot | s), \pi(\cdot | s))$ and ultimately by the return gap $|J(\pi') - J(\pi)|$.

To connect small parameter perturbations to changes in the controller's outputs, we rely on the following *deterministic* robustness result for single-layer pruning in feed-forward ReLU networks. It upper-bounds how a perturbation of layer k propagates to the policy output via a product of spectral norms of the *unpruned* layers.

Theorem 1 (Robustness of an OBD-pruned policy (Shamrai, 2025)). Let $\pi(\cdot; \Theta)$ be the L -layer MLP controller

$$x_0 = s, \quad x_\ell = \sigma(W_\ell x_{\ell-1} + b_\ell), \quad \ell = 1, \dots, L, \quad \Pi(s; \Theta) = x_L,$$

with ReLU-type activations σ_ℓ and weights $\Theta = \{W_\ell, b_\ell\}_{\ell=1}^L$. Suppose layer k is pruned, giving $\widehat{W}_k = W_k + \delta W_k$ and $\widehat{\Theta} = \Theta + \delta \Theta$. Then, for every input state $s \in X$,

$$\|\pi(s; \Theta) - \pi(s; \widehat{\Theta})\|_2 \leq \|\delta W_k\|_2 \left(\|s\|_2 \prod_{\substack{\ell=1 \\ \ell \neq k}}^L \|W_\ell\|_2 + \sum_{i=1}^{k-1} \prod_{\substack{\ell=i+1 \\ \ell \neq k}}^L \|W_\ell\|_2 \|b_i\|_2 \right) \leq \quad (1)$$

$$\leq \underbrace{\|\delta W_k\|_2 \left(\sup_{s \in X} \|s\|_2 \prod_{\substack{\ell=1 \\ \ell \neq k}}^L \|W_\ell\|_2 + \sum_{i=1}^{k-1} \prod_{\substack{\ell=i+1 \\ \ell \neq k}}^L \|W_\ell\|_2 \|b_i\|_2 \right)}_{=: C_{\max}}. \quad (2)$$

Consequently, the control-error bound in (1) can be summarized as $B_\pi(\delta\Theta) = C_{\max}\|\delta W_k\|_2$.

2. Results

Theorem 2 (Return difference via total variation). *In any finite MDP with $r \in [0, R_{\max}]$ and discount γ ,*

$$|J(\pi') - J(\pi)| \leq \frac{2R_{\max}}{(1-\gamma)^2} \mathbb{E}_{s \sim d^{\pi'}} \left[D_{\text{TV}}(\pi'(\cdot|s), \pi(\cdot|s)) \right].$$

Corollary 1 (KL-based version). *Combining Theorem 2 with Pinsker's inequality gives*

$$|J(\pi') - J(\pi)| \leq \frac{2R_{\max}}{(1-\gamma)^2} \mathbb{E}_{s \sim d^{\pi'}} \left[\sqrt{\frac{1}{2} D_{\text{KL}}(\pi'|\pi)} \right].$$

A simpler (conservative) worst-case bound is obtained by replacing the expectation with a supremum over s .

Corollary 2 (OBD performance bound). *Let π_{θ^*} be a locally optimal network policy with vanishing first derivatives, that is $\nabla_{\theta} \pi_{\theta^*}(a|s) = 0$. Prune weights $\mathcal{P} \subset [d]$ to obtain $\theta' = \theta^* + \delta$, $\delta_i = -\theta_i^*$ for $i \in \mathcal{P}$. If each $\theta \mapsto \pi_{\theta}(a|s)$ is C^3 with bounded third derivatives, then*

$$|J(\pi_{\theta'}) - J(\pi_{\theta^*})| \leq \frac{R_{\max}}{2(1-\gamma)^2} S_{\text{OBD}} + \mathcal{O}(\|\delta\|_2^3),$$

where $S_{\text{OBD}} := \sum_{i \in \mathcal{P}} \left[\sum_{(s,a)} h_i^{(a,s)} \right] \theta_i^{*2}$ is the sum of diagonal Hessian scores used by OBD (LeCun, Denker, & Solla, 1989).

Intuition. S_{OBD} is precisely what OBD minimises; the corollary states that, for small enough pruning, the drop in return is linear in that score.

The diagonal entries $h_i^{(a,s)}$ in Corollary 2 come from the global Hessian $H_{a,s} \in \mathbb{R}^{d \times d}$, where $d \approx 10^8$ for contemporary vision or language policies. Even storing $H_{a,s}$ is infeasible, let alone computing it for every (s, a) . SparseGPT (Frantar, & Alistarh, 2023) circumvents this by optimising a simple quadratic objective $\frac{1}{2} \|XW - Y\|_F^2$ layer by layer, whose Hessian $X^\top X$ is small and easy to invert. Thus curvature is captured locally while memory scales linearly in the number of parameters.

Corollary 3 (Return bound via layer-local TV (single pruned layer)). *Assume the setting of Theorem 1 and that the policy head mapping the final network output to action probabilities is 1-Lipschitz in ℓ_2 (e.g., a softmax over logits) so that $\pi(\cdot|s), \pi'(\cdot|s) \in \Delta(\mathcal{A})$ with $A := |\mathcal{A}| < \infty$. Then*

$$|J(\pi') - J(\pi)| \leq \frac{R_{\max} \sqrt{A} C_{\max}}{(1-\gamma)^2} \|\delta W_k\|_2.$$

The corollary leverages the layer-wise output perturbation control from Theorem 1 to upper bound the *statewise* policy shift in total variation, and then propagates this shift to a guaranteed return bound via Theorem 2. The certificate is: (i) *linear* in the pruning magnitude $\|\delta W_k\|_2$; (ii) *dimension-free* with respect to the total parameter count; and (iii) dependent on the action-space size only through $\sqrt{A}$. A less conservative, distributional variant replaces $\sup_s$ by $\mathbb{E}_{s \sim d^{\pi'}}$, yielding the same scaling but with $C_{\max}$ multiplied by an average instead of a worst-case factor. For multiple pruned layers, a triangle-inequality extension gives $\|\pi' - \pi\|_2 \leq \sum_{k \in \mathcal{K}} C_{\max}^{(k)} \|\delta W_k\|_2$, leading to an additive bound on the return gap.

Discussions and conclusions

We studied the effect of pruning on the *behavior* of deep RL policies, providing performance guarantees that are both interpretable and scalable. Our analysis starts from a policy-level quantity and ends with a certified bound on the return gap, avoiding any direct reasoning about nonconvex training losses. We derived a TV-based performance bound (Theorem 2) that upper-bounds $|J(\pi') - J(\pi)|$ through the state-wise policy shift $D_{\text{TV}}(\pi', \pi)$, with a KL variant via Pinsker (Corollary 1), connected TV to second-order parameter perturbations at an optimal policy, giving an OBD-style certificate (Corollary 2), eliminated the intractable global Hessian by invoking a layer-wise robustness result (Theorem 1), which yields a tight, *layer-local* estimate of probability shift; combined with Theorem 2, this produces a practical, dimension-free return bound (Corollary 3). Together, these results form a pipeline

$$\text{pruning magnitude } (\|\delta W_k\|_2) \Rightarrow D_{\text{TV}}(\pi', \pi) \Rightarrow |J(\pi') - J(\pi)|$$

that scales to modern policies without global curvature estimation.

The bound depends on (a) the pruning magnitude $\|\delta W_k\|_2$, (b) a computable robustness constant $C_{\max}$ (product of layer spectral norms and input bounds), and (c) the action-space factor $\sqrt{A}$. Each quantity can be estimated once per model (via power iteration and a calibration bound on $\|s\|_2$), enabling *pre-pruning* budgeting and *post-pruning* validation with negligible overhead. The certificate is additive across pruned layers by triangle inequality, which supports iterative or structured pruning schedules.

Our guarantees are worst-case and may be conservative if $C_{\max}$ is dominated by a small number of large spectral norms or by a loose input bound $\sup_{s \in \mathcal{X}} \|s\|_2$. The $\sqrt{A}$ dependence is inevitable for $\ell_2 \rightarrow \ell_1$ conversion, but can overstate degradation in highly peaked policies. Finally, the theory assumes a 1-Lipschitz policy head (e.g., softmax or bounded continuous control) and does not exploit environment-specific mixing or structure beyond bounded rewards.

Acknowledgments. The author expresses his gratitude to the Armed Forces of Ukraine for their protection, which has made this research possible.

Sources of funding. The author acknowledges financial support by the Simons Foundation grant (SFI-PD-Ukraine-00014586, M.S.) and the project 0125U000299 of the National Academy of Sciences of Ukraine.

References

- Andrychowicz, O. M., Baker, B., Chociej, M., Józefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., Schneider, J., Sidor, S., Tobin, J., Welinder, P., Weng, L., & Zaremba, W. (2020). Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1), 3–20. <https://doi.org/10.1177/0278364919887447>
- Berner, C., Brockman, G., Chan, B., Cheung, V., Debiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., de Oliveira Pinto, H. P., Raiman, J., Salimans, T., Schlatter, J., ... Zhang, S. (2019). *Dota 2 with large scale deep reinforcement learning*. <https://doi.org/10.48550/arXiv.1912.06680>
- Black, K., Brown, N., Darpinian, J., Dhabalia, K., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., Galliker, M. Y., Ghosh, D., Groom, L., Hausman, K., Ichter, B., Jakubczak, S., Jones, T., Ke, L., LeBlanc, C., Levine, S., ... Zhilinsky, U. (2025). $\pi_{0.5}$: a vision-language-action model with open-world generalization. <https://doi.org/10.48550/arXiv.2504.16054>
- Frantar, E., & Alistarh, D. (2023). Sparsegpt: massive language models can be accurately pruned in one-shot. In *Proceedings of the 40th International Conference on Machine Learning*. JMLR.org. <https://dl.acm.org/doi/10.5555/3618408.3618822>
- Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J., & Knoll, A. (2024). A review of safe reinforcement learning: Methods, theories, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 11216–11235. <https://doi.org/10.1109/TPAMI.2024.3457538>
- Kakade, S., & Langford, J. (2002). Approximately optimal approximate reinforcement learning. In *Proceedings of the 19th International Conference on Machine Learning* (p. 267–274). Morgan Kaufmann Publishers Inc. <https://dl.acm.org/doi/10.5555/645531.656005>
- LeCun, Y., Denker, J., & Solla, S. (1989). Optimal brain damage. In Touretzky (Ed.), *Advances in Neural Information Processing Systems* (Vol. 2). Morgan-Kaufmann. <https://shorturl.at/Sr2Ve>
- Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(1), 1334–1373. <https://dl.acm.org/doi/10.5555/2946645.2946684>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Hiedmiller, M. A., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust region policy optimization. In Bach & Blei (Eds.), *Proceedings of the 32nd International Conference on Machine Learning* (Vol. 37, pp. 1889–1897). PMLR. <https://proceedings.mlr.press/v37/schulman15.html>
- Shamrai, M. (2025). Closed-form robustness bounds for second-order pruning of neural controller policies. <https://doi.org/10.48550/arXiv.2507.02953>
- Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <https://doi.org/10.1038/nature16961>
- Todorov, E., Erez, T., & Tassa, Y. (2012). Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (p. 5026–5033). <https://doi.org/10.1109/IROS.2012.6386109>
- Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A. J., Chung, J., Choi, D., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., ... Silver, D. (2019). Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782), 350–354. <https://doi.org/10.1038/s41586-019-1724-z>

Отримано редакцією журналу / Received: 13.05.25

Прорецензовано / Revised: 24.09.25

Схвалено до друку / Accepted: 10.10.25

Максим ШАМРАЙ, асп.

ORCID ID: 0000-0003-0730-494X

e-mail: m.shamrai@imath.kiev.ua

Інститут математики НАН України, Київ, Україна

НЕАСИМПТОТИЧНІ МЕЖІ ПОГІРШЕННЯ ДИСКОНТОВАНОЇ КУМУЛЯТИВНОЇ ВИНАГОРОДИ ДЛЯ ОВД-ПРОРІДЖЕНИХ НЕЙРОННИХ КОНТРОЛЕРІВ

Останнім часом глибоке навчання з підкріпленням (RL) продемонструвало приголомшливі результати в галузях від ігор до робототехніки, однак отримані контролери часто містять мільйони параметрів – значно більше за обмеження пам'яті, затримки й енергоспоживання вбудованих платформ на кшталт квадрокоптерів, мобільних маніпуляторів і бортових мікроконтролерів. Прорідження (pruning) пропонує практичний шлях до розгортання, вилучаючи параметри без втрати точності, але для систем керування лишається відкрите фундаментальне запитання: наскільки прорідження погіршує замкнену (closed-loop) винагороду? У пропонуваній роботі розроблено теорію, що пов'язує збурення у просторі параметрів, спричинені прорідженням, із погіршенням винагороди в дисконтованій МППР (MDP), не покладаючись на глобальну кривизну функції втрат під час навчання. Відправною точкою є «щільна» нерівність на рівні політики: показано, що розрив у винагороді $|J(\pi') - J(\pi)|$ визначається варіаційною (TV) відстанню між початковою та прорідженою політиками для кожного стану. Ця межа на основі TV безпосередньо впливає з леми про різницю продуктивності й оцінки обмеженої переваги, а також має KL-варіант через нерівність Пінскера. Щоб пов'язати цю зміну політики з величиною прорідження, ми пропонуємо два комплементарні шляхи. По-перше, у разі локально оптимальної політики розклад Тейлора другого порядку для ймовірностей політики дає межу у стилі OBD. По-друге, визнаючи, що глобальний гесіан є непрактичним для сучасних моделей, ми залуцаємо теорему про поширену робастність для контролерів ReLU MLP. На практиці запропонована межа уможливує планування бюджету перед прорідженням, перевірку після прорідження та принциповий розподіл ступеня прорідження між шарами. На концептуальному рівні вона поєднує компресію та безпечне поліпшення політики: той самий апарат TV/KL, що лежить в основі trust-region методів, тепер надає сертифікацію кроків прорідження у глибокому RL. Загалом отримані результати пропонують першу наскрізну, масштабовану базу, що перетворює дії з прорідження на гарантії на рівні поведінки для глибоких RL-контролерів, забезпечуючи надійну компресію за жорстких бортових обмежень.

Ключові слова: глибоке навчання з підкріпленням, нейронні політики, прорідження optimal brain damage, сертифікати безпеки, стиснення.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження(у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

КОМП'ЮТЕРНІ НАУКИ ТА ІНФОРМАТИКА

UDC 51.681.3

DOI: <https://doi.org/10.17721/1812-5409.2025/2.25>

Igor ZAVADSKYI, DSc (Phys&Math), Prof.

ORCID ID: 0000-0002-4826-5265

e-mail: ihorzavadskyi@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Maksym KOVALCHUK, PhD Student

ORCID ID: 0009-0002-8450-905X

e-mail: max.koval4uk@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

OPTIMIZATION OF THE CANONICAL HUFFMAN CODES FOR WORD-BASED NATURAL LANGUAGE TEXT COMPRESSION

Canonical Huffman codes are widely used in data compression and allow fast decoding without compromising compression efficiency compared to classical Huffman codes. If the maximal length of a codeword does not exceed 11-12 bits or long codewords are very rare, any variable-length code, including Huffman codes, can be effectively decoded using a naive 'table' method. Canonical Huffman codes should be used primarily if the encoded text consists of a relatively large number of long codewords. For example, this is the case of word-level natural language text compression, where words are the symbols of an alphabet. A. Moffat and A. Turpin proposed the fastest state-of-the-art algorithm for decoding canonical codes. We offer several optimizations of this algorithm, which, according to our experiments, can speed it up by 20-30% without reducing the compression ratio. Some of these optimizations are based on low-level code parallelization using the SIMD architecture of modern processors, while others propose combining classical decoding of canonical codes with a universal 'table' decoding method. In addition to the above optimizations, the article discusses in detail the canonical Huffman codes, including the basic and optimized algorithms, as well as the universal 'table' method for decoding arbitrary variable-length codes. The results of an experimental comparison of the performance of the basic and optimized algorithms are presented. The results of the study can be used in information retrieval and data archiving systems.

Keywords: Huffman, encoding, codes, canonical, text compression.

AMS 2020 classification: 68P30, 94A29, 94-06.

Introduction

Relevance of research. Huffman codes (Huffman, 1952) are a keystone of lossless data compression. In semi-static statistical compression, for a given data source, they provide an optimal compression ratio if symbol probabilities are negative powers of 2 and a near-optimal ratio in many practical use cases. However, the default decoding algorithm for Huffman codes requires bit-by-bit processing of the encoded text with the corresponding traversal of a Huffman tree. This approach suffers from low decoding speed, which is a crucial characteristic in many practical data compression scenarios. In the last two decades, many other codes that trade the decoding speed for compression ratio have been proposed, e.g., Byte-Aligned codes (Brisaboa et al., 2003), Reverse Multi-Delimiter Codes (Anisimov, & Zavadskyi, 2017), Binary Mixed-Digit Codes (Zavadskyi, & Kovalchuk, 2023). Also, there are known solutions that can improve the Huffman decoding throughput without compromising the compression efficiency. Among them, one of the simplest and most widely used is the Canonical Huffman Codes (CHC), first proposed in (Schwartz, & Kallick, 1964).

The research objects are Canonical Huffman Codes and their fast decoding algorithms.

The aim of this research is to optimize the state-of-the-art canonical Huffman code decoding algorithm proposed in (Moffat, & Turpin, 1997), specifically for cases where the encoded text contains a large number of long codewords.

The structure of the paper is as follows. In Section 1, we describe the idea of the CHC and different known up-to-date approaches to their fast decoding, as well as the more general but also efficient 'table' method of a variable-length code fast decoding. In Section 2, we discuss our main contribution - several optimizations to Algorithm TABLE-LOOKUP from (Moffat, & Turpin, 1997), which we consider as the baseline. In Section 3, we provide the results of experiments, where we measured the execution time of different CHC decoding methods.

1. Canonical Huffman Codes

Before we discuss the CHC, let us note that there exists a universal and simple 'table' method of decoding classical Huffman codes, as well as any variable length codes. Let $maxLen$ be the bitlength of the longest codeword. Then, we use lookup tables D and L that are indexed by all possible binary $maxLen$ -bit vectors that start from a codeword. We call such bit vectors *buffer vectors*. Assume the bit representation of some buffer vector x starts from the codeword c . Then, the $D[x]$ is the result of the decoding of c (we also call it the result of the decoding of x), while $L[x]$ is the bitlength of c . The decoding process becomes almost trivial: at a given bitstream position, we read the $maxLen$ -bit value x , output the result of the decoding $D[x]$ and shift the read position by $L[x]$ bits forward. This method can hardly be outperformed if $maxLen$ does not exceed 11 or 12 bits. However, each of these tables consists of about 2^{maxLen} elements, and when the maximum codeword length becomes bigger, tables do not fit into the processor cache, which slows down the decoding dramatically.

Therefore, on short alphabets, such as ASCII characters, DNA nucleotides or protein bases, the fast decoding task can be considered successfully solved by the 'table' method described above. However, compression of data on larger alphabets is no less important task, e.g. compression of word-based natural language texts. As experiments show, the

semi-static compression of a word-based English text can be almost twice as efficient as a character-based text. This is why word-based natural language text compression is the main subject of interest to us.

The Canonical Huffman Codes allow us to avoid using both the large lookup tables and, to some extent, bit-by-bit code processing. They assume using the dictionary model of the encoded text. All source alphabet symbols reside in the dictionary – the array sorted by the descending symbol frequencies. During the encoding, we find the text symbol in the dictionary and encode its index. The decoding is reverse: we decode the index and then use it to get the symbol from the dictionary. Therefore, bigger encoded numbers correspond to less frequent symbols.

For a given source symbol distribution, codewords of the CHC have the same lengths as the classical Huffman codes, and thus, the compression ratios of these code classes are the same. However, the most important property of the canonical codes is that the codewords of the same length form a contiguous set of integers. Let $baseCwd[i]$ be the smallest value of the codeword of length i and $baseSym[i]$ be the result of its decoding. Then, if we know that the buffer vector starts from the codeword c of length i , the result of its decoding can be calculated as $baseSym[i] + c - baseCwd[i]$. The codeword c can be obtained by shifting the buffer by $maxLen - i$ bits to the right. The only problem that remains is to get the codeword length given the buffer vector. In the original canonical codes, the desired length is searched linearly, i.e., the length l is incremented, starting from the minimal possible value, until the buffer value is greater or equal to the threshold $Limit[l]$, which is the smallest buffer value that starts from the smallest codeword of length l . This search can be faster than traversing the Huffman tree; nonetheless, it remains too slow. In (Moffat, & Turpin, 1997), the authors propose to get the initial length value from the lookup table $Start$ of limited size. They use b leftmost bits from the buffer as the index V_b of that table. $Start[V_b]$ is the length of the shortest codeword c , which is consistent with V_b . The word 'consistent' here means that either some prefix of V_b coincides with c or some prefix of c coincides with V_b . Then, the obtained length value l is incremented until the buffer value is no less than the threshold $Limit[l]$.

The detailed algorithm is given below.

We process the input bit array E , which is indexed by bit positions. The variable pos holds the current read position. After loading bits from E to $buffer$, we address $buffer$ both as a variable (line 5) and a bit array (lines 4 and 7). Two more 'mask and shift' operations are needed to get prefixes of the $buffer$ bit array in lines 4 and 7.

input : Encoded bit sequence $E[1..n]$;
Lookup tables $Start$, $Limit$, $baseSym$, $baseCwd$ defined above;
Dictionary D ;
Maximal codeword length $maxLen$;
Length of the start lookup bit vector b .

```

1  pos ← 1;
2  while pos < n do
3      buffer ← E[pos..pos + maxLen - 1];
4      l ← Start[buffer[1..b]];
5      while buffer ≥ Limit[l] do
6          l ← l + 1;
7      end
8      codeword ← buffer[1..l];
9      symbol ← baseSym[l] + codeword - baseCwd[l];
10     Output(D[symbol]);
11     pos ← pos + l;
12 end

```

Algorithm 1: Decoding of the canonical codes - baseline algorithm

In (Liddell, & Moffat, 2006), an extension of Algorithm 1 was proposed (Algorithm 3 in the cited paper). The method combines a limited 'table' decoding procedure with the canonical Huffman decoder: p bits are first read from the bitstream, and if they already contain one or more complete codewords, decoding is performed using the 'table' method; otherwise, the canonical decoder is applied. Following authors, we denote this approach as **Ext- p** , where p is the parameter of the 'table' method, while the unmodified canonical decoder is referred to as the **Baseline**. However, the experimental results given in Tab. 1 demonstrate that this extension increases the decoding time compared to the baseline implementation when applied to texts on word-based alphabets, though it can be efficient for short character-based alphabets (the experimental environment is described in Section 3). A similar behavior was reported in (Liddell, & Moffat, 2006). For each text, we hereby take the best decoding time presented in Tab. 1 as the state-of-the-art baseline.

Table 1
Decoding time of the baseline canonical decoder and its extension (Ext- p) in seconds per 1000 iterations

File name	Baseline	Ext-6	Ext-8	Ext-10	Ext-11	Ext-12	Ext-13	Ext-14	Ext-15	Ext-20
Alice29	0.136	0.179	0.197	0.170	0.147	0.150	0.153	0.162	0.228	0.450
Harry Potter	0.385	0.482	0.550	0.565	0.503	0.506	0.506	0.527	0.764	2.665
Bible	3.738	5.934	5.515	4.760	4.219	4.194	4.139	4.226	5.769	16.918
Shakespeare	4.322	5.970	6.338	6.580	6.394	6.343	6.418	6.568	9.268	33.384
English.50MB	50.47	66.22	68.38	67.37	64.50	66.47	68.77	69.96	95.88	254.41
Shakespeare - char	2.671	2.297	1.593	1.165	1.184	1.261	1.416	1.574	2.292	5.060

Experiments on real-world word-based text compression show that, for $p = 12$, the loop in lines 5 – 6 never has more

than 1 iteration and most often is not executed at all. The loop was never entered for files of about 0.5MB; on the 4 – 5MB files, the probability of one iteration execution is less than 0.1%. At the same time, the size of the *Start* lookup table is $2^p = 2^{12}$, which is much less than $2^{maxLen} \in [2^{17}..2^{22}]$ for the tables of the default method discussed in the Introduction, and acceptable to fit into L1 processor cache. Thus, the Algorithm 1 can be considered a fast and memory-efficient decoding method with $O(1)$ per symbol time complexity. Nonetheless, we will discuss a number of further improvements that can accelerate this method significantly.

2. Baseline algorithm optimizations

1. In Algorithm 1, arrays *baseSym* and *baseCwd* are used in line 8 only, both with the same index, in the expression $baseSym[l] - baseCwd[l]$. Thus, we can replace both these arrays with one array *Base*, $Base[i] = baseSym[i] - baseCwd[i]$. This simple replacement saves one subtraction and one loading from memory per decoded symbol.
2. It would be better to get the decoded value itself from the *Start* table instead of the codeword length. However, if the alphabet size is bigger than 2^b , it is impossible to 'pack' all decoded values into the 2^b -element array. Nonetheless, some prefixes of the *buffer*[1..*b*] bitstring may coincide with a whole codeword. For word-based natural language text compression, the probability of this event is about 0.7 – 0.95, dependent on the value *p* and the type of text. This implies we can combine the 'table' method with the canonical codes if we store in the special array *Ncwd* the number of whole codewords in the bitstring *buffer*[1..*p*], where *p* is some predefined parameter; typically $p \leq b$. Then, if $Ncwd[buffer[1..p]] \geq 1$, we can extract one or two decoded numbers from *buffer*[1..*p*] and proceed to the next iteration; otherwise, we decode the canonical codeword as in Algorithm 1.

Let us note that the idea of retrieving several decoded results at each iteration of the decoded loop has repeatedly appeared, e.g., in (Choueka, Klein, & Perl, 1985) or (Milidiu et al., 2003). In application to CHC, it was investigated in detail in (Liddell, & Moffat, 2006). However, as we mentioned in 1, it appears not to be efficient in the practical decoding of texts on word-based alphabets. We claim that it is due to implementation shortcomings. The condition $Ncwd[buffer[1..p]] \geq 1$, which is checked in line 14 of Algorithm 3 in (Liddell, & Moffat, 2006), can be true or false with relatively high probability. Thus, the 'if' statement that checks it is unpredictable, and it slows down the program significantly. To avoid it, we exploit the advantage of the 64-bit architecture of modern processors, using the approach first implemented in the Binary Coded Ternary data compression code (Zavadskyi, 2022). Namely, we always output two 32-bit decoded symbols packed into a 64-bit word, even if the bitstring *buffer*[1..*p*] contains one or zero full codewords. Then, we increase the counter of output symbols by $Ncwd[buffer[1..p]]$. Therefore, if $Ncwd[buffer[1..p]] \leq 1$, we output one or two fictitious numbers, and at the next iteration, the actual numbers will be output to the same memory.

Both optimizations 1 and 2 are given in Algorithm 2. Apart from the lookup tables defined above, to shift the read position, the table *Shift* in line 8 is used to store the total length of codewords that coincide with the prefix of *buffer*[1..*p*]. Also, we explicitly show using the output array as the logic of writing values to it is important.

input : Encoded bit sequence $E[1..n]$;
 Lookup tables *Start*, *Limit*, *Base*, *Ncwd*, *Result*, *Shift* defined above;
 Dictionary *D*;
 Maximal codeword length – *maxLen*;
 Length of the start lookup bit vector – *b*;
 Length of the *buffer* prefix for table decoding – *p*.

output: Array of numbers *Out*.

```

1 pos ← 1;
2 outPos ← 1;
3 while pos < n do
4     buffer ← E[pos..pos + maxLen - 1];
5     prefix ← buffer[1..p];
6     Out[outPos, outPos + 1] ← Result[prefix];
7     outPos ← outPos + Ncwd[prefix];
8     pos ← pos + Shift[prefix];
9     l ← Start[buffer[1..b]];
10    while buffer ≥ Limit[l] do
11        | l ← l + 1;
12    end
13    codeword ← buffer[1..l];
14    Out[outPos] ← D[Base[l] + codeword];
15    outPos ← outPos + 1;
16    pos ← pos + l;
17 end

```

Algorithm 2: Hybrid table and canonical decoding

3. In Algorithms 1 and 2, we omitted the process of loading bits from the encoded bitstream *E* for simplicity. However, in the case of Canonical Huffman Codes, the decoding of individual symbols is already highly optimized, making the processing of the bitstream a significant contributor to the overall execution time. There are two baseline ways to load the chunk of the bitstream from memory to the variable *buffer* (line 3).

- (a) Byte-based buffer management. The bitstream is stored as a sequence of bytes. We need to divide pos by 8, load 8 bytes from memory to a *buffer* processor register, and perform a shift operation to remove up to 7 previously decoded bits. However, the applicability of this method depends on the machine's endianness and whether the leftmost bit is stored as the least or most significant bit of the integer. Due to these dependencies, this approach is unsuitable for CHC decoding.
- (b) Conditional buffer refilling. The buffer is always maintained with at least 32 valid bits. During decoding, the 'if' statement checks if fewer than 32 bits remain in the buffer. If so, a 32-bit integer is loaded into the buffer. This method is independent of endianness but introduces overhead due to frequent unpredictable conditional checks.

Both of these techniques refill the buffer for each decoded symbol. Now, we propose an optimization that minimizes the overhead associated with processing the bitstream E . Instead of refilling the buffer at every decoding iteration, we update it only once every $k = \lfloor \frac{64}{maxLen} \rfloor$ decoded symbols. Together with Optimization 1, this is presented in Algorithm 3. Each macro-iteration of the decoding proceeds as follows:

First, 64 bits are loaded into the buffer in line 5. In practice, the bitstream E is stored as a sequence of contiguous 64-bit integers. During the buffer update, the *buffer* is updated with two adjacent 64-bit integers from E using bitwise OR operation. If necessary, the first 64-bit integer is shifted left to update the most significant bits of the *buffer*. Then, the 64-bit integer is shifted right to update the least significant bits of the *buffer*.

In lines 6–14, k iterations of the standard decoding are performed. Each iteration processes l bits and finishes with line 14, where l bits are removed from the top of the buffer.

```

input : Encoded bit sequence  $E[1..n]$ ;
        Lookup tables  $Start, Limit, Base$  defined above;
        Dictionary  $D$ ;
        Maximal codeword length  $maxLen$ ;
        Length of the start lookup bit vector  $b$ .

output: Array of numbers  $Out$ .

1  $pos \leftarrow 1$ ;
2  $outPos \leftarrow 1$ ;
3  $k \leftarrow \lfloor \frac{64}{maxLen} \rfloor$ ;
4 while  $pos < n$  do
5    $buffer \leftarrow E[pos..pos + 63]$ ;
   for  $i \leftarrow 1$  to  $k$  do
6      $prefix \leftarrow buffer[1..b]$ ;
7      $l \leftarrow Start[prefix]$ ;
8     while  $buffer \geq Limit[l]$  do
9        $l \leftarrow l + 1$ ;
10    end
11     $codeword \leftarrow buffer[1..l]$ ;
12     $Out[outPos] \leftarrow D[Base[l] + codeword]$ ;
13     $outPos \leftarrow outPos + 1$ ;
14     $pos \leftarrow pos + l$ ;
     $buffer \leftarrow shiftLeft(buffer, l)$ ;
   end
end

```

Algorithm 3: Canonical decoding with fast buffer updates

4. Lines 7 and 8 of Algorithm 1 involve a shift operation to retrieve *codeword* from *buffer*, lookup into *Base* table (considering Optimization 1), and addition to get decoded symbol. The addition operation depends on the result of the preceding shift, potentially introducing latency. To optimize this, we propose moving all of the mentioned operations outside the main decoding loop. Both optimizations 1 and 4 are detailed in Algorithm 4. Instead of processing each symbol individually, the values of *buffer* and lengths l are stored in temporary arrays *tmpBuffer* and *tmpL* during the decoding process (lines 10 and 11). Once enough symbols (e.g. 512) are processed this way, these values are decoded in batches using SIMD instructions to perform the shifts, lookups, and additions simultaneously (lines 13–17). Experiments demonstrate execution time improvements on processors supporting 256-bit SIMD operations, such as those with AVX and AVX2 instruction sets. In lines 13–18, variables ending with 8 represent SIMD registers, each holding 8 32-bit integers. The *simd_load8* operation loads 8 consecutive integers from memory into a SIMD register. The *simd_subtract8* operation subtracts corresponding integers in one SIMD register from another. The *simd_shiftRight8* operation shifts each integer in one register to the right by the number of bits specified in the corresponding integer of another register. The *simd_lookup8* operation fetches 8 values simultaneously from the *Base* table, using integers in a SIMD register as indices. Finally, the *simd_add8* operation adds corresponding integers from two registers element-wise. These operations allow efficient parallel processing on 8 integers at once. Let us note that Optimization 2 is incompatible with Optimization 4. On processors limited to 128-bit SIMD without the required shift and table-lookup instructions, this optimization may actually degrade performance due to overhead. Conversely, on newer architectures with 512-bit SIMD (e.g., AVX-512), we expect higher speedup.

```

input : Encoded bit sequence  $E[1..n]$ ;
        Lookup tables  $Start$ ,  $Limit$ ,  $Base$  defined above;
        Dictionary  $D$ ;
        Maximal codeword length  $maxLen$ ;
        Length of the start lookup bit vector  $b$ .

output: Array of numbers  $Out$ .

1  $pos \leftarrow 1$ ;
2  $outPos \leftarrow 1$ ;
3  $maxLen8 \leftarrow simd\_fill8(maxLen)$ ;
4 while  $pos < n$  do
    for  $i \leftarrow 1$  to 512 do
6          $buffer \leftarrow E[pos..pos + maxLen - 1]$ ;
7          $prefix \leftarrow buffer[1..b]$ ;
8          $l \leftarrow Start[prefix]$ ;
9         while  $buffer \geq Limit[l]$  do
             $l \leftarrow l + 1$ ;
        end
10         $tmpBuffer[i] \leftarrow buffer$ ;
11         $tmpL[i] \leftarrow l$ ;
12         $pos \leftarrow pos + l$ ;
    end
    for  $i \leftarrow 1$  to 512 by 8 do
13         $buffer8 \leftarrow simd\_load8(tmpBuffer[i..i + 7])$ ;
14         $l8 \leftarrow simd\_load8(tmpL[i..i + 7])$ ;
15         $codeword8 \leftarrow simd\_shiftRight8(buffer8, simd\_subtract8(maxLen8 - l8))$ ;
16         $base8 \leftarrow simd\_lookup8(Base, l8)$ ;
17         $symbol8 \leftarrow simd\_add8(codeword8, base8)$ ;
        for  $j \leftarrow 0$  to 7 do
18             $Out[outPos] \leftarrow D[symbol8[j]]$ ;
19             $outPos \leftarrow outPos + 1$ ;
        end
    end
end

```

Algorithm 4: Canonical decoding with SIMD optimization. 256-bit SIMD operations are assumed.

3. Experimental results

We have implemented both the baseline algorithm and its optimizations in C++ language and compiled them with the Visual Studio compiler¹. We conducted testing on a Windows OS using an Intel i7-7700HQ processor (2.80 GHz) with 32 KB of L1 data cache per core and 16 GB of DDR4 RAM. Algorithms were tested by decoding five English texts encoded with the Canonical Huffman Codes on word-based alphabets. Text parameters are listed in Tab. 2.

The optimal value of the parameter b in all algorithms has been determined experimentally as $b = 12$, while the optimal value of the parameter p in Algorithm 2 has been determined as $p = 10$.

In Tab. 3, we give the total time of 1000 runs of different algorithms and their combinations. We did not evaluate the efficiency of optimizations 2, 3, and 4 without optimization 1, since the latter is straightforward and does not adversely affect any of the other improvements. To compare the productivity fairly, we did not measure the time of common text postprocessing operations, such as converting integers to strings or concatenating words to form the resultant text in memory.

Table 2

Text parameters

File	Size	Total symbols	Different symbols	Entropy H0 (bytes)	Encoded size	Average codeword length (bits)
Alice's Adventures in Wonderland by L. Carrol	145KB	26,446	3241	29,747	29,860	9.03
Harry Potter and the Philosopher's Stone	427KB	78,419	7067	95,450	95,792	9.77
The Bible, King James version	3593KB	766,111	13,754	850,668	853,812	8.91
Shakespeare's complete works	5320KB	899,822	36,693	1,189,654	1,192,828	10.6
50MB English text from Pizza&Chilie corpus	51200KB	9,435,637	148,865	12,389,798	12,421,116	10.53
Shakespeare's complete works - character-based alphabet	5320KB	5,447,171	84	3,131,931	3,149,568	4.63

¹The source code can be downloaded from <https://github.com/reeWorlds/Fast-Large-Huffman/tree/main/Test3>.

Table 3

Decoding time comparison in seconds per 1000 iterations

File	State-of-the-Art	Opt. 1	Opt. 1 + 2	Opt. 1 + 3	Opt. 1 + 4	Opt. 1 + 2 + 3	Opt. 1 + 3 + 4
Alice	0.136	0.124	0.113	0.101	0.118	0.106	0.098
Harry Potter	0.385	0.362	0.348	0.309	0.341	0.328	0.293
Bible	3.738	3.516	3.071	3.051	3.279	2.836	2.855
Shakespeare	4.322	4.114	4.172	3.507	3.873	3.828	3.260
English.50MB	50.47	45.04	42.54	41.93	43.30	41.74	38.58
Shakespeare - char	1.165	2.281	1.574	2.079	2.171	1.369	2.017

Combining different optimizations allows us to accelerate the baseline algorithm significantly. As optimizations 2 and 4 are incompatible, we tested optimizations 1 + 2 + 3 and 1 + 3 + 4. The latter option demonstrates the best results for all word-based texts except for The Bible, where optimizations 1 + 2 + 3 are the best. This is explained by the fact that the average codeword length for The Bible is the shortest, which means that two codewords can most often be processed within the same iteration of the decoding loop in Algorithm 2. By contrast, for Shakespeare's complete works, where the average codeword length is the largest, optimization 2 actually reduces the effect of optimization 1. The state-of-the-art decoding (Moffat, & Turpin, 1997) is 31 – 39% slower than the methods that use the full stack of our optimizations.

We also included one character-based text in our experimental set (the last line in all tables). Its average codeword length is roughly half that of the word-based texts. Consequently, Algorithm 3 from (Liddell, & Moffat, 2006) proves to be the most time-efficient, even for moderate values of the lookup bitvector length p . Thus, for $p = 10$, the condition $N_{cwd}[buffer[1..p]] \geq 1$, which is checked in line 14 of Algorithm 3 in (Liddell, & Moffat, 2006), is almost always satisfied, which makes the corresponding 'if' statement predictable and removes the main source of inefficiency in that algorithm. The Ext-10 method from (Liddell, & Moffat, 2006) outperforms not only the baseline algorithm from (Moffat, & Turpin, 1997), but also our most efficient stack of optimizations 1 + 2 + 3.

Discussion and conclusions

We offer optimizations of Huffman Canonical Codes that improve state-of-the-art decoding time for texts on word-based alphabets by more than 30%. The decoding algorithms with our optimizations are faster than any other known decoding method for large alphabets, where codeword lengths may exceed 11-12 bits, except for BCT codes (Zavadskyi, 2022), which are faster but produce about 5% larger files. The proposed methods can be used in a wide range of data compression and information retrieval systems.

Authors' contribution: Igor Zavadskyi – conceptualization, methodology, Algorithm 2; Maksym Kovalchuk – software, empirical data collection and validation, Algorithms 3 and 4.

Sources of funding. Funding is partially provided by Taras Shevchenko National University of Kyiv.

References

- Anisimov, A. V., & Zavadskyi, I. O. (2017). Variable-length prefix codes with multiple delimiters. *IEEE Transactions on Information Theory*, 63(5), 2885–2985. <https://doi.org/10.1109/TIT.2017.2674670>
- Brisaboa, N. R., Fariña, A., Navarro, G., & Esteller, M. F. (2003). (s,c)-dense coding: An optimized compression code for natural language text databases. In Nascimento, de Moura, & Oliveira (Eds.), *String Processing and Information Retrieval* (pp. 122–136). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-39984-1_10
- Choueka, Y., Klein, S. T., & Perl, Y. (1985). Efficient variants of Huffman codes in high level languages. In *Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (p. 122–130). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/253495.342777>
- Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9), 1098–1101. <https://doi.org/10.1109/JRPROC.1952.273898>
- Liddell, M., & Moffat, A. (2006). Decoding prefix codes. *Software Practice and Experience*, 36, 1687–1710. <https://doi.org/10.1002/spe.741>
- Milidiu, R., Laber, E., Moreno, L., & Duarte, J. (2003, March 18–20). A fast decoding method for prefix codes. In Storer & Cohn (Eds.), *Data Compression Conference* (p. 438). Snowbird, Utah, USA. <https://doi.org/10.1109/DCC.2003.1194057>
- Moffat, A., & Turpin, A. (1997, October). On the implementation of minimum-redundancy prefix codes. *IEEE Transactions on Communications*, 45(10), 1200–1207. <https://doi.org/10.1109/26.634683>
- Schwartz, E., & Kallick, B. (1964). Generating a canonical prefix encoding. *Communications of the ACM*, 7(3), 40–50. <https://doi.org/10.1145/363958.363991>
- Zavadskyi, I. (2022, March 19–21). Binary-coded ternary number representation in natural language text compression. In Bilgin, Marcellin, Serra-Sagristà, & Storer (Eds.), *Data Compression Conference* (pp. 419–428). Snowbird, Utah, USA: IEEE. <https://doi.org/10.1109/DCC52660.2022.00050>
- Zavadskyi, I., & Kovalchuk, M. (2023, September 26–28). Binary mixed-digit data compression codes. In Nardini, Pisanti, & Venturini (Eds.), *String Processing and Information Retrieval* (pp. 381–392). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43980-3_31

Отримано редакцією журналу / Received: 31.01.25

Прорецензовано / Revised: 24.09.25

Схвалено до друку / Accepted: 10.10.25

Ігор ЗАВАДСЬКИЙ, д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0002-4826-5265
e-mail: ihorzavadskyi@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

Максим КОВАЛЬЧУК, асп.
ORCID ID: 0009-0002-8450-905X
e-mail: max.koval4uk@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

ОПТИМІЗАЦІЯ КАНОНІЧНИХ КОДІВ ХАФМАНА ДЛЯ ПОСЛІВНОГО СТИСНЕННЯ ПРИРОДНОМОВНИХ ТЕКСТІВ

Канонічні коди Хафмана — це метод стискального кодування джерела даних, що дає можливість швидкого декодування, не зменшуючи ефективності стиснення порівняно з класичними кодами Хафмана. Якщо максимальна довжина кодового слова не перевищує 11–12 бітів або довгі кодові слова трапляються дуже рідко, будь-який код змінної довжини, зокрема коди Хафмана, можна ефективно декодувати найвним "табличним" методом. Канонічні коди Хафмана доцільно використовувати насамперед тоді, коли стиснутий текст містить достатньо багато довгих кодових слів. Наприклад, це характерно для природномовних текстів, якщо елементами алфавіту вважати слова. Найшвидший на сьогодні алгоритм декодування канонічних кодів розроблено М. Лідлем та А. Мофатом. Ми пропонуємо кілька оптимізацій цього алгоритму, що, як свідчать проведені нами експерименти, дають змогу прискорити його на 20–30 %, не знижуючи коефіцієнта стиснення. Деякі з цих оптимізацій базуються на низькорівневому розпаралелюванні коду з використанням SIMD-архітектури сучасних процесорів, в інших пропонуємо поєднувати класичне декодування канонічних кодів із універсальним "табличним" методом. Крім згаданих оптимізацій, у статті детально розглянуто канонічні коди Хафмана, включно з базовим та оптимізованим М. Лідлем і А. Мофатом алгоритмами, а також універсальний "табличний" метод декодування кодів змінної довжини. Представлено результати експериментального порівняння швидкодії базового та оптимізованих алгоритмів. Результати дослідження можуть бути використані в системах інформаційного пошуку й архівування даних.

Ключові слова: Хафман, кодування, коди, канонічні, стискання тексту.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження (у зборі, аналізі чи інтерпретації даних, якщо це мало місце), у написанні рукопису та в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study (in the collection, analyses or interpretation of data if applicable), in the writing of the manuscript as well as in the decision to publish the results.

УДК 004.415 #

DOI: <https://doi.org/10.17721/1812-5409.2025/2.26>Даниїл ЛІПСЬКИЙ, асп.
ORCID ID: 0009-0000-4068-9453
e-mail: lipsky@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

АРХІТЕКТУРА ТА ФУНКЦІОНАЛЬНІ ОСОБЛИВОСТІ ПЛАТФОРМИ ДЛЯ АВТОМАТИЗОВАНОГО ТЕСТУВАННЯ ВЕБЗАСТОСУНКІВ

Представлено платформу для автоматизованого тестування вебзастосунків на основі об'єктно-орієнтованого підходу із застосуванням С#. Платформа призначена для масштабованого та гнучкого тестування вебзастосунків, підтримуючи можливості UI- та API-тестування. Описано основні компоненти системи, що проаналізовані з позиції їхньої функціональності та ролі в автоматизації. Детально розглянуто структуру та взаємодію компонентів, що забезпечують ефективне управління браузерним драйвером, налаштування тестового середовища, збір й обробку тестових даних. Особливістю є використання патерну "singleton" для роботи із браузерним драйвером і підтримка кросбраузерного тестування, що дає змогу перевіряти вебзастосунки на різних платформах. Платформа також підтримує динамічне налаштування та розширення можливостей завдяки конфігураційним файлам і централізованому управлінню параметрами. Система забезпечує зручний пошук вебелементів та їхніх колекцій і розширює функціональність для інтеграційного тестування API та роботи з локальним сховищем браузера. Платформа орієнтована на автоматизацію як сценарного, так й утилітарного тестування, сприяючи ефективному підходу до перевірки функціональності та стабільності вебзастосунків.

Ключові слова: автоматизоване тестування, вебзастосунки, С#, кросбраузерне тестування, конфігурація середовища, UI-тестування, API-тестування.

Класифікація відповідно до AMS 2020: 68M15, 68N19.

Вступ

В умовах сучасного швидкого розвитку вебтехнологій автоматизація тестування вебзастосунків стає ключовим фактором для забезпечення якості програмного забезпечення. Ручне тестування вже не здатне повністю відповідати вимогам ефективності, надійності та швидкості виведення продукту на ринок. Автоматизація дає змогу швидше виявляти помилки, охоплювати широке коло тестових сценаріїв і значно скорочувати час, необхідний для регресійних перевірок.

Традиційне тестування вебзастосунків передбачає перевірку функціональності інтерфейсу користувача, поведінки сторінок та їхніх компонентів у різних браузерах і середовищах. Однак, крім інтерфейсного тестування, важливо також проводити тестування API для забезпечення коректного обміну даними між фронтендом і бекендом. Комплексний підхід до автоматизованого тестування, що охоплює як UI-, так і API-тести, дає змогу ефективно перевіряти всі рівні вебзастосунку, що є особливо важливим для складних багатослов'язних систем (Forgacs, & Kovacs, 2024).

Метою представленої статті є опис запропонованої автором платформи у вигляді бібліотеки для автоматизації тестування вебзастосунків, що об'єднує можливості UI- та API-тестування, підтримує кросбраузерність і надає інструменти для детальної валідації й генерації звітів. В основі платформи використано сучасні інструменти і технології, що забезпечують її легке масштабування й адаптацію під вимоги конкретного проекту.

1. Архітектура платформи

Платформа для автоматизованого тестування є комплексним інструментом, що забезпечує точне, стабільне та гнучке виконання тестів. Її архітектура побудована так, щоб інтегрувати всі необхідні компоненти в єдиний потік виконання, охоплюючи такі ключові аспекти, як робота з вебдрайвером, обробка елементів, верифікація, логування, генерація звітів та багато ін. Завдяки запропонованій структурі, платформа дає змогу виконувати тестування навіть у найскладніших сценаріях, забезпечуючи масштабованість, простоту у використанні та високу ефективність. Наведемо опис основних модулів, з яких складається система.

Модуль WebDriverManager забезпечує запуск і контроль браузера під час виконання тестів, а його основними класами є IWebDriverManager і WebDriverManager (Homes, 2024). Інтерфейс IWebDriverManager задає контракт для роботи з драйвером. Він охоплює властивості та методи для отримання екземпляру вебдрайвера, використання механізмів очікування (WebDriverWait), пошуку елементів на сторінках (ElementFinder, ElementsFinder) та ведення логів через LogWorker. Метод OpenApplicationStartPage відповідає за відкриття стартової сторінки тестованого вебзастосунку.

Клас WebDriverManager реалізує цей інтерфейс і відповідає за створення драйвера відповідно до конфігурацій. При ініціалізації він отримує об'єкт TestSettings, де зберігаються всі параметри, включаючи тип браузера, версію, режим виконання та URL вебзастосунку. Локальний запуск драйвера здійснюється через метод GetWebDriver. У ньому враховують вибраний браузер (Chrome, Firefox, Safari), і застосовують відповідні налаштування. Для віддаленого виконання використовують метод GetRemoteWebDriver, що створює драйвер на основі Grid URI, заданого в конфігурації.

WebDriverWait реалізується через властивість Lazy<WebDriverWait> для ефективного управління очікуваннями. Це дає змогу виконувати динамічні дії з елементами без зайвого навантаження.

Метод OpenApplicationStartPage здійснює масштабування вікна браузера та навігацію на початкову сторінку, використовуючи URL із налаштувань. Він забезпечує готовність середовища для виконання тестів. Крім того, клас реалізує інтерфейс IDisposable для забезпечення коректного завершення сесій драйвера та збереження логів після виконання тестів. Цей модуль забезпечує гнучкість у налаштуванні браузера й адаптованість до різних середовищ виконання, що робить платформу універсальним інструментом для автоматизованого тестування.

Модуль обробки вебелементів у платформі реалізований через набір класів, серед яких Element, Elements, ElementFinder, ElementsFinder, а також методи розширення WebElementExtension (Homes, 2024). Клас Element

© Ліпський Даниїл, 2025

представляє окремий вебелемент і надає властивості й методи для взаємодії, включно з доступом до основного об'єкта типу `IWebElement` через `webElementCore`, отримання текстового значення елемента через `Text`, а також методи для кліку, очищення текстового поля та введення тексту. Додатково він підтримує функції автоматичного прокручування до елемента `ScrollToElement` та підсвічування елемента перед взаємодією `PerformActionWithHighlighting`, що підвищує стабільність і зручність тестування.



Рис. 1. Модулі ElementHandle та DriveManager

Клас `Elements` дає змогу працювати з колекціями вебелементів, забезпечуючи, наприклад, отримання кількості елементів у колекції через метод `Count`. Класи `ElementFinder` й `ElementsFinder` реалізують пошук одного або декількох елементів за допомогою CSS-селекторів, XPath або ID. Метод `Css` у `ElementFinder` шукає перший елемент за заданим CSS-селектором, повертаючи об'єкт `Element`, тоді як у `ElementsFinder` аналогічний метод повертає всі знайдені елементи у вигляді об'єкта `Elements` (Homes, 2024).

Методи розширення `WebElementExtension` додають функціональність до `IWebElement`, зокрема автоматичне прокручування до елемента, підсвічування під час взаємодії, а також переміщення курсора до центра елемента. Усі ці можливості доповнюють базовий функціонал Selenium `WebDriver`, забезпечуючи підвищену стабільність і зручність роботи з елементами. Завдяки модульній структурі цей компонент легко інтегрується з іншими частинами платформи, що робить його незамінним для автоматизації тестування.

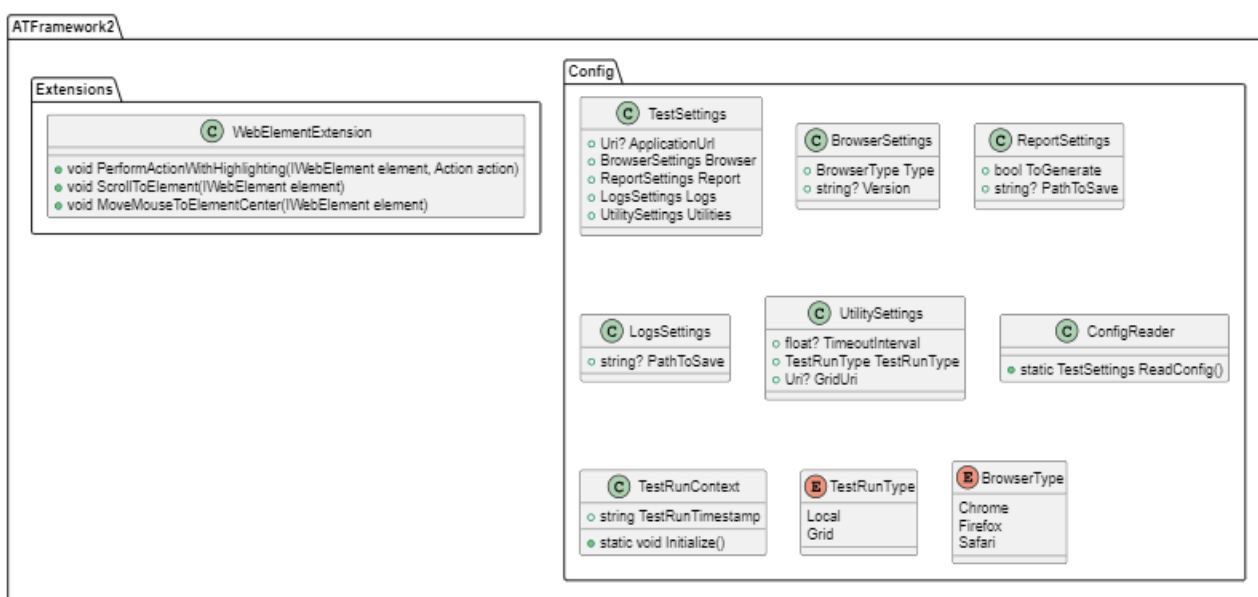


Рис. 2. Модулі Extensions і Configuration

Модуль Wait забезпечує стабільність тестів і коректну роботу з динамічними вебсторінками, реалізуючи механізм очікувань. Клас містить набір статичних методів для керування очікуваннями на основі WebDriverWait. Ініціалізація виконується через метод Initialize, що приймає екземпляр вебдрайвера, активує внутрішній об'єкт – driver і надає доступ до всіх функцій очікування.

Метод UntilTrue дає змогу чекати виконання заданої умови, що передається у вигляді функції condition, з можливістю встановлення тайм-ауту. У разі перевищення тайм-ауту генерується виняток із відповідним повідомленням. Методи UntilElementIsVisible й UntilElementIsClickable створені для роботи з вебелементами. Перший очікує появу елемента на сторінці та повертає його у вигляді об'єкта Element, другий перевіряє, коли елемент стане доступним для кліку.

Функціональність UntilAttributeContains дає змогу відстежувати зміни атрибутів елемента, очікуючи, поки атрибут отримає потрібне значення, що особливо корисно для контролю стану динамічних елементів.

Інтеграція цього механізму з іншими компонентами, зокрема з ElementFinder, сприяє ефективному поєднанню пошуку елементів з очікуваннями, що зменшує ймовірність помилок, спричинених асинхронною природою вебзастосунків. Завдяки цьому тестування стає надійнішим навіть у найскладніших сценаріях.

Модуль HtmlReportGenerator відповідає за створення детальних HTML-звітів про виконання тестів, надаючи інструменти для зручної й інтерактивної візуалізації результатів тестування. Ініціалізацію виконують методом Instance, що створює або повертає вже наявний екземпляр цього класу. Реалізація за допомогою патерну Singleton гарантує, що для одного тестового запуску буде створено лише один екземпляр. Конфігураційні параметри зчитують із TestSettings, що дає змогу визначити необхідність генерації звітів і вказати шлях для їхнього збереження.

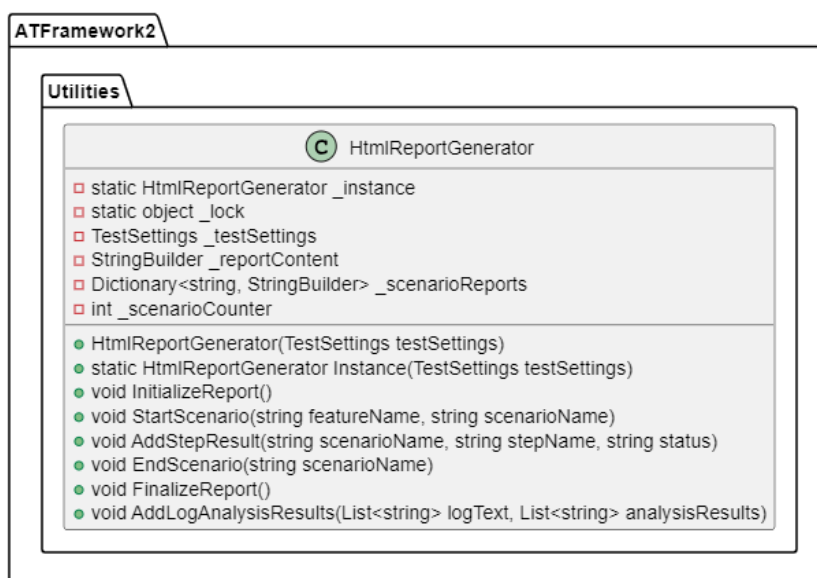


Рис. 3. Модуль HtmlReportGenerator

Формування базової структури HTML-документа виконують, використовуючи метод InitializeReport. Він додає заголовок, стилі CSS для оформлення таблиць, кнопок й інших елементів, а також JavaScript-код для інтерактивних функцій. Наприклад, collapsible-блоки дають можливість згорнути й розгорнути розділи для зручного перегляду деталей тестових сценаріїв.

Додавання нового сценарію до звіту здійснюється через метод StartScenario, що створює заголовок сценарію, ініціалізує таблицю для запису тестових кроків і присвоює унікальний ідентифікатор. Метод AddStepResult додає інформацію про виконані кроки, включно з їхніми назвами, статусом (Passed або Failed) і часовою міткою. Використовується кольорове виділення: зелений колір для успішних кроків і червоний – для невдалих. Після завершення сценарію метод EndScenario закриває відповідну таблицю.

Коли всі тестові сценарії виконані, метод FinalizeReport збирає всі дані в єдиний HTML-документ і зберігає його у вказаній директорії. Ім'я файлу містить часову мітку, що полегшує ідентифікацію конкретного тестового запуску.

Завдяки цій реалізації звіти є не лише інформативними, а й інтерактивними, надаючи детальну інформацію про кожен тест і спрощуючи аналіз результатів тестування.

Модулі CheckWorker і VerifyWorker забезпечують перевірку очікуваних результатів під час тестування, надаючи інструменти для контролю відповідності фактичних й очікуваних значень у тестових сценаріях. Його функціональність спрямована на підвищення гнучкості та стабільності тестів. Клас CheckWorker містить набір статичних методів для перевірки умов, таких як Equals, що дає змогу перевіряти рівність двох значень із підтримкою чутливості до регістру, зокрема: NotEquals, який перевіряє нерівність значень; Contains, що визначає наявність підрядка в рядку з опціональною підтримкою нечутливості до регістру; та Matches, який виконує перевірку за регулярними виразами. Специфічні методи, як DateTimeEquals і DateTimeNotEquals, забезпечують порівняння дат. Для динамічних середовищ метод ExecuteWithRetry допомагає повторювати перевірки до успіху або завершення часу очікування. Після завершення тестів функція FinalizeChecks збирає всі невдалі перевірки та надає виняток із деталізованим списком помилок.

Клас VerifyWorker надає розширені можливості, зокрема перевірку списків методом ListEqualsIgnoringOrder, який порівнює списки незалежно від порядку елементів. Метод Multiple дає змогу виконувати кілька перевірок одночасно, збираючи всі помилки в один результат. Логування результатів перевірок через LogWorker робить аналіз прозорим і зручним (Loubser, 2021).



Рис. 4. Модуль VerifyWorker



Рис. 5. Модуль CheckWorker

Обидва класи інтегровані з іншими компонентами платформи, такими як LogWorker для запису логів і Wait для повторних перевірок. Модуль верифікації є невід'ємною частиною платформи, забезпечуючи надійний контроль якості тестових сценаріїв і результатів, що сприяє стабільності й ефективності тестування.

Модуль ApiWorker забезпечує взаємодію з REST API, надаючи зручний і гнучкий інструмент для виконання HTTP-запитів й обробки відповідей у тестових сценаріях. Основою модуля є об'єкт HttpClient, що інкапсулює логіку виконання HTTP-запитів і допомагає гнучко налаштувати поведінку, наприклад, через додавання middleware або налаштування тайм-аутів. Для створення екземпляра ApiWorker потрібно передати ініціалізований HttpClient, що дає можливість легко адаптувати модуль до специфічних вимог тестування (Mohan, 2022).

Підтримка основних CRUD-операцій у модулі реалізована через асинхронні методи, що забезпечують ефективну взаємодію з API. Метод GetAsync виконує HTTP-запит до вказаного endpoint і повертає результат у вигляді рядка, перевіряючи успішність операції за допомогою EnsureStatusCode. Для надсилання даних використовують PostAsync, який приймає JSON-дані у форматі StringContent із відповідним Content-Type. Оновлення ресурсів на сервері виконують методом PutAsync, що має аналогічний механізм роботи. Видалення даних реалізовано через DeleteAsync, який надсилає запит на заданий endpoint.

Асинхронна реалізація всіх методів сприяє паралельному виконанню запитів у тестових сценаріях, підвищуючи продуктивність і гнучкість інтеграції. Завдяки реалізації інтерфейсу IDisposable клас ApiWorker дає змогу коректно завершувати роботу з HttpClient, вивільняючи ресурси після завершення тестування.

Зазначений модуль відіграє важливу роль у платформі, забезпечуючи можливість інтеграційного тестування, перевірки роботи API та легкого налаштування запитів відповідно до вимог тестування. Його взаємодія з іншими компонентами, такими як логування та верифікація, підвищує ефективність автоматизації та розширює функціональні можливості тестової платформи.

Модуль `LocalStorageWorker` надає інструменти для роботи з локальним сховищем браузера, забезпечуючи можливість маніпулювання даними, збереженими в локальному сховищі вебзастосунків. Його функціональність базується на виконанні JavaScript-кодів через `JavaScriptExecutor`, що забезпечує прямий доступ до сховища.

Метод `GetLocalStorage` виконує JavaScript-код для зчитування всіх ключів і їхніх значень у `Local Storage` та повертає результат у вигляді словника `Dictionary<string, string>`. Це допомагає перевіряти дані, що зберігаються застосунком під час роботи. Метод `AddToLocalStorage` додає нові записи у вигляді пари ключ – значення, використовуючи JavaScript-функцію `window.localStorage.setItem`, що дає змогу підготувати стан браузера для тестів. Оновлення даних у сховищі виконується методом `UpdateLocalStorageValue`, що змінює значення для заданого ключа. Видалення записів реалізовано через метод `DeleteFromLocalStorage`, що використовує функцію `removeItem`, а повне очищення `Local Storage` виконується за допомогою методу `ClearLocalStorage`, який викликає `window.localStorage.clear`.

Цей модуль є важливим для тестування застосунків, що використовують `Local Storage` для збереження даних користувачів, налаштувань або сесій. Наприклад, перед виконанням тестів можна змінювати дані в сховищі, щоб створити необхідні початкові умови для тестових сценаріїв.

`LocalStorageWorker` легко інтегрується з іншими модулями платформи. Отримані дані можуть бути перевірені за допомогою модулів верифікації, таких як `CheckWorker` або `VerifyWorker`, або збережені в тестових звітах через `HtmlReportGenerator`. Це забезпечує платформу додатковою гнучкістю і допомагає тестувати складні сценарії роботи з вебзастосунками.

Модуль `TestDataWorker` сприяє створенню унікальних і динамічних тестових даних, що адаптуються до різних сценаріїв тестування, надаючи гнучкість і варіативність вхідних параметрів. Його використання зменшує залежність від статичних даних і підвищує гнучкість тестів.

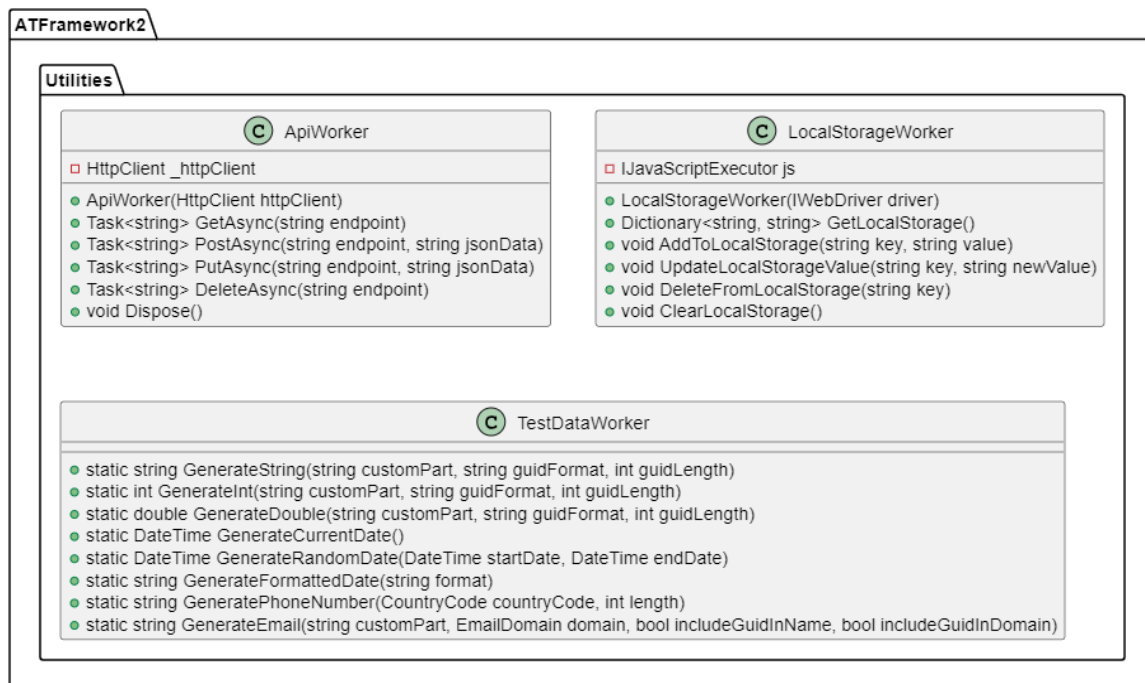


Рис. 6. Утилітарні компоненти платформи

Для генерації рядків і чисел метод `GenerateString` створює унікальний рядок, поєднуючи користувацьку частину та GUID із можливістю налаштування формату та довжини. Метод `GenerateInt` генерує унікальні цілі числа на основі хешування рядка з GUID, а `GenerateDouble` створює унікальні числа із плаваючою крапкою, комбінуючи рядок із хешованими значеннями. Для роботи з датами `GenerateCurrentDate` повертає поточну дату та час, `GenerateFormattedDate` дає змогу отримувати дати в заданому форматі, а `GenerateRandomDate` генерує випадкову дату в межах заданого інтервалу, перевіряючи коректність діапазону.

Генерація телефонних номерів здійснюється методом `GeneratePhoneNumber`, який створює номери із вказаним кодом країни, такими як +1 для США чи +380 для України, з можливістю налаштування довжини номера. Метод `GenerateEmail` продукує унікальні email-адреси, додаючи GUID до користувацької частини чи домену за необхідності, використовуючи вибраний домен, наприклад, `example.com` чи `test.com`.

Модуль зручний для автоматизації створення даних у різних тестових сценаріях, таких як реєстрація користувачів із використанням згенерованих email-адрес, тестування форм із випадковими датами або імітація унікальних телефонних номерів для перевірки локалізації. Згенеровані дані легко інтегруються з іншими модулями платформи, такими як модуль обробки елементів для заповнення форм чи `ApiWorker` для тестування API-запитів.

`TestDataWorker` є універсальним інструментом для динамічного налаштування тестових сценаріїв, що значно скорочує час їхньої підготовки й підвищує ефективність тестування.

Модуль `LogWorker` забезпечує відстеження процесу виконання тестів, збір діагностичної інформації та аналіз помилок, відіграючи вагомий роль у платформі (Loubser, 2021). Реалізація цього компонента інтегрована у важливі модулі, зокрема `CheckWorker`, `VerifyWorker`, `WebDriverManager` і `HtmlReportGenerator`.



Рис. 7. Модуль логуювання

Основна функціональність LogWorker охоплює запис текстових повідомлень у лог-файли з додаванням часових міток, рівнів важливості (напр., Info, Error) та контексту виконання, що містить метод і клас, які викликали лог. Це допомагає створювати детальні записи подій для легкого ідентифікування джерела проблем. Повідомлення мають структурований вигляд, наприклад: [2024-12-14 12:34:56] [INFO] [WebDriverManager. OpenApplicationStartPage] Відкрито стартову сторінку: <https://example.com>, що забезпечує зручність читання й аналізу.

Для різних компонентів створюються окремі лог-файли, наприклад, VerifyWorkerLog.txt для перевірок або GlobalTestLogs_{timestamp}.txt для роботи з драйвером. Такий підхід дає змогу чітко відокремлювати інформацію про різні частини системи.

Логуювання тісно інтегроване з іншими модулями платформи. У CheckWorker і VerifyWorker записують як успішні перевірки, так і помилки із зазначенням очікуваних і фактичних значень. HtmlReportGenerator використовує логуювання для запису процесу створення звітів, фіксації помилок й інших ключових подій. У WebDriverManager логуються дії, пов'язані із запуском браузера, навігацією між сторінками та завершенням сесій драйвера. Усі логи зберігаються у директорії, зазначеній у LogsSettings.PathToSave, що спрощує доступ до них для подальшого аналізу.

Завдяки структурованим логам тестувальники можуть швидко ідентифікувати проблеми, оптимізувати сценарії тестування та надати докази результатів замовникам. Цей компонент підвищує надійність і підтримуваність платформи, роблячи її ефективною навіть у масштабних проектах.

Інтеграція описаних модулів платформи в єдиний потік виконання є завершальним, але одним із найважливіших аспектів її роботи. Кожен модуль – від роботи з вебдрайвером, обробки елементів, верифікації, генерації звітів і до логуювання – органічно взаємодіє з іншими, забезпечуючи цілісність тестового процесу.

Потік починається з ініціалізації об'єкта TestSettings за допомогою модуля ConfigReader, який зчитує параметри з JSON-файлу для налаштування платформи відповідно до тестового запуску. Паралельно створюється контекст виконання TestRunContext із часовою міткою для подальшого відстеження дій. Після цього WebDriverManager створює вебдрайвер, вибираючи локальний чи віддалений режим роботи, а метод OpenApplicationStartPage відкриває стартову сторінку, визначену у TestSettings.ApplicationUrl. Усі дії логуються через LogWorker.

На наступному етапі модулі ElementFinder і ElementsFinder забезпечують пошук елементів для взаємодії, а Wait гарантує, що вони доступні. Через Element виконуються дії, як-от кліки чи введення тексту, із застосуванням методів розширення, що забезпечують стабільність і точність операцій. Результати дій перевіряються за допомогою модулів CheckWorker і VerifyWorker, які логують успішні перевірки та помилки.

Для інтеграційних тестів використовують ApiWorker, що надсилає HTTP-запити до бекенд-сервісів й аналізує відповіді. Усі важливі події, результати та статуси тестів фіксуються у лог-файлах через LogWorker, що забезпечує прозорість процесу. Завершальним етапом є генерація звітів через HtmlReportGenerator, який формує HTML-звіт із детальною інформацією про сценарії, кроки та результати виконання. Після завершення всіх тестів драйвер закривається, а зібрані дані зберігаються в заданих директоріях.

Інтеграція модулів платформи реалізована так, щоб забезпечити їхню взаємозалежність й узгоджену роботу, створюючи єдиний скоординований процес автоматизованого тестування. Наприклад, дані, знайдені ElementFinder, передаються в VerifyWorker для перевірки, результати верифікацій входять до звітів HtmlReportGenerator, а LogWorker записує всі ключові етапи роботи платформи.

Ця інтеграція забезпечує чіткий потік виконання, автоматизуючи всі аспекти тестування. Завдяки цьому платформа залишається гнучкою, масштабованою та простою у використанні навіть для складних проєктів, відповідаючи вимогам різних тестових сценаріїв.

2. Функціональні особливості платформи

Запропонована платформа для автоматизації тестування вебзастосунків є сучасним рішенням, що поєднує ефективність, стабільність і масштабованість у єдиному підході до тестування. Вона створена з урахуванням проблем, з якими стикаються тестувальники під час роботи з традиційними інструментами, та пропонує такі можливості для більш гнучкої та точної автоматизації:

1. Гнучка модульна архітектура є важливою особливістю платформи, оскільки вона дає змогу інтегрувати нові функціональні можливості без необхідності зміни основного коду. На відміну від багатьох рішень, що мають жорстку архітектурну структуру та вимагають значних змін при адаптації до нового середовища, ця платформа забезпечує легкість налаштувань і масштабування завдяки чітко визначеним компонентам. Це особливо важливо для великих проєктів, де необхідно швидко адаптувати тестове середовище до змін у вебзастосунку.

2. Підтримка різних типів тестування є однією із фундаментальних переваг платформи. Платформа дає можливість виконувати не лише тестування користувацького інтерфейсу (UI-тести), але й API-тестування, що є критичним для багаторівневих вебзастосунків. Завдяки цьому вона допомагає тестувальникам перевіряти як безпосередню взаємодію користувача з вебінтерфейсом, так і правильність роботи викликів API, що забезпечує комплексний підхід до перевірки програмного забезпечення.

3. Динамічне налаштування параметрів тестування через JSON-файли дає змогу легко змінювати середовище виконання тестів без необхідності вносити зміни до коду, що забезпечує гнучкість і зручність у використанні платформи. Це особливо корисно під час виконання тестування в різних середовищах, таких як локальне, тестове, проміжне або клієнтське, на відміну від традиційних платформ, що вимагають ручного внесення змін у код або використовують складні системи конфігурацій, що значно уповільнює процес налаштування тестів.

4. Наявність різних режимів запуску тестів, включно з локальним виконанням і запуском через Selenium Grid, сприяє ефективному використанню платформи в розподілених системах тестування. Це особливо важливо для великих команд або організацій із великою кількістю тестових сценаріїв, оскільки забезпечує можливість паралельного виконання тестів на різних пристроях і браузерах.

5. Механізм управління браузерними драйверами є перевагою платформи, оскільки використання патерну Singleton дає змогу уникнути зайвого створення браузерних сесій, що значно підвищує стабільність тестів та оптимізує використання ресурсів. У конкурентних платформах часто зустрічається проблема надмірного запуску браузерів, що призводить до перенавантаження тестового середовища та нестабільної роботи тестів. Завдяки правильному керуванню ресурсами й інтелектуальному очікуванню (Lazy WebDriverWait) платформа не перенавантажує систему і працює значно швидше.

6. Модуль обробки вебелементів забезпечує можливість взаємодії з елементами сторінок, спрощуючи виконання автоматизованих тестів. Убудовані механізми автоматичного прокручування до елементів та їхнє підсвічування під час виконання дій зменшують кількість тестових помилок, пов'язаних із невидимими або недоступними елементами. Це особливо корисно для складних динамічних вебзастосунків, у яких стандартні інструменти взаємодії з елементами можуть працювати нестабільно. Розроблювана платформа забезпечує глибший контроль над роботою з елементами, що підвищує точність і надійність тестування, на відміну від багатьох інших платформ, які обмежуються лише базовою взаємодією з інтерфейсом.

7. Наявна система очікувань не використовує жорстко задані таймінги, а дає можливість встановлювати динамічні умови очікування. Це особливо важливо для вебзастосунків, що активно використовують асинхронні запити та динамічний рендеринг контенту. На відміну від інших платформ, що часто вимагають ручного налаштування очікувань, цей підхід спрощує підтримку тестів і підвищує їхню стабільність.

8. Автоматично згенеровані інтерактивні HTML-звіти містять детальну інформацію про кожен тестовий сценарій, включно зі статусами виконання, скріншотами помилок та інтерактивними елементами для навігації. Це суттєво спрощує аналіз тестових запусків, у той час як багато інших рішень обмежуються лише текстовими або статичними звітами.

9. API-тестування є важливою особливістю платформи, оскільки забезпечує можливість автоматизованої перевірки роботи вебсервісів без необхідності ручного втручання. На відміну від платформ, які вимагають інтеграції з окремими інструментами для тестування API, запропонована система має вбудований API-модуль, що надає можливість проводити повноцінне тестування REST-інтерфейсів без додаткових налаштувань. Завдяки реалізації підтримки асинхронного виконання запитів, API-тести виконуються значно швидше, ніж під час використання традиційних рішень.

10. Можливості верифікації даних у платформі є розширеними, що сприяє здійсненню детальної перевірки коректності отриманих результатів у тестах. Убудовані модулі перевірки результатів підтримують автоматичний повтор перевірок у випадку невідповідностей, що забезпечує гнучкість під час тестування динамічних інтерфейсів. Це допомагає зменшити кількість хибно виявлених помилок у тестах і забезпечує точнішу верифікацію функціональності.

11. Підтримка роботи з локальним сховищем браузера дає змогу ефективно тестувати сценарії, пов'язані зі збереженням і використанням даних на рівні браузера. Це усуває потребу у використанні зовнішніх скриптів або ручного маніпулювання даними у сховищі.

Завдяки модульності архітектури, інтеграція штучного інтелекту не потребуватиме суттєвих змін у кодовій базі та дає змогу поступово розширювати інтелектуальні можливості платформи, додаючи нові алгоритми для поліпшення автоматизації тестування. Подальший розвиток платформи передбачає впровадження штучного інтелекту для аналізу тестових логів. Використання машинного навчання надасть можливість автоматично класифікувати та пріоритизувати помилки, аналізувати закономірності у збоях і прогнозувати потенційні проблеми. Це принесе додаткові можливості у сферу автоматизації тестування, оскільки допоможе зменшити навантаження на тестувальників і підвищити ефективність усунення дефектів.

Запропонована платформа забезпечує стабільність, продуктивність і зручність у використанні. Завдяки своїй гнучкості, інтегрованості та можливостям розширення, очікується, що вона може стати ефективним інструментом як для малих команд, так і для великих проєктів. Ураховуючи описану функціональність, ця система має значний потенціал для подальшого розвитку й упровадження нових інновацій у сфері автоматизації тестування.

Дискусія і висновки

Запропонована платформа для автоматизованого тестування вебзастосунків є комплексним рішенням, що поєднує можливості UI- та API-тестування, підтримує кросбраузерну взаємодію та допомагає масштабовано проводити перевірки функціональності вебзастосунків. Основною її особливістю є:

- модульна архітектура – легка адаптація, інтеграція нових компонентів і зміна параметрів тестування без редагування коду;

- комплексний підхід – поєднання UI- та API-тестування в єдиному середовищі;
- гнучкість – проста інтеграція нових функцій без змін у базовій структурі;
- централізоване управління конфігураціями – можливість швидкої зміни параметрів тестування без втручання в код;
- автоматичні HTML-звіти – детальна аналітика, скріншоти та логування для прозорості тестування.

Зазначена платформа також вирізняється високою стабільністю. Завдяки продуманій системі очікувань, розширеним механізмам роботи з елементами та налаштуванням параметрам очікування, вона надає можливість ефективно працювати навіть із динамічними вебзастосунками. Це робить її придатною для тестування складних вебзастосунків, що активно використовують асинхронні запити і динамічний контент.

Масштабованість платформи забезпечується підтримкою локального та віддаленого запуску тестів, що дає змогу використовувати її як у невеликих командах, так і в масштабних корпоративних проєктах. Вона легко інтегрується з іншими інструментами, що розширює можливості її використання.

Завдяки зручному налаштуванню параметрів тестування та простоті у використанні платформа підходить для тестування механізмів із різним рівнем підготовки. Вона дає змогу швидко розпочати автоматизацію тестування без необхідності глибокого занурення в технічні деталі, що значно скорочує час упровадження автоматизованих перевірок у процес розроблення.

Представлена автором платформа є сучасним, функціональним і перспективним інструментом для автоматизації тестування, який може бути ефективно використаний у широкому спектрі проєктів. Подальший розвиток, зокрема інтеграція методів штучного інтелекту для аналізу логів та автоматичної класифікації помилок, допоможе значно підвищити продуктивність процесу тестування, прискорити виявлення дефектів і поліпшити якість програмного забезпечення. Завдяки модульності архітектури, така інтеграція не вимагатиме суттєвих змін у кодовій базі, що сприятиме поступовому розширенню можливостей платформи відповідно до нових викликів у сфері тестування.

Джерела фінансування. Дослідження частково профінансоване Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Forgacs, I., & Kovacs, A. (2024). *Modern Software Testing Techniques: A Practical Guide for Developers and Testers* (1st ed.). Apress.
 Homes, B. (2024). *Fundamentals of Software Testing* (2nd ed.). Revised and Updated. Wiley-ISTE.
 Mohan, G. (2022). *Full Stack Testing: A Practical Guide for Delivering High Quality Software*. O'Reilly Media.
 Loubser, N. (2021). *Software Engineering for Absolute Beginners: Your Guide to Creating Software Products* (1st ed.). Apress.

References

- Forgacs, I., & Kovacs, A. (2024). *Modern Software Testing Techniques: A Practical Guide for Developers and Testers* (1st ed.). Apress.
 Homes, B. (2024). *Fundamentals of Software Testing* (2nd ed.). Revised and Updated. Wiley-ISTE.
 Mohan, G. (2022). *Full Stack Testing: A Practical Guide for Delivering High Quality Software*. O'Reilly Media.
 Loubser, N. (2021). *Software Engineering for Absolute Beginners: Your Guide to Creating Software Products* (1st ed.). Apress.

Отримано редакцією журналу / Received: 23.03.25
 Прорецензовано / Revised: 15.08.25
 Схвалено до друку / Accepted: 10.10.25

Danyil LIPSKYI, PhD Student
 ORCID ID: 0009-0000-4068-9453
 e-mail: lipskyi@knu.ua
 Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

ARCHITECTURE AND FUNCTIONAL CAPABILITIES OF A PLATFORM FOR AUTOMATED WEB APPLICATION TESTING

The article presents a platform for automated web application testing based on an object-oriented approach using C#. The platform is designed for scalable and flexible testing of web applications, supporting both UI and API testing capabilities. The main system components are described and analyzed in terms of their functionality and role in automation. The structure and interaction of components are examined in detail, highlighting efficient browser driver management, test environment configuration, and the collection and processing of test data. A distinctive feature is the use of the singleton pattern for browser driver management and support for cross-browser testing, allowing validation of web applications across different platforms. The platform also supports dynamic configuration and extensibility through configuration files and centralized parameter management. The system provides convenient mechanisms for locating web elements and their collections and extends functionality for API integration testing and working with the browser's local storage. The platform is geared toward automating both scenario-based and utility testing, contributing to an effective approach to verifying the functionality and stability of web applications.

Keywords: automated testing, web applications, C#, cross-browser testing, environment configuration, UI testing, API testing.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 519.21

DOI: <https://doi.org/10.17721/1812-5409.2025/2.27>

Ігор МАКУШЕНКО, канд. фіз.-мат. наук

ORCID ID: 0000-0002-4313-4936

e-mail: igormakushenko@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Ірина УСАР, канд. фіз.-мат. наук

ORCID ID: 0000-0002-3639-5505

e-mail: iryna.usar@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Ганна ЛІВІНСЬКА, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0001-9676-7932

e-mail: hanna.livinska@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

ГІСТЕРЕЗИСНА СТРАТЕГІЯ ДЛЯ ОДНОКАНАЛЬНОЇ СИСТЕМИ ІЗ ЧЕРГОЮ ТА ПОВТОРНИМИ ВИКЛИКАМИ

Розглянуто систему обслуговування марковського типу з повторними викликами, що складається з одного обслуговуючого приладу і в якій передбачено одне місце в черзі. Для оптимізації роботи системи використано гістерезисну стратегію. Основна ідея цього типу керування пов'язана із залежністю інтенсивності вхідного потоку від кількості джерел повторних викликів у системі. Роботу такої системи змодельовано тривимірним процесом міграції. Для нього встановлено умови існування стаціонарного режиму та досліджено імовірнісні характеристики процесу. Метод дослідження базується на апроксимації вихідного процесу процесом з обмеженим простором станів. Для ймовірнісних характеристик процесу у стаціонарному режимі одержано явні формули скалярного типу, для знаходження яких використано метод прирівнювання потоків імовірностей через границі областей, що обираються спеціальним чином. Для ілюстрації отриманих результатів наведено приклад розрахунку стаціонарних імовірностей.

Ключові слова: стохастична система, повторні виклики, змінна інтенсивність вхідного потоку, стаціонарний розподіл, процес міграції.

Класифікація відповідно до AMS 2020: 62H30, 34C60.

Вступ

Нині дослідження систем з повторними викликами є одним з найактуальніших напрямків сучасної теорії масового обслуговування, що активно розвивається. Ці системи детально розглянуті в монографії (Falin, & Templeton, 1997). Їхню популярність пояснюють тим, що такі системи мають широке застосування як математичні моделі для реальних систем у практиці проектування комп'ютерних мереж, дослідженні систем обробки інформації та сучасних систем зв'язку, для опису процесу посадки повітряних суден тощо (Artalejo, & Gomez-Corral, 2008; Yang, & Templeton, 1987). Характерна особливість систем з повторними викликами полягає в тому, що у випадку зайнятості всіх обслуговуючих приладів вимоги надходять на віртуальну орбіту і потім знову через випадковий проміжок часу повторюють спробу потрапити на обслуговування.

У роботі розглянуто систему із чергою та повторними викликами. У таких системах, якщо всі обслуговуючі прилади зайняті, вимоги, що надходять у систему, стають у чергу обмеженої довжини. Якщо всі місця в черзі зайняті, то такі вимоги надходять на віртуальну орбіту і потім, через випадковий проміжок часу повторюють спробу потрапити на обслуговування. Такі вимоги стають джерелами повторних викликів і генерують вторинний потік. Одне з важливих застосувань моделей із чергою та повторними викликами полягає в тому, що з їхньою допомогою можна ставити і розв'язувати задачі оптимального керування параметрами системи (Makushenko et al., 2023). Складність дослідження таких систем пов'язана з їхнім необмеженим фазовим простором. У роботі (Jonin, & Sedol, 1970) використовували ідею заміни нескінченної системи на скінченну. Проте постає питання про точність наближення характеристик нескінченної системи за допомогою отриманого стаціонарного розподілу урізаної системи (Falin, & Gomez-Corral, 2000). Розв'язання цієї задачі сприяє знаходженню для стаціонарних імовірностей явних формул скалярного типу.

1. Математична модель

У представленому дослідженні розглянуто систему обслуговування типу $[M / M / 1 / 1 + 1]^{(\infty)}$ з повторними викликами, що складається з одного обслуговуючого приладу і в якій передбачено одне місце у черзі. Час обслуговування має показниковий розподіл із параметром μ . На вхід обслуговуючого пристрою надходить пуассонівський потік вимог, інтенсивність якого функціонально залежить від числа джерел повторних викликів. Якщо вимога надійшла до системи, але виявилось, що всі прилади зайняті, то вона стає в чергу, як у класичних системах з очікуванням. Після звільнення обслуговуючого приладу, вимога із черги одразу займає його. Тобто зайнятість місця в черзі означає і зайнятість обслуговуючого приладу. У такій ситуації вимога, що надходить до системи, стає джерелом повторних викликів і генерує вторинний пуассонівський потік вимог інтенсивності ν .

Алгоритм керування роботою такої системи визначається на основі гістерезисної стратегії, яку можна описати так. Нехай $h_1 \leq h_2$ – два натуральні числа, які ми будемо називати порогами. Якщо кількість повторних викликів $j \leq h_1 - 1$, то система буде працювати в першому режимі й інтенсивність вхідного потоку дорівнює $\lambda_1(j)$. Якщо $j \geq h_2$, то в другому режимі з інтенсивністю вхідного потоку $\lambda_2(j)$. А якщо $h_1 \leq j < h_2$, то система зберігає той режим, у якому вона працювала в

© Макушенко Ігор, Усар Ірина, Лівінська Ганна, 2025

попередній момент часу. Клас порогових стратегій є підмножиною класу гістерезисних стратегій, який можна отримати за $h_1 = h_2$. Такі системи розглянуто в роботах (Lebedev, & Usar, 2013; Lebedev et al., 2017; Makushenko et al., 2023).

Процес обслуговування будемо моделювати тривимірним процесом міграції $Q(t) = (Q_1(t), Q_2(t), Q_3(t))$ у фазовому просторі $S = S^{(1)} \cup S^{(2)}$, $S^{(1)} \cap S^{(2)} = \emptyset$, де

$$S^{(1)} = \{(i, j, 1) : i = 0, 1, 2, j = 0, 1, \dots, h_2 - 1\}, \quad S^{(2)} = \{(i, j, 2) : i = 0, 1, 2, j = h_1, \dots\},$$

$Q_1(t)$ – сумарне число зайнятих приладів та місць у черзі в момент часу t , $Q_2(t)$ – число джерел повторних вимог, $Q_3(t)$ – режим роботи системи в момент часу t . Якщо $Q_3(t) = 1$, то система працює в першому режимі, якщо $Q_3(t) = 2$, то у другому режимі. Інфінітезимальні характеристики $q_{(i,j,r)}^{(i',j',r')}$ процесу $Q(t)$ визначаємо так:

1) якщо $(i = 0, 1, j = 0, 1, \dots, h_2 - 1, r = 1) \vee (i = 0, 1, j = h_1 + 1, \dots, r = 2)$, то

$$q_{(i,j,r)}^{(i',j',r')} = \begin{cases} \lambda_r(j), & \text{за } (i', j', r') = (i + 1, j, r), \\ \mu, & \text{за } (i', j', r') = (i - 1, j, r), \\ j\nu, & \text{за } (i', j', r') = (i + 1, j - 1, r), \\ -(\lambda_r(j) + i\mu + j\nu), & \text{за } (i', j', r') = (i, j, r), \\ 0, & \text{в інших випадках;} \end{cases}$$

2) якщо $(i = 2, j = 0, 1, \dots, h_2 - 2, r = 1) \vee (i = 2, j = h_1 + 2, \dots, r = 2)$, то

$$q_{(2,j,r)}^{(i',j',r')} = \begin{cases} \lambda_r(j), & \text{за } (i', j', r') = (2, j + 1, r), \\ \mu, & \text{за } (i', j', r') = (1, j, r), \\ -(\lambda_r(j) + \mu), & \text{за } (i', j', r') = (2, j, r), \\ 0, & \text{в інших випадках;} \end{cases}$$

3) якщо $i = 2, j = h_2 - 1, r = 1$, то

$$q_{(2,h_2-1,1)}^{(i',j',r')} = \begin{cases} \lambda_1(h_2 - 1), & \text{за } (i', j', r') = (2, h_2, 2), \\ \mu, & \text{за } (i', j', r') = (1, h_2 - 1, 1), \\ -(\lambda_1(h_2 - 1) + \mu), & \text{за } (i', j', r') = (2, h_2 - 1, 1), \\ 0, & \text{в інших випадках;} \end{cases}$$

4) якщо $i = 0, 1, j = h_1, r = 2$, то

$$q_{(i,h_1,2)}^{(i',j',r')} = \begin{cases} \lambda_2(h_1), & \text{за } (i', j', r') = (i + 1, h_1, 2), \\ \mu, & \text{за } (i', j', r') = (i - 1, h_1, 2), \\ h_1\nu, & \text{за } (i', j', r') = (i + 1, h_1 - 1, 1), \\ -(\lambda_2(h_1) + \mu + h_1\nu), & \text{за } (i', j', r') = (i, h_1, 2), \\ 0, & \text{в інших випадках.} \end{cases}$$

2. Умови існування та явні формули для стаціонарного розподілу

З'ясуємо умови існування стаціонарного режиму для тривимірного процесу Маркова $Q(t)$, $t \geq 0$.

Лема 1. Нехай $\lambda_2 = \lim_{j \rightarrow \infty} \lambda_2(j) < \infty$. Тоді за $\frac{\lambda_2}{\mu} < 1$ існує стаціонарний режим для процесу $Q(t)$.

Доведення. Доведення цієї леми базується на застосуванні теореми Твіді (Falin, & Templeton, 1997, р. 97). Розглянемо в ролі тест-функцій Ляпунова такі функції, як:

$$\Phi(i, j, r) = ai + j + r, \quad (i, j, r) \in S,$$

де параметр "a" буде визначено пізніше.

Якщо ми покажемо, що параметр "a" можна вибрати так, щоб середній перенос для цих функцій, який визначається за формулою

$$y_{ijr} = \sum_{(i',j',r') \neq (i,j,r)} q_{(i,j,r)}^{(i',j',r')} (\Phi(i', j', r') - \Phi(i, j, r)) = \sum_{(i',j',r') \neq (i,j,r)} q_{(i,j,r)}^{(i',j',r')} (a(i' - i) + j' - j + r' - r),$$

буде рівномірно менше нуля відносно (i, j, r) , за винятком, можливо, скінченної кількості точок, то із цього буде випливати наша лема.

Не обмежуючи загальності, будемо вважати, що $\lambda_2(j) \leq \mu\lambda_2$, де $\lambda_2 < 1$. Маємо

$$y_{ijr} = \sum_{(i',j',r') \neq (i,j,r)} q_{(i,j,r)}^{(i',j',r')} (ai' + j' - j) = \lambda_2(j)a + j\nu(a - 1) \leq \lambda_2\mu a + j\nu(a - 1), \quad j = h_1 + 1, \dots$$

$$y_{ijr} = \sum_{(i',j',r') \neq (i,j,r)} q_{(i,j,r)}^{(i',j',r')} (ai' + j' - a - j) = \lambda_2(j)a - \mu a + j\nu(a - 1) =$$

$$= a(\lambda_2(j) - \mu) + jv(a-1) \leq a\mu(2\lambda_2 - 1) + jv(a-1), \quad j = h_1 + 1, \dots$$

$$y_{ijr} = \sum_{(i', j', r') \neq (i, j, r)} q_{(i, j, r)}^{(i', j', r')} (a(i' - i) + j' - j) = \lambda_2(j) - \mu a \leq \mu(\lambda_2 - a), \quad j = h_1 + 1, \dots$$

Умови теореми Твіді будуть виконані, якщо ми вкажемо, що $\varepsilon > 0$ та $a > 0$ і що будуть виконуватися нерівності

$$\begin{cases} \lambda_2 a + jv(a-1) \leq -\varepsilon; \\ a\mu(2\lambda_2 - 1) + jv(a-1) \leq -\varepsilon; \\ \mu(\lambda_2 - a) \leq -\varepsilon, \end{cases}$$

для всіх j за винятком, можливо, скінченної їхньої кількості.

Нехай $\lambda_2 < a < 1$. Тоді легко зрозуміти, що для таких a ліві частини нерівностей у системі будуть меншими за $-\varepsilon$ з деяким $\varepsilon > 0$ для всіх достатньо великих j .

Лемі 1 доведено.

Для знаходження стаціонарного розподілу спочатку розглянемо урізану систему, у якій кількість джерел повторних викликів буде обмеженою деяким числом $N > h_2$. Таку систему ми позначатимемо як $[M / M / 1 / 1 + 1]^{(N)}$. Якщо $N < \infty$, то стаціонарний розподіл $\pi_{ij}^{(r)}(N) = \lim_{t \rightarrow \infty} P(Q_1(t) = i, Q_2(t) = j, Q_3(t) = r)$ для процесу $Q^{(N)}(t)$ завжди існує. Для спрощення позначень вважатимемо $\pi_{ij}^{(r)}(N) = \pi_{ij}^{(r)}$.

Введемо такі позначення:

$$\beta_i(j) = \lambda_i(j) + jv; \quad \alpha_i(j) = \lambda_i(j) [\beta_i^2(j) + jv\mu]; \quad \gamma_i(j) = jv\mu(\lambda_i(j-1) + 2\beta_i(j) + \mu);$$

$$B_i(j) = \begin{cases} \prod_{k=0}^{j-1} \frac{\alpha_i(k)}{\gamma_i(k+1)}, & \text{для } i=1, j=1, \dots, h_2-1, \\ \prod_{k=h_1}^{j-1} \frac{\alpha_2(k)}{\gamma_2(k+1)}, & \text{для } i=2, j=h_1+1, \dots \end{cases}$$

Позначимо матрицю розміром $(h_2 - h_1) \times (h_2 - h_1)$

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & \dots & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & \dots & 0 & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} & \dots & 0 & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & a_{h_2-h_1-1, h_2-h_1-2} & a_{h_2-h_1-1, h_2-h_1-1} & a_{h_2-h_1-1, h_2-h_1} \\ 0 & 0 & 0 & 0 & \dots & 0 & a_{h_2-h_1, h_2-h_1-1} & a_{h_2-h_1, h_2-h_1} \end{pmatrix}, \text{ де } a_{ij} = \begin{cases} \gamma_1(h_1-1+i) + \alpha_1(h_1-1+i), & \text{якщо } i=j, \\ -\alpha_1(h_1-2+i), & \text{якщо } i=j+1, \\ -\gamma_1(h_1-1+j), & \text{якщо } j=i+1, \\ 0, & \text{в інших випадках.} \end{cases}$$

Нехай $\bar{1}$ позначає вектор-стовпчик з $h_2 - h_1$ координатами, перша координата якого дорівнює одиниці, а решта – нулю. Уведемо величини $A(j), j = h_1, \dots, h_2 - 1$ так:

$$A^{-1}\bar{1} = (A(h_1), \dots, A(h_2-1))'.$$

Покладемо

$$b(h_1) = \alpha_1(h_1)A(h_1) - \gamma_1(h_1+1)A(h_1+1),$$

$$Q(j) = \alpha_1(h_1-1)B_1(h_1-1)B_2(j) \left[\frac{1 - \alpha_1(h_1)A(h_1)}{\mu h_1 v(\mu + \beta_2(h_1))} + b(h_1) \sum_{l=h_1}^{j-1} \frac{1}{\gamma_2(l+1)B_2(l+1)} \right],$$

$$D(h_1, h_2) = \frac{\alpha_2(h_2-1)Q(h_2-1) + \alpha_1(h_1-1)\alpha_1(h_2-1)A(h_2-1)B_1(h_1-1)}{\gamma_2(h_2)}.$$

Теорема 1. Якщо $\lambda_1(j), \lambda_2(j) > 0, j = 0, 1, \dots, N$, то для системи $[M / M / 1 / 1 + 1]^{(N)}$ справедливі наступні формули для стаціонарного розподілу

$$\frac{\pi_{0j}^{(1)}}{\pi_{00}^{(1)}} = \begin{cases} B_1(j), & \text{за } j=1, \dots, h_1-1; \\ \alpha_1(h_1-1)B_1(h_1-1)A(j), & \text{за } j=h_1, \dots, h_2-1; \end{cases}$$

$$\frac{\pi_{1j}^{(1)}}{\pi_{00}^{(1)}} = \begin{cases} \frac{\beta_1(j)B_1(j)}{\mu}, & \text{за } j=0, \dots, h_1-1; \\ \frac{\alpha_1(h_1-1)B_1(h_1-1)\beta_1(j)A(j)}{\mu}, & \text{за } j=h_1, \dots, h_2-1; \end{cases}$$

$$\frac{\pi_{2j}^{(1)}}{\pi_{00}^{(1)}} = \begin{cases} \frac{(j+1)v(\mu + \beta_1(j+1))B_1(j+1)}{\mu\lambda_1(j)}, \text{ } \text{за } j = 0, \dots, h_1 - 2; \\ \frac{\alpha_1(h_1 - 1)B_1(h_1 - 1)[1 - h_1 v \mu \lambda_1(h_1 - 1)A(h_1)]}{\mu^2 \lambda_1(h_1 - 1)}, \text{ } \text{за } j = h_1 - 1; \\ \frac{\alpha_1(h_1 - 1)B_1(h_1 - 1)[\alpha_1(j)A(j) - (j+1)v\mu\lambda_1(j)A(j+1)]}{\mu^2 \lambda_1(j)}, \\ \text{за } j = h_1, \dots, h_2 - 2; \\ \frac{\alpha_1(h_1 - 1)B_1(h_1 - 1)\alpha_1(h_2 - 1)A(h_2 - 1)}{\mu^2 \lambda_1(h_2 - 1)}, \text{ } \text{за } j = h_2 - 1; \end{cases}$$

$$\frac{\pi_{0j}^{(2)}}{\pi_{00}^{(1)}} = \begin{cases} Q(j), \text{ } \text{за } j = h_1, \dots, h_2 - 1; \\ \frac{B_2(j)D(h_1, h_2)}{B_2(h_2)}, \text{ } \text{за } j = h_2, \dots, N; \end{cases}$$

$$\frac{\pi_{1j}^{(2)}}{\pi_{00}^{(1)}} = \begin{cases} \frac{\beta_2(j)Q(j)}{\mu}, \text{ } \text{за } j = h_1, \dots, h_2 - 1; \\ \frac{\beta_2(j)B_2(j)D(h_1, h_2)}{\mu B_2(h_2)}, \text{ } \text{за } j = h_2, \dots, N; \end{cases}$$

$$\frac{\pi_{2j}^{(2)}}{\pi_{00}^{(1)}} = \begin{cases} \frac{\alpha_2(j)Q(j) - (j+1)v\mu\lambda_2(j)Q(j+1)}{\mu^2 \lambda_2(j)}, \text{ } \text{за } j = h_1, \dots, h_2 - 2; \\ \frac{\alpha_2(h_2 - 1)Q(h_2 - 1) - h_2 v \mu \lambda_2(h_2 - 1)D(h_1, h_2)}{\mu^2 \lambda_2(h_2 - 1)}, \text{ } \text{за } j = h_2 - 1; \\ \frac{(j+1)v(\mu + \beta_2(j+1))B_2(j+1)D(h_1, h_2)}{\mu\lambda_2(j)B_2(h_2)}, \text{ } \text{за } j = h_2, \dots, N - 1; \\ \frac{\alpha_2(N)B_2(N)D(h_1, h_2)}{\mu^2 \lambda_2(N)B_2(h_2)}, \text{ } \text{за } j = N, \end{cases}$$

де

$$\begin{aligned} \pi_{00}^{(1)} = & \left[1 + \mu^{-1}\beta_1(0) + \sum_{j=1}^{h_1-1} (1 + \mu^{-1}\beta_1(j))B_1(j) + \mu^{-1}v \sum_{j=0}^{h_1-2} \frac{(j+1)(\mu + \beta_1(j+1))B_1(j+1)}{\lambda_1(j)} + \right. \\ & + \frac{\alpha_1(h_1 - 1)B_1(h_1 - 1)}{\mu^2} \left\{ \mu^2 \sum_{j=h_1}^{h_2-1} (1 + \mu^{-1}\beta_1(j))A(j) + \sum_{j=h_1}^{h_2-2} \frac{[\alpha_1(j)A(j) - (j+1)v\mu\lambda_1(j)A(j+1)]}{\lambda_1(j)} + \right. \\ & + \frac{1 - h_1 v \mu \lambda_1(h_1 - 1)A(h_1)}{\lambda_1(h_1 - 1)} + \frac{\alpha_1(h_2 - 1)A(h_2 - 1)}{\lambda_1(h_2 - 1)} \left. \right\} + \sum_{j=h_1}^{h_2-1} (1 + \mu^{-1}\beta_2(j))Q(j) + \frac{1}{\mu^2} \sum_{j=h_1}^{h_2-2} \frac{\alpha_2(j)Q(j) - (j+1)v\mu\lambda_2(j)Q(j+1)}{\lambda_2(j)} \\ & + \frac{\alpha_2(h_2 - 1)Q(h_2 - 1) - h_2 v \mu \lambda_2(h_2 - 1)D(h_1, h_2)}{\mu^2 \lambda_2(h_2 - 1)} + \frac{D(h_1, h_2)}{B_2(h_2)} \left\{ \sum_{j=h_2}^N (1 + \mu^{-1}\beta_2(j))B_2(j) + \right. \\ & \left. + \frac{v}{\mu} \sum_{j=h_2}^{N-1} \frac{(j+1)(\mu + \beta_2(j+1))B_2(j+1)}{\lambda_2(j)} + \frac{\alpha_2(N)B_2(N)}{\mu^2 \lambda_2(N)} \right\} \left. \right]^{-1}. \end{aligned}$$

Доведення. Запишемо рівняння для стаціонарних імовірностей системи, використовуючи теорему про рівність потоків імовірностей через замкнений контур (Walrand, 1988, Section II). Проте окремо потрібно розглядати три області: область, де діє лише перший режим, тобто $0 \leq j \leq h_1 - 1$; область де діє лише другий режим, тобто $h_2 \leq j \leq N$ і область, де можуть діяти обидва режими, тобто $h_1 \leq j \leq h_2 - 1$. Для спрощення позначень будемо вважати $\pi_{ij}^{(r)}(N) = \pi_{ij}^{(r)}$. Отримаємо такі рівняння:

$$(j+1)v(\pi_{0j+1}^{(1)} + \pi_{1j+1}^{(1)}) = \lambda_1(j)\pi_{2j}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2, \quad (1)$$

$$(\lambda_1(j) + jv)\pi_{0j}^{(1)} = \mu\pi_{1j}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2, \quad (2)$$

$$(\lambda_1(j) + \mu + jv)\pi_{1j}^{(1)} = \lambda_1(j)\pi_{0j}^{(1)} + (j+1)v\pi_{0j+1}^{(1)} + \mu\pi_{2j}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2. \quad (3)$$

З рівнянь (1), (3) маємо

$$(\lambda_1(j) + \mu + jv)\pi_{1j}^{(1)} = \lambda_1(j)\pi_{0j}^{(1)} + \frac{(j+1)v(\lambda_1(j) + \mu)}{\lambda_1(j)}\pi_{0j+1}^{(1)} + \frac{\mu(j+1)v}{\lambda_1(j)}\pi_{1j+1}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2.$$

Якщо тепер урахувати (2) та введені позначення, то останнє рівняння можемо переписати так:

$$\alpha_1(j) \pi_{0j}^{(1)} = \gamma_1(j+1) \pi_{0j+1}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2.$$

Позначимо $a(j) = \frac{\alpha_1(j)}{\gamma_1(j+1)}$. Тоді отримаємо рекурентне співвідношення

$$\pi_{0j+1}^{(1)} = a(j) \pi_{0j}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2.$$

Звідки

$$\pi_{0j}^{(1)} = \prod_{k=1}^j a(j-k) \pi_{00}^{(1)} = \prod_{k=0}^{j-1} a(k) \pi_{00}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2. \quad (4)$$

Тепер із (2) маємо

$$\pi_{1j}^{(1)} = \frac{\lambda_1(j) + j\nu}{\mu} \prod_{k=0}^{j-1} a(k) \pi_{00}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2. \quad (5)$$

Аналогічно з (1) та (4), (5) випливає

$$\pi_{2j}^{(1)} = \frac{(j+1)\nu(\lambda_1(j+1) + \mu + (j+1)\nu)}{\lambda_1(j)\mu} \prod_{k=0}^j a(k) \pi_{00}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2. \quad (6)$$

Ураховуючи введені позначення, отримаємо

$$\begin{aligned} \pi_{0j}^{(1)} &= B_1(j) \pi_{00}^{(1)}, \quad j = 1, \dots, h_1 - 1, \\ \pi_{1j}^{(1)} &= \frac{\beta_1(j) B_1(j)}{\mu} \pi_{00}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2, \\ \pi_{2j}^{(1)} &= \frac{(j+1)\nu(\mu + \beta_1(j+1)) B_1(j+1)}{\lambda_1(j)\mu} \pi_{00}^{(1)}, \quad j = 0, 1, \dots, h_1 - 2. \end{aligned} \quad (7)$$

Напишемо рівняння для точок області, де можуть діяти обидва режими.

$$(\lambda_1(j) + j\nu) \pi_{0j}^{(1)} = \mu \pi_{1j}^{(1)}, \quad j = h_1 - 1, \dots, h_2 - 1. \quad (8)$$

$$(\lambda_1(j) + \mu + j\nu) \pi_{1j}^{(1)} = \lambda_1(j) \pi_{0j}^{(1)} + (j+1)\nu \pi_{0j+1}^{(1)} + \mu \pi_{2j}^{(1)}, \quad j = h_1, \dots, h_2 - 2. \quad (9)$$

$$(\lambda_1(h_1 - 1) + \mu + (h_1 - 1)\nu) \pi_{1h_1-1}^{(1)} = \lambda_1(h_1 - 1) \pi_{0h_1-1}^{(1)} + h_1 \nu (\pi_{0h_1}^{(1)} + \pi_{0h_1}^{(2)}) + \mu \pi_{2h_1-1}^{(1)}. \quad (10)$$

$$(\lambda_1(h_2 - 1) + \mu + (h_2 - 1)\nu) \pi_{1h_2-1}^{(1)} = \lambda_1(h_2 - 1) \pi_{0h_2-1}^{(1)} + \mu \pi_{2h_2-1}^{(1)}. \quad (11)$$

$$(\lambda_1(j) + \mu) \pi_{2j}^{(1)} = \lambda_1(j) \pi_{1j}^{(1)} + \lambda_1(j-1) \pi_{2j-1}^{(1)} + (j+1)\nu \pi_{1j+1}^{(1)}, \quad j = h_1, \dots, h_2 - 2. \quad (12)$$

$$(\lambda_1(h_2 - 1) + \mu) \pi_{2h_2-1}^{(1)} = \lambda_1(h_2 - 1) \pi_{1h_2-1}^{(1)} + \lambda_1(h_2 - 2) \pi_{2h_2-2}^{(1)}. \quad (13)$$

$$(\lambda_1(h_1 - 1) + \mu) \pi_{2h_1-1}^{(1)} = \lambda_1(h_1 - 1) \pi_{1h_1-1}^{(1)} + \lambda_1(h_1 - 2) \pi_{2h_1-2}^{(1)} + h_1 \nu (\pi_{1h_1}^{(1)} + \pi_{1h_1}^{(2)}). \quad (14)$$

Другий режим задіяний лише в рівняннях (10) та (14). Тому спочатку розглянемо інші рівняння. З (8) маємо

$$\pi_{1j}^{(1)} = \frac{\beta_1(j)}{\mu} \pi_{0j}^{(1)}, \quad j = h_1 - 1, \dots, h_2 - 1 \quad (15)$$

і, вилучивши за допомогою (15) з решти рівнянь $\pi_{1j}^{(1)}$, після нескладних перетворень отримаємо такі рівняння

$$\frac{\alpha_1(j)}{\lambda_1(j)\mu} \pi_{0j}^{(1)} = (j+1)\nu \pi_{0j+1}^{(1)} + \mu \pi_{2j}^{(1)}, \quad j = h_1, \dots, h_2 - 2, \quad (16)$$

$$\frac{\alpha_1(h_2 - 1)}{\lambda_1(h_2 - 1)\mu} \pi_{0h_2-1}^{(1)} = \mu \pi_{2h_2-1}^{(1)}, \quad (17)$$

$$(\lambda_1(j) + \mu) \pi_{2j}^{(1)} = \frac{\lambda_1(j) \beta_1(j)}{\mu} \pi_{0j}^{(1)} + \lambda_1(j-1) \pi_{2j-1}^{(1)} + \frac{(j+1)\nu \beta_1(j+1)}{\mu} \pi_{0j+1}^{(1)}, \quad j = h_1, \dots, h_2 - 2, \quad (18)$$

$$(\lambda_1(h_2 - 1) + \mu) \pi_{2h_2-1}^{(1)} = \frac{\lambda_1(h_2 - 1) \beta_1(h_2 - 1)}{\mu} \pi_{0h_2-1}^{(1)} + \lambda_1(h_2 - 2) \pi_{2h_2-2}^{(1)}. \quad (19)$$

За допомогою формули (16) вилучимо $\pi_{2j}^{(1)}$ з рівняння (18). Після відповідних перетворень отримаємо

$$\gamma_1(j+1) \pi_{0j+1}^{(1)} - (\alpha_1(j) + \gamma_1(j)) \pi_{0j}^{(1)} + \alpha_1(j-1) \pi_{0j-1}^{(1)} = 0, \quad j = h_1, \dots, h_2 - 2,$$

що дає нам $h_2 - h_1 - 1$ рівнянь для $h_2 - h_1 + 1$ невідомих $\pi_{0h_1-1}^{(1)}, \dots, \pi_{0h_2-2}^{(1)}$. Нам потрібно ще одне рівняння, яке ми отримаємо із (17), (19) та (16) за $j = h_2 - 2$. Після відповідних обчислень матимемо:

$$\alpha_1(h_2 - 2) \pi_{0h_2-2}^{(1)} - (\alpha_1(h_2 - 1) + \gamma_1(h_2 - 1)) \pi_{0h_2-1}^{(1)} = 0.$$

Отже, ми отримали таку систему лінійних рівнянь

$$\begin{cases} (\alpha_1(h_1) + \gamma_1(h_1))\pi_{0h_1}^{(1)} - \gamma_1(h_1 + 1)\pi_{0h_1+1}^{(1)} = \alpha_1(h_1 - 1)\pi_{0h_1-1}^{(1)}, \\ -\alpha_1(j - 1)\pi_{0j-1}^{(1)} + (\alpha_1(j) + \gamma_1(j))\pi_{0j}^{(1)} - \gamma_1(j + 1)\pi_{0j+1}^{(1)} = 0; \quad j = h_1 + 1, \dots, h_2 - 2, \\ -\alpha_1(h_2 - 2)\pi_{0h_2-2}^{(1)} + (\alpha_1(h_2 - 1) + \gamma_1(h_2 - 1))\pi_{0h_2-1}^{(1)} = 0, \end{cases} \quad (20)$$

де $\pi_{0h_1-1}^{(1)}$ будемо вважати величиною відомою.

Позначимо матрицю

$$A(h_1, h_2) = \begin{pmatrix} \hat{\alpha}(h_1) & -\gamma_1(h_1 + 1) & 0 & 0 & \dots & 0 & 0 \\ -\alpha_1(h_1) & \hat{\alpha}(h_1 + 1) & -\gamma_1(h_1 + 2) & 0 & \dots & 0 & 0 \\ 0 & -\alpha_1(h_1 + 1) & \hat{\alpha}(h_1 + 2) & \gamma_1(h_1 + 3) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -\alpha_1(h_2 - 2) & \hat{\alpha}(h_2 - 1) \end{pmatrix}, \text{ де } \hat{\alpha}(j) = \alpha_1(j) + \gamma_1(j).$$

Нехай

$$A^{-1}(h_1, h_2) \bar{1} = (A(h_1), A(h_1 + 1), \dots, A(h_2 - 1))', \quad \pi_0^{(1)} = (\pi_{0h_1}^{(1)}, \pi_{0h_1+1}^{(1)}, \dots, \pi_{0h_2-1}^{(1)})'.$$

Тепер систему (20) запишемо так:

$$A(h_1, h_2) \pi_0^{(1)} = \alpha(h_1 - 1) \pi_{0h_1-1}^{(1)} \bar{1}.$$

Звідси, з урахуванням (7) маємо

$$\pi_{0j}^{(1)} = \begin{cases} B_1(j) \pi_{00}^{(1)}, & \text{за } j = 1, \dots, h_1 - 1; \\ \alpha_1(h_1 - 1) B_1(h_1 - 1) A(j) \pi_{00}^{(1)}, & \text{за } j = h_1, \dots, h_2 - 1. \end{cases} \quad (21)$$

Із формул (7) і (15) отримаємо

$$\pi_{1j}^{(1)} = \begin{cases} \frac{\beta_1(j) B_1(j)}{\mu} \pi_{00}^{(1)}, & \text{за } j = 0, \dots, h_1 - 1; \\ \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1) \beta_1(j) A(j)}{\mu} \pi_{00}^{(1)}, & \text{за } j = h_1, \dots, h_2 - 1. \end{cases} \quad (22)$$

Запишемо тепер формули для $\pi_{2j}^{(1)}$ за $j = h_1 - 1, \dots, h_2 - 2$.

Із (16), (21) та (22) маємо

$$\pi_{2j}^{(1)} = \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1) [\alpha_1(j) A(j) - (j + 1) \nu \mu \lambda_1(j) A(j + 1)]}{\mu^2 \lambda_1(j)} \pi_{00}^{(1)}, \quad j = h_1, \dots, h_2 - 2.$$

Залишилося знайти $\pi_{2h_1-1}^{(1)}$ та $\pi_{2h_2-1}^{(1)}$. Із (18) за $j = h_1$ одержимо

$$\pi_{2h_1-1}^{(1)} = \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1)}{\mu^2 \lambda_1(h_1 - 1)} \{ [\gamma_1(h_1) + \alpha_1(h_1) - h_1 \nu \mu \lambda_1(h_1 - 1)] A(h_1) - \gamma_1(h_1 + 1) A(h_1 + 1) \} \pi_{00}^{(1)}. \quad (23)$$

З очевидної рівності

$$A(h_1, h_2) (A(h_1), A(h_1 + 1), \dots, A(h_2 - 1))' = \bar{1}$$

отримаємо, що

$$\begin{aligned} & [\gamma_1(h_1) + \alpha_1(h_1)] A(h_1) - \gamma_1(h_1 + 1) A(h_1 + 1) = 1, \\ & -\alpha_1(j - 1) A(j - 1) + (\alpha_1(j) + \gamma_1(j)) A(j) - \gamma_1(j + 1) A(j + 1) = 0, \quad j = h_1 + 1, \dots, h_2 - 2, \\ & -\alpha_1(h_2 - 2) A(h_2 - 2) + (\alpha_1(h_2 - 1) + \gamma_1(h_2 - 1)) A(h_2 - 1) = 0. \end{aligned} \quad (24)$$

З урахуванням (17) та (24) можемо записати (23) так:

$$\pi_{2h_2-1}^{(1)} = \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1) \alpha_1(h_2 - 1) A(h_2 - 1)}{\mu^2 \lambda_1(h_2 - 1)} \pi_{00}^{(1)}, \quad \text{за } j = h_2 - 1.$$

Отже, маємо

$$\pi_{2j}^{(1)} = \begin{cases} \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1) [1 - h_1 \nu \mu \lambda_1(h_1 - 1) A(h_1)]}{\mu^2 \lambda_1(h_1 - 1)} \pi_{00}^{(1)}, & \text{за } j = h_1 - 1; \\ \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1) [\alpha_1(j) A(j) - (j + 1) \nu \mu \lambda_1(j) A(j + 1)]}{\mu^2 \lambda_1(j)} \pi_{00}^{(1)}, & \text{за } j = h_1, \dots, h_2 - 2; \\ \frac{\alpha_1(h_1 - 1) B_1(h_1 - 1) \alpha_1(h_2 - 1) A(h_2 - 1)}{\mu^2 \lambda_1(h_2 - 1)} \pi_{00}^{(1)}, & \text{за } j = h_2 - 1. \end{cases}$$

Виходячи з аналогічних міркувань, шляхом переходу до інших частин фазового простору, ми отримаємо подання для всіх стаціонарних імовірностей через $\pi_{00}^{(1)}$. Для того, щоб знайти $\pi_{00}^{(1)}$ використовуємо умову нормування.

Теорему 1 доведено.

Якщо $N = \infty$, то за умови $\lim_{j \rightarrow \infty} \frac{\lambda_2(j)}{\mu} < 1$ існує стаціонарний розподіл нашої системи. Щоб отримати формули для стаціонарного розподілу, потрібно в теоремі 1 перейти до границі за $N \rightarrow \infty$. У зв'язку із цим покажемо, що ряд

$$\sum_{j=h_2}^{\infty} (\mu + \beta_2(j)) B_2(j)$$

є збіжний. Маємо, що

$$\lim_{j \rightarrow \infty} \frac{(\mu + \beta_2(j+1)) B_2(j+1)}{(\mu + \beta_2(j)) B_2(j)} = \lim_{j \rightarrow \infty} \frac{\lambda_2(j+1) + \mu + (j+1)v}{\lambda_2(j) + \mu + jv} \times \frac{\lambda_2(j) [(\lambda_2(j) + jv)^2 + \mu jv]}{\mu (j+1)v [2(j+1)v + \mu + \lambda_2(j) + 2\lambda_2(j)]} = \lim_{j \rightarrow \infty} \frac{\lambda_2(j)}{\mu} < 1,$$

що й означає збіжність ряду. Аналогічно доводимо збіжність ряду

$$\sum_{j=h_2}^{\infty} \frac{(j+1)(\mu + \beta_2(j+1))}{\lambda_2(j)} B_2(j+1).$$

3. Приклад

Наведемо приклад розрахунків для конкретної системи. Для цього розглянемо систему $[M / M / 1 / 1 + 1]^{(20)}$ з такими параметрами:

$$h_1 = 6, h_2 = 15, \mu = 1, v = 0, 3, \lambda_1(j) = 1, \lambda_2(j) = 2j / (1 + 1,5j).$$

Для цієї системи отримано такі результати розрахунків для стаціонарних імовірностей:

$\pi_{0j}^{(1)}$	0.133	0.079	0.042	0.021	0.011	0.005	0.002	0.001	0.000	0.000	0	0	0	0	0	0	0	0	0	0
$\pi_{1j}^{(1)}$	0.133	0.102	0.067	0.041	0.024	0.013	0.007	0.004	0.001	0.000	0	0	0	0	0	0	0	0	0	0
$\pi_{2j}^{(1)}$	0.054	0.066	0.057	0.042	0.029	0.019	0.011	0.007	0.002	0.001	0	0	0	0	0	0	0	0	0	0

$\pi_{0j}^{(2)}$	0	0	0	0	0	0.00001	0.00002	0.00002	0.00002	0.00002	0.00002	0.00001	0.00001	0.00001	0.00001	0	0	0	0	0
$\pi_{1j}^{(2)}$	0	0	0	0	0	0.00005	0.00007	0.00009	0.00009	0.00009	0.00009	0.00009	0.00008	0.00008	0.00005	0	0	0	0	0
$\pi_{2j}^{(2)}$	0	0	0	0	0	0.00007	0.00012	0.00016	0.00018	0.00020	0.00021	0.00022	0.00023	0.00024	0.00017	0	0	0	0	0

Дискусія і висновки

Розглянуто одноканальну стохастичну систему із чергою і повторними викликами та вхідним потоком змінної інтенсивності. Зміна інтенсивності вхідного потоку відбувається на основі гістерезисної стратегії. Робота такої системи моделюється тривимірним процесом міграції з неперервним часом. Для нього встановлено умови існування стаціонарного режиму та досліджено імовірнісні характеристики процесу. Метод дослідження базується на апроксимації вихідного процесу процесом з обмеженим простором станів. Дослідження стаціонарного режиму використовує рівність потоків імовірностей через границю замкненого контуру фазового простору. За відповідного вибору системи контурів для стаціонарних імовірностей отримано явні формули скалярного типу. Це дає змогу розв'язати задачу оцінки точності наближення характеристик нескінченної системи за допомогою отриманого стаціонарного розподілу урізаної системи. Одержані результати надають можливість ставити і розв'язувати задачі оптимального керування параметрами системи.

Внесок авторів: Ігор Макушенко – концептуалізація, методика, теоретичні основи дослідження; Ірина Усар – методика, збір і перевірка даних, емпіричне дослідження; Ганна Лівінська – методика, аналіз джерел, підготовка огляду літератури.

Джерела фінансування. Дослідження частково профінансоване Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Artalejo, J. R., & Gómez-Corral, A. (2008). *Retrial queueing systems*. Springer-Verlag. <https://doi.org/10.1007/978-3-540-78725-9>
- Falin, G. I., & Gómez-Corral, A. (2000). On a bivariate Markov process arising in the theory of single-server retrial queues. *Statistica Neerlandica*, 54, 67–78. <https://doi.org/10.1111/1467-9574.00126>
- Falin, G. I., & Templeton, J. G. C. (1997). *Retrial queues*. Chapman & Hall.
- Jonin, G. L., & Sedol, J. J. (1970). *Telephone systems with repeated calls*. Proceedings of 6 International Teletraffic Congress.
- Lebedev, E., Makushenko, I., Livinska, H., & Usar, I. (2017). On steady-state analysis of $[M|M|m|m + n]$ -type retrial queueing systems. In: Dudin, A., Nazarov, A., Kirpichnikov, A. (Eds), *Information Technologies and Mathematical Modelling. Queueing Theory and Applications: Vol. 800. Communications in Computer and Information Science* (pp. 133–146). Springer. https://doi.org/10.1007/9783-319-68069-9_11
- Lebedev, E., Usar, I. (2013). Retrial queueing systems with variable arrival rate, *Cybernetics and Systems Analysis*, 49(3), 457–464. <https://doi.org/10.1007/s10559-013-9529-9>
- Makushenko, I., Usar, I., Livinska, H., & Sharapov, M. (2023). Optimal threshold strategies for retrial systems with a queue. *Journal of Computational and Applied Mathematics*, 427, 115–136. <https://doi.org/10.1016/j.cam.2023.115136>
- Walrand, J. (1988). *An introduction to queueing networks*. Prentice Hall, Englewood Cliffs.
- Yang, T., & Templeton, J. G. C. (1987). A survey on retrial queues. *Queueing Systems*, 2, 201–233. <https://doi.org/10.1007/BF01158899>

References

- Artalejo, J. R., & Gómez-Corral, A. (2008). *Retrial queueing systems*. Springer-Verlag. <https://doi.org/10.1007/978-3-540-78725-9>
- Falin, G. I., & Gómez-Corral, A. (2000). On a bivariate Markov process arising in the theory of single-server retrial queues. *Statistica Neerlandica*, 54, 67–78. <https://doi.org/10.1111/1467-9574.00126>
- Falin, G. I., & Templeton, J. G. C. (1997). *Retrial queues*. Chapman & Hall.
- Jonin, G. L., & Sedol, J. J. (1970). *Telephone systems with repeated calls*. Proceedings of 6 International Teletraffic Congress.
- Lebedev, E., Makushenko, I., Livinska, H., & Usar, I. (2017). On steady-state analysis of $[M|M]m|m + n$ -type retrial queueing systems. In: Dudin, A., Nazarov, A., Kirpichnikov, A. (Eds), *Information Technologies and Mathematical Modelling. Queueing Theory and Applications: Vol. 800. Communications in Computer and Information Science* (pp. 133–146). Springer. https://doi.org/10.1007/9783-319-68069-9_11
- Lebedev, E., Usar, I. (2013). Retrial queueing systems with variable arrival rate, *Cybernetics and Systems Analysis*, 49(3), 457–464. <https://doi.org/10.1007/s10559-013-9529-9>
- Makushenko, I., Usar, I., Livinska, H., & Sharapov, M. (2023). Optimal threshold strategies for retrial systems with a queue. *Journal of Computational and Applied Mathematics*, 427, 115–136. <https://doi.org/10.1016/j.cam.2023.115136>
- Walrand, J. (1988). *An introduction to queueing networks*. Prentice Hall, Englewood Cliffs.
- Yang, T., & Templeton, J. G. C. (1987). A survey on retrial queues. *Queueing Systems*, 2, 201–233. <https://doi.org/10.1007/BF01158899>

Отримано редакцією журналу / Received: 07.04.25

Прорецензовано / Revised: 23.06.25

Схвалено до друку / Accepted: 10.10.25

Igor MAKUSHENKO, PhD (Phys. & Math.)

ORCID ID: 0000-0002-4313-4936

e-mail: igormakushenko@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Iryna USAR, PhD (Phys. & Math.)

ORCID ID: 0000-0002-3639-5505

e-mail: iryna.usar@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Hanna LIVINSKA, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0001-9676-7932

e-mail: hanna.livinska@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

HYSTERESIS STRATEGY FOR A SINGLE-CHANNEL SYSTEM WITH A QUEUE AND REPEATED CALLS

A Markov-type service system with repeated calls is considered. The system consists of one service device, and one place in the queue is provided. A hysteresis strategy is used to optimize the system operation. The main idea of this type of control is related to the dependence of the intensity of the input flow on the number of sources of repeated calls in the system. The operation of such a system is modeled by a three-dimensional migration process. The conditions for the existence of a stationary regime are established for it, and the probabilistic characteristics of the process are investigated. The research method is based on the approximation of the initial process by a process with a limited state space. Explicit formulas of the scalar type are found for the probabilistic characteristics of the process in the stationary regime. With this aim, the method of equalizing probability flows through the boundaries of regions that are chosen in a special way is used. To illustrate the obtained results, an example of calculating stationary probabilities is given.

Keywords: stochastic system, repeated calls, variable rate of external flow, stationary distribution, migration process.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 519.83:519.217.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.28>

Sergiy MARTYNIUK, PhD (Phys. & Math.), Assist.

ORCID ID: 0009-0009-5118-0658

e-mail: s.martyniuk@chnu.edu.ua

Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine

Viacheslav TSURKAN, PhD Student

ORCID ID: 0009-0005-4025-4392

e-mail: tsurkan.viacheslav.i@chnu.edu.ua

Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine

CHOOSING THE OPTIMAL STRATEGY FOR A DISCRETE DYNAMIC COOPERATIVE GAME

Game theory steadily expands across economics, sociology, and engineering. Improvements in computing power have facilitated the creation of mathematical models, the discovery of optimal solutions for diverse and complex problems, and the handling of large datasets. In recent years, game theory concepts have also been applied in the military domain and across all levels of security, including cybersecurity. A game involves the interaction of two or more rational players or teams, each pursuing specific strategies to compete for a given reward. Game theory provides the construction of an optimal strategy for such games. Game theory provides the framework for constructing an optimal strategy for such games. Moreover, game-theoretic approaches can be extended to developing algorithms that allow system developers to predict game outcomes in favor of a group of players, using complex game structures. This article explores research in game theory. A group of players, performing actions sequentially, aims to achieve the best possible result within a limited number of steps. Collected statistical data makes it possible to analyze optimal strategies practically. Based on game theory, the Markov model is also employed as a research method to achieve better outcomes in dynamic corporate games.

Keywords: *game theory, statistics, Markov model, cooperative games, strategy.*

AMS 2020 classification: 91A06, 91A12, 91A25, 60J10, 60J85, 05B20, 37M05, 03H05.

Introduction

In today's highly competitive and rapidly changing market environment, discrete dynamic cooperative games are becoming an essential tool for modeling the strategic behavior of companies and groups in the business environment. In real business, decisions are often made not continuously, but at specific intervals (for example, budgets are determined quarterly, strategic plans once a year). The discrete approach allows you to model the successive stages of decision-making clearly. The dynamism of the market environment is characterized by rapid changes in technology, legislation, and the emergence of new competitors. Dynamic games allow you to consider changes in the state of the system from stage to stage and predict their impact on the players' strategies. In a corporate environment, it is essential to consider the consequences of actions in the long term, and not just instant benefits, so with the help of discrete dynamic games, you can model situations where current decisions affect future opportunities. In addition, modern companies have access to significant amounts of statistical information (Big Data), which allows them to create more realistic dynamic models, and the application of game theory helps to optimize decisions based on the analysis of accumulated data. Game models help companies build adaptive strategies in conditions of uncertainty, when the behavior of competitors or external factors is unpredictable. Discrete dynamic games are used in economics, finance, cybersecurity, energy, logistics, and the development of IT solutions.

A dynamic discrete cooperative game is a model in game theory used to analyze strategic interactions between companies or divisions within an organization. In this model, time is viewed as discrete, that is, divided into distinct periods (e.g., quarters or years), and participants make decisions at each of these stages, taking into account both the current situation and the possible future consequences of their actions.

Let us consider the characteristics of dynamic discrete cooperative games, such as: *discreteness of time, multi-stage, and strategic interaction.*

Discreteness of time means that players make decisions at separate moments in time, which allows for modeling a sequence of strategic steps. Actions are separated by time intervals, and the results of actions accumulate or change step by step. This will enable you to structure the game, easily model strategies, and predict the consequences of actions. Facilitates the application of mathematical methods such as dynamic programming, Markov processes, and optimization theory. Corresponds to real-world scenarios where decisions are made not continuously, but periodically (for example, daily trading on the stock exchange, staged negotiations, turn-by-turn in the game).

Multi-stage (or multi-step) - game participants make decisions sequentially at several stages, and their decisions' results affect the game's further course. At each stage, players can change their strategies, adapt to new circumstances, and take into account both their own past decisions and the actions of other players and opponents. In this case, the game is divided into a certain number of stages or moves, and decisions at each stage depend on the current state of the game and the history of previous steps. The goal of such games is to optimize the total win or achieve a certain result by the end of the game, and players must take into account not only the current win but also the future consequences of their actions. Two types of multi-stage games are considered: with a fixed number of stages (finite game), when players know how many moves there will be, and with an infinite number of stages (infinite or stochastic game), when the game continues indefinitely or with a certain probability of ending at each stage.

In cooperative games that simulate the activities of companies or groups in a business environment, *strategic interaction* means a decision-making process by several participants (companies, divisions, groups), where each participant takes into account not only their own goals, but also the behavior of competitors, partners or customers. The gain of one player depends not only on his actions, but also on the actions of other players. Therefore, each participant tries to predict how others will

© Martyniuk Sergiy, Tsurkan Viacheslav, 2025

react to his strategy and adjusts his behavior accordingly. Players can either compete (for example, for market share) or cooperate (for example, create alliances). For such games, it is important to formalize the players, namely to determine who the participants are in the game, describe the strategies that are available to each player, and find an equilibrium state where none of the players has an incentive to change their strategy unilaterally (for example, Nash equilibrium). Among the features of cooperative games, it is worth noting the important role of information (complete, incomplete, asymmetric), long-term consequences of decisions, the frequent occurrence of mixed strategies (planning actions with certain probabilities), and possible coalitions or collaborations between players. Thus, strategic interaction in cooperative games is a process in which companies or groups make decisions, taking into account the possible actions of competitors and partners, to maximize their own gain in conditions of interdependence.

1. Analysis of the latest research and publications

Recent research in the field of game theory demonstrates its effectiveness in solving various corporate problems.

The basic principles of game theory were laid down by E. Borel, A. Cournot, J. Bertrand, J. Neumann, O. Morgenstern, J. Harshani, R. Selten, T. Schelling, R. Aumann. A significant contribution to its development was made by V. V. Itilinsky, V. Algin, T. Boydel, E. Wentzel, E. Vilkas, O. Granaturov, S. Nakonechny, V. Sokolov and others. However, it was thanks to J. Neumann that the final formation of this theory took place (Neumann, 1928). He managed to mathematically substantiate the general strategy for a game between two participants in conditions of minimization and maximization. The book by J. Neumann and O. Morgenstern's "Game Theory and Economic Behavior" demonstrate the possibility of applying game theory to a certain number of participants, which made it possible to use it in economics to model the behavior of enterprises in a competitive environment, that is, to form and choose a strategy for the development of enterprises. The "minimax" strategy, or minimization of maximum losses, proposed in this book is especially important. Such a strategy allows rationalizing the costs of enterprises in conditions of uncertainty (Baranovska, & Bukovskiy, 2017; Chan Kim, & Mauborgne, 2016; Dixit, & Nalebuff, 1991).

Zoryana Koval in her work proposes a methodology for selecting and evaluating enterprise strategies using game theory (Koval, 2021). This approach allows you to take into account competitors' strategies or the state of the market, which contributes to making optimal decisions in conflict situations. The author analyzes the advantages and disadvantages of using game theory methods in this area, and also explores the features of using strategy selection criteria. She considered the selection and search for opportunities to use game theory methods to analyze the effectiveness of enterprise strategies, developed methods for assessing the effectiveness of enterprise strategies by modeling the situation using game theory, and forming conclusions and recommendations on the application of game theory methods and tools in the field of assessing the effectiveness of enterprise strategies.

V. V. Kazimko explores the application of game theory to modeling information security problems (Kazimko, 2022). He notes that game theory concepts can be used to develop mechanisms that allow system developers to change the balance and predict outcomes in favor of defenders using complex game designs. Using games, it is possible to develop and analyze optimal player actions and find possible mathematical solutions to security problems. From a security perspective, the combination of cyberspace and physical space leads to cyber-physical security, or the combination of security and economics creates cyber insurance. To solve security and privacy problems in new areas, the most suitable tools are game-theoretical methods, as they provide a variety of proven mathematical methods for creating multi-user strategies using different ways to cover the confidentiality and security aspects of player interaction. The presented theories show the effectiveness and feasibility of using game theory and differential game theory in the field of information security.

2. Research purpose

The research is aimed at constructing an optimal strategy for a discrete dynamic cooperative game. In accordance with the set goal, the following main tasks are solved:

1. Analysis of the literature on research in the field of game theory, namely discrete dynamic cooperative games.
2. Construction of a mathematical model of the game.
3. Choice of strategy for the player.
4. Construction of an optimal strategy.

3. Presentation of the basic material and substantiation of the obtained research results

Consider a dynamic discrete cooperative game. There is a team of N players who perform actions sequentially. The initial player is selected randomly, after which the players take turns deciding to perform an action or pass the move to the next player (passing the move to oneself is prohibited). The game can have a maximum of M stages. If no one has performed an action by the last stage, then the player to whom the move was passed at the last stage is obliged to perform the action.

Rules of the game. Each player has two main options: *to perform an action* or *pass the move* to another player.

Performing an action can lead to:

- winning with probability P_{win} , which brings the player 1 point and ends the game;
- losing with probability P_{loss} , which leads to the loss of 1 point and also ends the game.

Passing the move to another player. A player passes the move to one of $N-1$ others with probability P_{pass} , and continues the game. Passing the move to oneself is prohibited. Passing the move to a player who has already participated in the game is allowed.

Winning chances for each player are variable at each stage.

Sequence of moves. After each player's move, the next player also chooses between performing an action or passing the move. The player can pass the move to any of the other players (except himself).

Multi-stage game structure. The game takes place in M stages:

Stage 1: The first player (randomly chosen) decides to perform an action or pass the move to another player.

Stage 2: The next player makes a decision. He can perform an action or pass the move to another player, including returning the move to the previous player.

...

Stage M: In the last stage, the player is obliged to perform an action.

Stopping the game. The game stops after any win or loss.

The goal of the game. It is necessary to calculate the probability of winning for players at any stage when performing an action, which involves taking into account all possible decisions of players at previous stages. Choose a strategy for each player at each stage and build the optimal game strategy to obtain the maximum chance of success.

Mathematical formulation.

Let us denote the set of players by the vector $G = (G_1 \ G_2 \ G_3 \dots G_N)$.

Probabilities of actions. As initial data, we take complete statistical data of previous games in terms of players and their quality of action execution. From these, we can form relative outcome matrices to construct the initial game matrices.

Let $P_{win\ j}^k$ be the probability of winning at the stage k ($k \in [1; M]$) of player G_j ($j \in [1; N]$). That is, in matrix form

$P_{win} = \begin{pmatrix} P_{win}^1 \\ P_{win}^2 \\ \dots \\ P_{win}^M \end{pmatrix}$, where P_{win}^k is the vector of success probability, when performing an action at stage k ($k \in [1; M]$), for each player of dimension N .

Let $P_{lost\ j}^k$ be the probability of losing at the stage k ($k \in [1; M]$) of player G_j ($j \in [1; N]$). In matrix form $P_{lost} = \begin{pmatrix} P_{lost}^1 \\ P_{lost}^2 \\ \dots \\ P_{lost}^M \end{pmatrix}$,

where P_{lost}^k is the vector of failure probabilities, when performing an action at stage k ($k \in [1; M]$), for each player of dimension N .

Let $P_{pass\ j}^k$ be the probability that player G_j ($j \in [1; N]$) decides to pass the move instead of performing an action at stage k ($k \in [1; M]$).

$$P_{pass\ j}^k = 1 - P_{win\ j}^k - P_{lost\ j}^k.$$

In matrix form $P_{pass} = \begin{pmatrix} P_{pass}^1 \\ P_{pass}^2 \\ \dots \\ P_{pass}^M \end{pmatrix}$, where P_{pass}^k is a vector of probabilities of passing a move for each player of dimension N ,

and ($k \in [1; M]$).

The probability of passing a move by player G_j ($j \in [1; N]$) to player G_i ($i \in [1; N]$) is the same for any of the stages, denoted by $P_{pass\ ji}$. Since passing a move to oneself is forbidden, then $P_{pass\ jj} = 0$. Moreover, the sum of all probabilities is equal to 1, that is:

$$\sum_{i=1}^N P_{pass\ ji} = 1.$$

Accordingly, for player G_j the following equality is valid at each stage:

$$P_{win\ j}^k + P_{lost\ j}^k + P_{pass\ j}^k \times (\sum_{i=1}^N P_{pass\ ji}) = 1; \text{ where } k \in [1; M] \text{ is the stage number, } j, i \in [1; N].$$

Objective:

1) Calculate the probability of winning for players at any stage when making certain moves, which involves taking into account all possible decisions of players at previous stages.

2) Choose a strategy for each player at each stage.

3) Construct the optimal game strategy to obtain the maximum winning chances.

Solution Approach.

To calculate the final probability of a player winning at the last stage, one can plot all possible scenarios and apply a decision tree with conditional probabilities at each step. This will ensure an accurate calculation of the success probability.

The probability of a move reaching a player at a given stage and the formation of a payoff matrix.

In order for the game to reach a player, it is necessary that the players at the previous stages pass the move on without performing any action. This scheme is a first-order Markov chain. We will write this sequence in a transition probability matrix, where the matrix data represent the probability of a move passing from one player to another and, finally, achieving a result. Thus, a matrix of probabilities of a move passing from one player to another was created, where the rows correspond to the player passing the move, and the columns to the player receiving the move. Since a player cannot pass the move to himself, the central diagonal of the matrix will be zero.

$$T = \begin{pmatrix} 0 & p_{12} & p_{13} & \dots & p_{1N} \\ p_{21} & 0 & p_{23} & \dots & p_{2N} \\ p_{31} & p_{32} & 0 & \dots & p_{3N} \\ \dots & \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & p_{N3} & \dots & 0 \end{pmatrix}.$$

For a given transition matrix, the statement that the sum of the probabilities in each row is equal to 1 is valid:

$$\sum_{i=1}^N p_{ij} = 1, \quad j \in [1; N]$$

where $i \in [1; N]$ is the row number of the player who passes the move, $j \in [1; N]$ is the column number of the player who receives the move.

To determine the probability of a player getting a move at a certain stage, we define the initial state vector.

$$V^1 = (p_1 \ p_2 \ p_3 \ \dots \ p_N),$$

where the elements are the probabilities of the player getting the first move.

Let's consider the probabilities of the player getting a move at the second stage, for this we multiply the initial state vector by the transition matrix. The resulting vector will be the resulting vector.

$$V^2 = (V^1 \circ P_{pass}^1) \times T,$$

where the elements are the probabilities of the player getting a move in the second stage. Using the Hadamard product, the vectors V^1 and P_{pass}^1 are multiplied element by element.

For the stage k , the equality will take the form:

$$V^k = (V^{k-1} \circ P_{pass}^{k-1}) \times T = (V^1 \circ P_{pass}^1 \circ P_{pass}^2 \circ \dots \circ P_{pass}^{k-1}) \times T^{k-1}.$$

To determine the probability of winning players at the stage k (W^k), we calculate as the Hadamard product the probability of reaching the player's move at the stage k and the probability of winning the player when performing an action at this stage:

$$W^k = V^k \circ P_{win}^k.$$

Therefore, the total probability of winning for players at all stages can be represented as a matrix: $W = \begin{pmatrix} W^1 \\ W^2 \\ \dots \\ W^M \end{pmatrix}$, where

W^k is a vector of success probability for each player of dimension N at stage k ($k \in [1; M]$).

The matrix found is the odds of winning for players at any stage when performing an action that involves taking into account all possible decisions of players at previous stages.

Choosing a strategy for a player at each stage.

To find the optimal strategy, you need to decide when players should perform an action and when to pass the move to the next player in order to maximize the overall success probability.

Note that:

- At each stage, players have a choice: perform an action or pass the move.
- The optimal strategy will depend on the probabilities of winning and losing each player at each stage, as well as on the probabilities of passing the move.

Determining the criteria for choosing an action or passing the move at stage k for player G_j :

If player G_j received a move at stage k , the expected gain from performing the action for him is W_j^k , where the probability of receiving a move at this stage and the probability of winning are taken into account.

The criterion for choosing to perform an action will be the correspondence of the value of W_j^k to the maximum value of the probability of winning for this stage. That is, if $W_j^k = \max \{W_i^m, \forall i \in [1, N], \forall m \in [k, M]\}$ is the maximum value among the elements of the payoff matrix vectors W , of all rows of stages starting from the k .

If this value is less, the player must pass the move to the next player, i.e. continue the game. The priority player for passing the move will be the player with the maximum odds of winning at the next stage. If such a player is the player passing the move, then pass the move to the next player with the highest probability.

$$G_{next} = \max_i \{W_i^{k+1}, \forall i \in [1, N], i \neq next, k < M\}.$$

When receiving a move at the last stage M , the player is obliged to perform an action.

After constructing the vectors of success probability at each stage, it is possible to determine the potentially most optimal players at each stage. Accordingly, the transfer of the move to the next stage, namely the choice of the player, can be chosen precisely according to this criterion.

Calculation of the optimal strategy.

Using the above criteria, we can determine when each player should pass the move or perform an action to maximize the overall odds of winning.

To achieve a cooperative win, it is necessary to determine the stage and the player with the maximum probability of winning, and construct the optimal path to reach the given player at the given stage for him to perform the action.

$$\max_{i,k} \{W_i^k, \forall i \in [1, N], \forall k \in [1, M]\}.$$

From the maximum value of the probability of winning, the index i will correspond to the player G_i , and k - to the stage at which he should perform the action.

Discussions and conclusions

We have formulated and proposed a solution to a dynamic discrete cooperative game. The strategies of the players and the team are determined for the optimal solution of this game.

This dynamic discrete cooperative game is the basic part for building a mathematical model of the game and building optimal strategies for players. Using the constructed outcome matrices (Martyniuk, & Tsurkan, 2024), it is possible to form input data for the considered game. Since probabilistic data are formed based on the analysis of players' actions during previous similar games, they will clearly set the initial conditions for the game. In further studies, it is important to consider the qualitative impact of performing an action or passing a move on the constructed game, as well as the dynamics of changing the initial data for the game according to the mathematical model.

The complete game model can be broken down into parts that can be represented and solved using the game in question. Also, an important area of research will be the practical use. What areas of problems can be represented by such games, and how can we form the initial data for them?

Authors' contribution: Sergiy Martyniuk – conceptualization; methodology; Viacheslav Tsurkan – empirical data collection and validation; empirical research; analysis of sources, preparation of literature review or theoretical foundations of research.

Sources of funding. This study did not receive any grant from a financial institution in the public, commercial, or non-profit sectors.

References

- Baranovska, L. V., & Bukovskiy, O. M. (2017). Mixed strategy Nash equilibrium in one game and rationality. In *Proceedings of the III International Scientific and Practical Conference "Scientific Issues of the Modernity"* (pp. 4–8). ROSTranse Trade F Z C company, Scientific and Educational Consulting Group "WORLD Science".
- Chan Kim, W., & Mauborgne, R. (2016). *Blue ocean strategy*. Family Leisure Club.
- Dixit, A. K., & Nalebuff, B. J. (1991). *Thinking strategically: The competitive edge in business, politics, and everyday life*. Norton.
- Kazimko, V. V. (2022). Application of game theory to model information security problems. *Telecommunications and Information Technologies*, 1(74), 15–28 [in Ukrainian]. [Kazimko, B. B. (2022). Застосування теорії ігор для моделювання інформаційних проблем безпеки. *Телекомунікаційні та інформаційні технології*, 1(74), 15–28]. <https://doi.org/10.31673/2412-4338.2022.011524>
- Koval, Z. O. (2021). Evaluating enterprise strategy using game theory. *Management and Entrepreneurship in Ukraine: Stages of Formation and Development Problems*, 3(2), 47–55 [in Ukrainian]. [Коваль, З. О. (2021). Оцінювання стратегії підприємства методом теорії ігор. *Менеджмент та підприємництво в Україні: етапи становлення і проблеми розвитку*, 3(2), 47–55]. <https://doi.org/10.23939/smeu2021.02.047>
- Martyniuk, S., & Tsurkan, V. (2024). Construction of a mathematical model of a player using a matrix of outcomes. *Mathematics. Information technologies. Education*, 11, 53–55 [in Ukrainian]. [Мартинюк, С., & Цуркан, В. (2024). Побудова математичної моделі гравця з використанням матриці результатів. *Математика. Інформаційні технології. Освіта*, 11, 53–55]
- Neumann, J. von. (1928). Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100, 295–320. <https://doi.org/10.1007/BF01448847>

Отримано редакцією журналу / Received: 07.08.25

Прорецензовано / Revised: 30.09.25

Схвалено до друку / Accepted: 10.10.25

Сергій МАРТИНЮК, канд. фіз.-мат. наук, асист.

ORCID ID: 0009-0009-5118-0658

e-mail: s.martyniuk@chnu.edu.ua

Чернівецький національний університет імені Юрія Федьковича, Чернівці, Україна

Вячеслав ЦУРКАН, асп.

ORCID ID: 0009-0005-4025-4392

e-mail: tsurkan.viacheslav.i@chnu.edu.ua

Чернівецький національний університет імені Юрія Федьковича, Чернівці, Україна

ВИБІР ОПТИМАЛЬНОЇ СТРАТЕГІЇ ДЛЯ ДИСКРЕТНОЇ ДИНАМІЧНОЇ КООПЕРАТИВНОЇ ГРИ

З кожним роком відбувається розширення використання теорії ігор в економіці, соціології, інженерних науках. Розвиток обчислювальних ресурсів дає змогу будувати математичні моделі та знаходити оптимальні стратегії для різноманітних задач і різних рівнів їхньої складності, обробляти великі об'єми інформації. Останнім часом концепції теорії ігор застосовують також до військової сфери, до всіх рівнів безпеки, включно з кіберпростором. Гра передбачає дії двох або більше раціональних гравців чи команд, що мають певну стратегію і змагаються за певну винагороду. Теорія ігор забезпечує побудову оптимальної стратегії для таких ігор. Крім того, теоретико-ігрові підходи можна поширити на розроблення алгоритмів, що допомагають розробникам систем передбачати результати ігор на користь групи гравців, використовуючи складні ігрові конструкції.

У пропонованій статті розглянуто дослідження теорії ігор. Група гравців, виконуючи послідовно дії, прагнуть досягнути максимального результату за обмежену кількість кроків. Використання зібраних статистичних даних дає змогу на практичному рівні розглянути оптимальні стратегії. Також для досягнення кращих результатів у динамічних кооперативних іграх із використанням теорії ігор як метод дослідження застосовано модель Маркова.

Ключові слова: теорія гри, статистика, модель Маркова, кооперативні ігри, стратегія.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 007:004.9

DOI: <https://doi.org/10.17721/1812-5409.2025/2.29>

Михайло МАХНО¹, канд. техн. наук, доц.
ORCID ID: 0000-0001-5045-1705
e-mail: mykhailo.makhno@knu.ua

Олексій ФЕДОРУС¹, канд. техн. наук, асист.
ORCID ID: 0009-0006-5496-4544
e-mail: oleksii.fedorus@knu.ua

Олександр БОРИСЕНКО¹, асп.
ORCID ID: 0009-0009-7852-3227
e-mail: borysenko@knu.ua

Максим ВЕРЕМЧУК², асп.
ORCID ID: 0009-0003-9787-2787
e-mail: mrveremc@uwaterloo.ca

¹Київський національний університет імені Тараса Шевченка, Київ, Україна

²Університет Ватерлоо, Ватерлоо, Онтаріо, Канада

АРХІТЕКТУРА СИСТЕМИ ВИЯВЛЕННЯ БОТІВ У СОЦІАЛЬНИХ МЕРЕЖАХ

Сучасні інформаційні системи потребують ефективної архітектури для забезпечення високої продуктивності, масштабованості та надійності. У представленій статті розглянуто підхід до проєктування архітектури системи, що враховує новітні технологічні рішення та методи оптимізації роботи з великими масивами даних. Запропоновано власну архітектуру системи виявлення ботів, засновану на мікросервісному підході та сучасних методах обробки даних. На відміну від наявних рішень, система акцентує увагу не на розробленні радикально нового методу класифікації, а на ефективному поєднанні вже відомих підходів у єдиній архітектурі.

Розвиток інформаційних технологій вимагає створення ефективних архітектурних рішень, що забезпечують високу продуктивність і надійність програмних систем. Зі зростанням обсягів даних і вимог до швидкодії, класичні підходи до побудови архітектури потребують удосконалення. Дослідження в цій галузі є важливими для розробників й архітекторів програмного забезпечення. Мета пропонованої роботи – розроблення концепції архітектури, що відповідає сучасним вимогам продуктивності, масштабованості та безпеки. Основними завданнями є аналіз наявних підходів, ідентифікація їхніх переваг і недоліків, а також вибудовування ефективної архітектури, що мінімізує витрати ресурсів і підвищує швидкість обробки даних.

У дослідженні застосовано методи аналізу архітектурних рішень, моделювання системи, тестування продуктивності та порівняльний аналіз ефективності різних підходів. Для реалізації архітектури використано сучасні технології, зокрема мікросервісний підхід, контейнеризація та розподілені обчислення.

Запропонована архітектура дає змогу підвищити продуктивність системи шляхом оптимізації обробки запитів та розподілу навантаження між сервісами. Використання контейнеризації та оркестрації допомагає досягти гнучкості у масштабуванні та поліпшує стабільність роботи системи. Аналіз продуктивності показав зменшення затримок обробки запитів й ефективне використання ресурсів серверів.

Розроблена архітектура підтвердила свою ефективність у тестових середовищах і може бути застосована для високонавантажених систем. Перспективи подальших досліджень передбачають інтеграцію штучного інтелекту для автоматичного масштабування й оптимізації роботи сервісів, а також дослідження впливу різних підходів до кешування даних на продуктивність системи.

Ключові слова: архітектура систем, бот, API, NLP.

Класифікація відповідно до AMS 2020: 68P05, 68P10, 68P15, 68P20, 68W32.

Вступ

Зі зростанням популярності соціальних мереж активність бот-акаунтів значно збільшилася. Вони використовуються для маніпулювання громадською думкою, поширення дезінформації та фішингу. Згідно з дослідженнями, значна частина інтернет-трафіку генерується автоматизованими системами, проте шкідливі боти займають 32 % від усього трафіку, згідно з (Imprava, 2024), що містить загрозу інформаційній безпеці.

Існування ботів у соціальних мережах створює низку проблем, пов'язаних із поширенням дезінформації та маніпуляцією громадською думкою. Боти можуть імітувати поведінку реальних користувачів, що ускладнює їхнє виявлення та сприяє неконтрольованому розповсюдженню неправдивої інформації. З розвитком генеративних моделей, створення фейкових новин стало доступнішим, що підсилює проблему дезінформації в соціальних мережах.

Виявлення ботів є складним завданням, що потребує використання різних методів аналізу. Традиційно для цього застосовували аналіз метаданих профілю та лінгвістичні особливості повідомлень. Однак дослідження показують, що аналіз структурних особливостей мережі, зокрема використання структурних убудовувань, може бути ефективнішим у виявленні ботів. Це підтверджує гіпотезу, що локальна соціальна мережа, сформована навколо бот-акаунтів, містить цінну інформацію для їхньої ідентифікації.

Крім того, сучасні системи виявлення ботів стикаються з проблемами масштабованості й узагальненості. Більшість із них базуються на методах машинного навчання з використанням вручну позначених даних, що обмежує їхню здатність виявляти скоординовані дії бот-мереж. Існує потреба в розробленні більш універсальних підходів, що враховують різні аспекти діяльності ботів, включно з аналізом зображень та міжплатформною активністю. Проблему виявлення скоординованих бот-мереж також розглянуто в роботах, присвячених BotNet-детекції (Devle et al., 2021).

Отже, для ефективної боротьби з ботами та пов'язаними з ними загрозами необхідно поєднувати різні методи аналізу й розробляти нові підходи, що враховують сучасні тенденції в розвитку технологій і поведінки ботів у соціальних мережах. Існуючі методи виявлення ботів передбачають:

- **аналіз поведінкових патернів** (частота публікацій, часові інтервали, аномалії);
- **обробку природної мови (NLP)** для аналізу контенту та визначення аномалій у стилістиці;
- **графові методи** (аналіз мережевої структури зв'язків);
- **гібридні підходи**, що комбінують кілька підходів для підвищення точності (Sen et al., 2018).

Останні дослідження (Pote, 2024) та (Antispoofing Wiki, 2024) показують, що методи на основі графових ембедингів забезпечують вищу точність порівняно з традиційними підходами, що ґрунтуються на NLP та мета-даних акаунтів.

Головна мета дослідження – створення масштабованої системи виявлення ботів у соціальних мережах, що здатна працювати в режимі реального часу, забезпечуючи високу точність класифікації. Важливими аспектами системи є:

- підтримка потокової обробки великих обсягів даних,
- використання графових ембедингів для підвищення точності детекції,
- інтеграція із сучасними базами даних і технологіями контейнеризації,
- масштабування системи при збільшенні навантаження.

Перед початком розроблення системи виявлення ботів необхідно визначити основні загрози, типи ботів і методи, які вони використовують для маскування своєї активності. Боти можуть мати різну мету, зокрема:

- **Маніпулювання громадською думкою** – використання ботів для поширення дезінформації, політичної пропаганди або соціального впливу через масове створення та поширення контенту.

- **Фінансове шахрайство** – передбачає накрутку лайків, переглядів, підписників, а також шахрайські схеми із криптовалютними та фінансовими операціями.

- **Кіберзагрози** – використання ботів для фішингових атак, спаму та розповсюдження шкідливого ПЗ.

Сучасні боти дедалі частіше використовують складні техніки маскування своєї активності, що ускладнює їхнє виявлення:

- **Імітація людської поведінки** – нерівномірні інтервали між публікаціями, взаємодія з іншими користувачами, перемикування автоматичних та ручних дій.

- **Динамічна зміна IP-адрес** – використання VPN, проксі-серверів або ботнет-інфраструктури для приховування справжнього місцезнаходження бота.

- **Персоналізовані облікові записи** – використання реальних або зламаних акаунтів, що мають історію активності та друзів.

- **Автоматизація через ШІ** – використання моделей GPT та аналогічних нейромереж для створення правдоподібного контенту, що важко відрізнити від людського. Також варто враховувати можливість ботів генерувати візуальний контент самостійно, наприклад, за допомогою таких моделей, як Stable Diffusion.

Щоб ефективно виявляти ботів, необхідно використовувати кілька взаємопов'язаних підходів, кожен з яких дає змогу ідентифікувати різні характеристики ботів і їхньої активності. Застосування комплексного аналізу допомагає суттєво підвищити точність класифікації ботів. Таким чином ми отримуємо ще одну вимогу до системи – можливість застосування одночасно кількох підходів для аналізу поведінки. Аналітичний сервіс матиме кілька модулів, зокрема: модуль аналізу поведінкових патернів, аналізу текстів та аналізу мережевих зв'язків.

Перший модуль відповідає за аналіз поведінкових патернів користувачів, зокрема часових рядів їхньої активності та виявлення аномалій у публікаціях.

Другий модуль здійснює обробку контенту акаунтів для виявлення ботів на основі мовних характеристик їхніх дописів. Основні напрямки аналізу:

- **лексичний і синтаксичний** – застосування методів обробки природної мови (NLP) для аналізу структури речень, виявлення надмірного використання шаблонних текстів і повторюваних ключових слів (Tzoumanekas et al., 2024);

- **семантичний** – визначення смислових зв'язків між словами та текстами, що допомагає розпізнавати автоматично згенерований контент. Використовуються моделі Word2Vec, BERT, RoBERTa;

- **стилістичний** – оцінювання характерних особливостей текстів, таких як довжина речень, складність граматики, частота використання певних слів. Боти часто демонструють нестандартну стилістику;

- **тональність й емоційний аналіз** – оцінювання емоційної спрямованості текстів (позитивна, нейтральна, негативна) та виявлення шаблонної емоційної поведінки ботів.

Гібридні глибокі нейронні архітектури, що поєднують рекурентні мережі та векторні подання слів (напр., BiGRU-LSTM із використанням GloVe), демонструють високу ефективність у задачах виявлення ботів у соціальних мережах (Ellaky et al., 2024).

Реалізація другого модуля передбачає використання бібліотек NLP, таких як spaCy, Hugging Face Transformers та алгоритмів машинного навчання (LSTM, Transformer-based models).

Третій модуль дає змогу аналізувати структуру соціальних зв'язків користувачів для виявлення бот-мереж. Основні методи передбачають:

- **графовий аналіз** – використання алгоритмів Node2Vec, Struc2Vec для виявлення схожих патернів взаємодії серед акаунтів;

- **ідентифікація координаційних груп** – дослідження поведінки груп акаунтів, що синхронно взаємодіють із певним контентом (масові ретвіти, спільні підписки, одномоментні публікації);

- **моделювання соціальних графів** – аналіз зв'язків між акаунтами з використанням таких метрик, як центральність, кластеризація, коефіцієнт спільних сусідів;

- **виявлення аномальних зв'язків** – визначення акаунтів з аномально високою взаємодією з іншими ботами або великою кількістю однотипних зв'язків.

1. Опис архітектури системи

Архітектура системи виявлення ботів реалізована у вигляді набору взаємодіючих вебсервісів, що дає змогу забезпечити масштабованість, гнучкість і незалежне оновлення компонентів. В основі архітектури лежить кілька ключових сервісів, кожен з яких виконує спеціалізовані функції (рис. 1).

Система починається із сервісу збору даних, що відповідальний за отримання інформації з різних соціальних мереж, таких як Twitter, Telegram та Facebook. Для цього використовують офіційні API-платформ, а також методи вебскрапінгу. Цей сервіс також застосовує попередню фільтрацію, щоб зменшити обсяг оброблюваної інформації.

Далі інформація проходить через сервіс ETL-обробки, що виконує очищення, нормалізацію та трансформацію даних для їх подальшого аналізу. Це передбачає видалення дублікатів (отримання одного повідомлення з різних джерел), стандартизацію форматів й агрегацію даних, що дає змогу підготувати їх для подальшого аналізу.

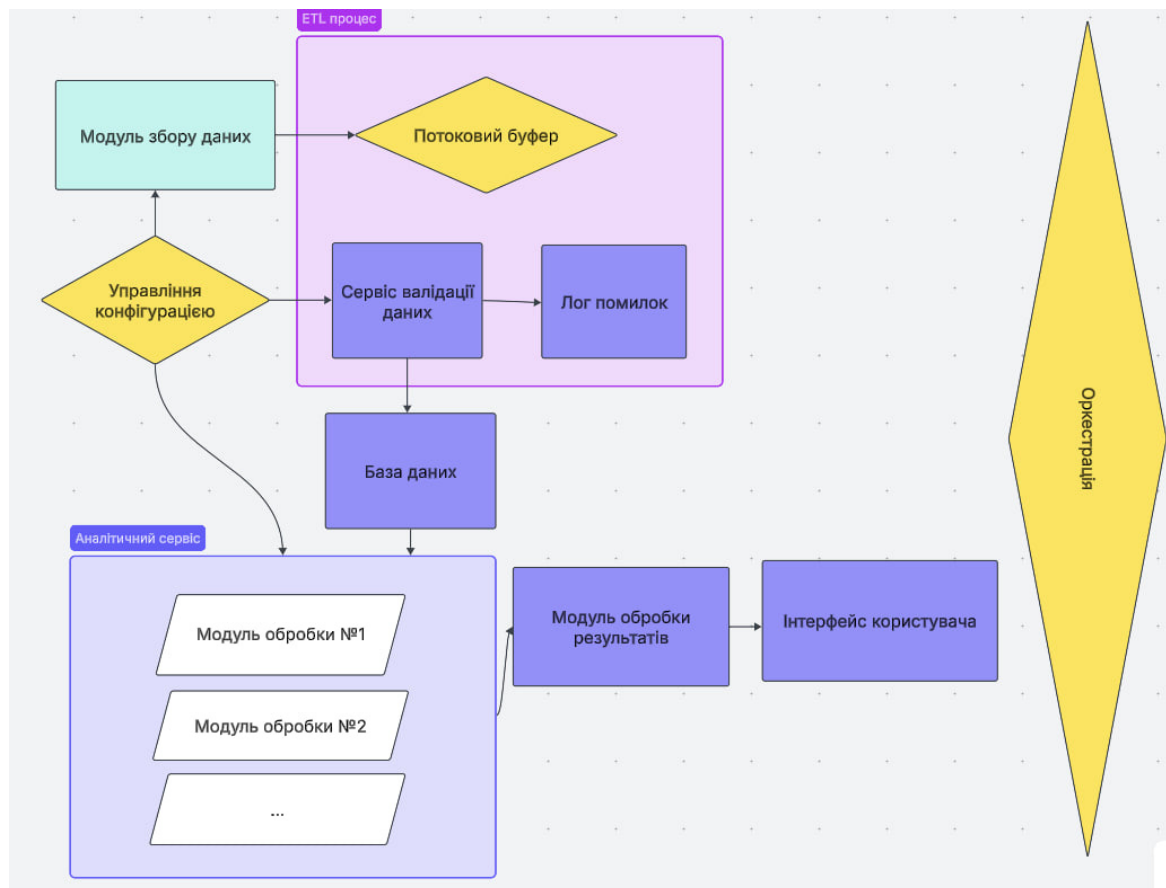


Рис. 1. Загальна схема архітектури системи виявлення ботів у соціальних мережах

Центральним елементом є аналітичний сервіс, який застосовує алгоритми машинного навчання для класифікації ботів. Він складається з декількох основних підмодулів: аналізу поведінкових патернів, аналізу тексту та аналізу мережових зв'язків. Перший модуль аналізує часові ряди активності акаунтів, виявляючи аномалії та повторювані шаблони, другий застосовує NLP-методи для оцінки змісту та стилістики повідомлень, тоді як третій використовує графові алгоритми для визначення підозрілих зв'язків між акаунтами. Ще в одному з модулів аналізуємо, які особливості користувачів можна використовувати для виявлення ботів. Ці особливості розділяємо на чотири групи:

- **NLP та P** – дані профілю й ознаки, отримані з твітів за допомогою обробки природної мови.
- **GF** – прості графові характеристики, такі як центральність вузлів.
- **EMB** – ознаки, отримані із соцмережі за допомогою алгоритмів убудовування (поділяються на класичні та структурні).

Метою цього модуля є не створення нового класифікатора ботів, а порівняння ефективності різних наборів ознак, особливо структурних EMB. Ми показуємо, що боти часто формують локальні соціальні структури, які можна виявити саме за допомогою таких ознак.

Вибір мікросервісної архітектури пояснюють необхідністю забезпечити масштабованість системи та можливість незалежного оновлення окремих компонентів без зупинки роботи всієї інфраструктури. На відміну від монолітних систем, мікросервіси дають змогу швидко адаптуватися до зростання навантаження та спрощують інтеграцію з новими сервісами. Практичні підходи до розгортання масштабованих мікросервісних систем із використанням контейнеризації та оркестрації за допомогою Docker і Kubernetes детально описані в роботі (Lin, 2019).

Використання REST API зумовлене потребою уніфікованої взаємодії між компонентами системи та зовнішніми застосунками. Такий підхід забезпечує простоту інтеграції та високу сумісність із наявними інструментами аналітики.

Поеднання PostgreSQL і DuckDB обрано для оптимізації роботи з різними типами даних: PostgreSQL гарантує надійність і транзакційну цілісність, тоді як DuckDB забезпечує ефективну обробку великих аналітичних запитів.

Elasticsearch інтегровано як інструмент для швидкого пошуку й аналітики в режимі реального часу, що є критично важливим у роботі з високошвидкісними потоками даних.

Для аналітичних модулів використано алгоритми графових убудовувань (Node2Vec, Struc2Vec та ін.), оскільки вони дають змогу автоматично знаходити приховані соціальні структури та взаємозв'язки, що є складним завданням для класичних підходів на основі лише текстового аналізу чи простих статистичних характеристик. Отже, обрана комбінація технологій дає змогу поєднати гнучкість у масштабуванні, надійність зберігання та високу точність виявлення ботів.

Ключові результати:

- Усі чотири набори ознак (NLP, P, GF та EMB) мають предиктивну силу у виявленні ботів.
- Поєднання класичних і структурних ознак підвищує точність моделей.
- Алгоритми EMB можуть автоматично знаходити корисні особливості, які складно створити вручну.
- Структурні EMB мають вищу точність, ніж класичні методи вбудовування.
- Підвищення розмірності EMB незначно впливає на результат, адже навіть низькорозмірні вбудовування ефективні.
- EMB-моделі демонструють стійкість до шуму в соціальній мережі.

Запропонована система базується на принципах Каппа-архітектури, що передбачає потокову обробку даних у реальному часі та використання єдиної обчислювальної моделі для всіх типів обробки. На відміну від Lambda-архітектури, де існує поділ на пакетний і потоковий рівні, Каппа-архітектура спрощує підтримку системи, оскільки усуває дублювання логіки обробки та надає можливість уникнути повторних обчислень. Це особливо важливо для систем виявлення ботів, що працюють із великими потоками даних із соціальних мереж, де швидкість реакції є критичною. Потокова обробка даних у реальному часі в подібних архітектурах зазвичай реалізується з використанням брокерів повідомлень, таких як Apache Kafka, що підтверджується сучасними промисловими практиками (Statsig, 2024). Використання Каппа-архітектури забезпечує масштабованість, спрощує інтеграцію нових джерел даних і підвищує ефективність оброблення без необхідності додаткової пакетної обробки.

Список модулів, що можуть бути інтегрованими в сервіс, детально описані в (Antispoofing Wiki, 2024). Аналітичний сервіс інтегрується з іншими частинами системи через REST API, що забезпечує його масштабованість і легку адаптацію до нових загроз.

Результати аналізу передаються до сервісу обробки інформації та сховища даних, де вони зберігаються у структурованому форматі для подальшого оброблення. Використання реляційних баз даних, таких як PostgreSQL та DuckDB (інтеграція реалізується за допомогою окремого модулю pg_duckdb), забезпечує ефективне зберігання та швидкий доступ до даних, тоді як Elasticsearch дає змогу виконувати пошук й аналітику в реальному часі.

Додатково система використовує MinIO (відкрите S3-сумісне сховище даних) для збереження великих обсягів історичних даних.

Підсумковим результатом аналізу є оцінка ризику, що визначає ймовірність того, що користувач є ботом. Якщо значення оцінки перевищує 95 %, система автоматично відносить акаунт до категорії ботів. Якщо ж оцінка перебуває в діапазоні 70–95 %, то рішення передається людині (оператору системи) для подальшого розгляду. Аналогічно до сучасних систем автопілотів у такій, коли алгоритм може самостійно керувати у стандартних ситуаціях, а у складних випадках підключається оператор, тут система прагне мінімізувати кількість випадків, коли адміністратору необхідно втручатися вручну.

Для взаємодії з користувачами розроблено UI-сервіс, що надає графічний інтерфейс для адміністраторів та аналітиків. Він розроблений на основі сучасного frontend-фреймворку і дозволяє керувати налаштуваннями системи, переглядати звіти й моніторити активність ботів у соціальних мережах.

Інтеграція UI з backend відбувається через REST API. Завдяки поділу на окремі вебсервіси, система підтримує високу продуктивність, масштабованість і можливість безперервного оновлення без зупинення роботи.

Дискусія і висновки

Отримані результати свідчать, що поєднання різних груп ознак (NLP, профільні характеристики, прості графові метрики та структурні вбудовування) забезпечує вищу точність порівняно з використанням окремих методів, що узгоджується з висновками інших дослідників (Dehghan et al., 2023; Tzoumanekas et al., 2024). На відміну від традиційних рішень, які акцентують увагу на аналізі текстів або метаданих (Sen et al., 2018), запропонована архітектура інтегрує кілька підходів одночасно та забезпечує роботу в режимі реального часу. Важливим аспектом є те, що архітектура реалізує принципи Каппа-архітектури, де всі дані обробляються у потоковому режимі, а результати зберігаються для подальшого аналізу. Це дає змогу уникнути дублювання обчислень і спрощує підтримку системи порівняно з Lambda-архітектурою. Подальші дослідження можуть бути спрямовані на оптимізацію апаратних вимог, що допоможе зменшити залежність від високопродуктивного обладнання. Перспективним напрямком є удосконалення моделей машинного навчання, зокрема впровадження сучасних архітектур нейронних мереж для точнішого аналізу тексту й поведінкових даних.

Важливим завданням також є розширення можливостей модулів збору даних для роботи з новими платформами та впровадження методів динамічного оновлення моделей, що забезпечить їхню адаптацію до нових загроз. Реалізація таких удосконалень допоможе підвищити точність виявлення ботів, розширити функціональні можливості системи й інтегрувати її з іншими технологіями та платформами.

Виявлення ботів у соціальних мережах є складним завданням, що вимагає комплексного підходу. Запропонована система на основі Каппа-архітектури забезпечує високу продуктивність, масштабованість і гнучкість.

Подальший розвиток цієї системи дасть змогу ефективніше протидіяти сучасним ботам, підтримуючи безпеку та якість взаємодії у соціальних мережах. Використання передових методів машинного навчання та потокової обробки даних відкриває нові можливості для виявлення ботів навіть у складних умовах.

Внесок авторів: Олександр Борисенко – написання, оригінальна чернетка; Михайло Махно – концептуалізація, методологія; Олексій Федорус – модифікація підходів і редагування; Максим Веремчук – модифікація підходів і редагування.

Джерела фінансування. Публікацію частково профінансовано Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Antispoofing Wiki. (2024). *Detecting Bot-Generated Fake News in Social Media*. Antispoofing Wiki. <https://antispoofing.org/detecting-bot-generated-fake-news-in-social-media/>
- Dehghan, A., Siuta, K., Skorupka, A., Dubey, A., Betlen, A., Miller, D., Xu, W., Kamiński, B., & Pralat, P. (2023). *Detecting bots in social networks using node and structural embeddings*. Journal of Big Data, 10(119). <https://doi.org/10.1186/s40537-023-00796-3>
- Devle, A. C., Jose, J. A., Saraswathula, A. S., Mehta, S., Srivastava, S., Kona, S., & Daggumalli, S. (2021). *BotNet detection on social media*. arXiv. <https://doi.org/10.48550/arXiv.2110.05661>
- Ellaky Z., Benabbou F., Matrane Y., & Qaqa S. (2024). *A Hybrid Deep Learning Architecture for Social Media Bots Detection Based on Bigru-LSTM and Glove Word Embedding*. IEEE Access, 12, 100278-100294. <https://doi.org/10.1109/ACCESS.2024.3430859>
- Imperva. (2024). *Bad Bot Report*. Imperva. <https://www.imperva.com/resources/resource-library/reports/2024-bad-bot-report/>
- Lin J. (2019). *Deploying a scalable web application with Docker and Kubernetes*. Medium. <https://medium.com/better-practices/deploying-a-scalable-web-application-with-docker-and-kubernetes-a5000a06c4e9>
- Pote, M. (2024). *Computational Propaganda Theory and Bot Detection System: Critical Literature Review*. arXiv. <https://doi.org/10.48550/arXiv.2404.05240>
- Sen, I., Aggarwal, A., Mian, S., & Singh, S. (2018). Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram. In Proceedings of the 10th ACM Conference on Web Science, *WebSci 2018: The Reality of Social Media* (pp. 205–209). Association for Computing Machinery. <http://dx.doi.org/10.1145/3201064.3201105>
- Statsig. (2024). *Real-time data processing with Apache Kafka*. Satsig. <https://www.statsig.com/perspectives/real-time-data-processing-with-apache-kafka>
- Tzoumanekas, G., Chatzianastasis, M., Ilias, L., Kiokes, G., Psarras, J., & Askounis, D. (2024). *A Graph Neural Architecture Search Approach for Identifying Bots in Social Media*. Front. Artif. Intell. <https://doi.org/10.48550/arXiv.2411.16285>

References

- Antispoofing Wiki. (2024). *Detecting Bot-Generated Fake News in Social Media*. Antispoofing Wiki. <https://antispoofing.org/detecting-bot-generated-fake-news-in-social-media/>
- Dehghan, A., Siuta, K., Skorupka, A., Dubey, A., Betlen, A., Miller, D., Xu, W., Kamiński, B., & Pralat, P. (2023). *Detecting bots in social networks using node and structural embeddings*. Journal of Big Data, 10(119). <https://doi.org/10.1186/s40537-023-00796-3>
- Devle, A. C., Jose, J. A., Saraswathula, A. S., Mehta, S., Srivastava, S., Kona, S., & Daggumalli, S. (2021). *BotNet detection on social media*. arXiv. <https://doi.org/10.48550/arXiv.2110.05661>
- Ellaky Z., Benabbou F., Matrane Y., & Qaqa S. (2024). *A Hybrid Deep Learning Architecture for Social Media Bots Detection Based on Bigru-LSTM and Glove Word Embedding*. IEEE Access, 12, 100278-100294. <https://doi.org/10.1109/ACCESS.2024.3430859>
- Imperva. (2024). *Bad Bot Report*. Imperva. <https://www.imperva.com/resources/resource-library/reports/2024-bad-bot-report/>
- Lin J. (2019). *Deploying a scalable web application with Docker and Kubernetes*. Medium. <https://medium.com/better-practices/deploying-a-scalable-web-application-with-docker-and-kubernetes-a5000a06c4e9>
- Pote, M. (2024). *Computational Propaganda Theory and Bot Detection System: Critical Literature Review*. arXiv. <https://doi.org/10.48550/arXiv.2404.05240>
- Sen, I., Aggarwal, A., Mian, S., & Singh, S. (2018). Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram. In Proceedings of the 10th ACM Conference on Web Science, *WebSci 2018: The Reality of Social Media* (pp. 205–209). Association for Computing Machinery. <http://dx.doi.org/10.1145/3201064.3201105>
- Statsig. (2024). *Real-time data processing with Apache Kafka*. Satsig. <https://www.statsig.com/perspectives/real-time-data-processing-with-apache-kafka>
- Tzoumanekas, G., Chatzianastasis, M., Ilias, L., Kiokes, G., Psarras, J., & Askounis, D. (2024). *A Graph Neural Architecture Search Approach for Identifying Bots in Social Media*. Front. Artif. Intell. <https://doi.org/10.48550/arXiv.2411.16285>

Отримано редакцією журналу / Received: 31.03.25

Прорецензовано / Revised: 07.10.25

Схвалено до друку / Accepted: 10.10.25

Mykhailo MAKHNO¹, PhD (Engin.) Assoc. Prof.

ORCID ID: 0000-0001-5045-1705

e-mail: mykhailo.makhno@knu.ua

Oleksii FEDORUS¹, PhD (Engin.), Assist.

ORCID ID: 0009-0006-5496-4544

e-mail: oleksii.fedorus@knu.ua

Oleksandr BORYSENKO¹, PhD Student

ORCID ID: 0009-0009-7852-3227

e-mail: borysenko@knu.ua

Maksym VEREMCHUK², PhD Student

ORCID ID: 0009-0003-9787-2787

e-mail: mrveremc@uwaterloo.ca

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

²University of Waterloo, Waterloo, Ontario, Canada

ARCHITECTURE OF A SOCIAL MEDIA BOT DETECTION SYSTEM

Modern information systems require efficient architectures to ensure high performance, scalability, and reliability. This article presents an approach to system architecture design that incorporates the latest technological solutions and methods for optimizing the processing of large data sets. The paper proposes an original architecture of a bot detection system based on the microservices paradigm and modern data processing techniques. Unlike existing solutions, the proposed system does not aim to develop a radically new classification method but focuses on the effective integration of well-established approaches within a unified architecture.

The advancement of information technologies requires the development of architectural solutions that guarantee high performance and reliability of software systems. With the increasing volume of data and growing demands for processing speed, traditional architectural approaches require refinement. Research in this field is important for software developers and system architects.

The aim of this study is to develop an architectural concept that meets modern requirements for performance, scalability, and security. The main objectives include analyzing existing approaches, identifying their advantages and drawbacks, and designing an efficient architecture that minimizes resource consumption and increases data processing speed.

The study employed methods of architectural analysis, system modeling, performance testing, and comparative evaluation of different approaches. For the implementation of the architecture, modern technologies were used, including the microservices paradigm, containerization, and distributed computing.

The proposed architecture improves system performance by optimizing request processing and distributing workloads across services. The use of containerization and orchestration enables flexible scalability and enhances system stability. Performance analysis has shown reduced request processing latency and efficient utilization of server resources.

The developed architecture has proven its effectiveness in test environments and can be applied to high-load systems. Future research directions include the integration of artificial intelligence for automatic scaling and service optimization, as well as studying the impact of different caching strategies on overall system performance.

Keywords: *system architecture, bot, API, NLP.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 004.89

DOI: <https://doi.org/10.17721/1812-5409.2025/2.30>

Руслан ПРАВОСУД, асп.

ORCID ID: 0009-0008-9450-5515

e-mail: pronod9999@gmail.com

Київський національний університет імені Тараса Шевченка, Київ, Україна

Олександр МАРЧЕНКО, д-р фіз.-мат. наук, проф.

ORCID ID: 0000-0002-5408-5279

e-mail: omarchenko@univ.kiev.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

ПЕРЕНЕСЕННЯ СТИЛЮ ДІАЛОГУ ЗА ДОПОМОГОЮ ПАРАМЕТРИЧНО ЕФЕКТИВНОГО ДОНАВЧАННЯ

Розглянуто застосування сучасних методів донавчання, що характеризуються високою параметричною ефективністю, до задачі генерації діалогових реплік у стилі певного вигаданого або реального персонажа. Основна мета дослідження полягає в розробленні підходу, що допомагає адаптувати вже наявні мовні моделі до специфічного стилю спілкування або мовлення, властивого конкретній особистості, з використанням мінімальної кількості нових параметрів. Це дає змогу уникнути повного перенавчання великих моделей і зменшити обчислювальні витрати.

Однією із ключових цілей є можливість реалізації компактної, але функціональної мовної моделі, яку можна буде запускати на окремому комп'ютері без необхідності використання серверів чи потужних графічних процесорів. Такий підхід відкриває нові можливості для персоналізованих або розважальних застосувань, зокрема створення чат-ботів, які можуть взаємодіяти з користувачами в образі певного персонажа з книги, фільму чи історичної особи.

Для реалізації такого підходу запропоновано використання методу LoRA (Low-Rank Adaptation), що допомагає ефективно донавчати наявні трансформерні архітектури – зокрема модель GPT-2 – без повного перенавчання її всіх шарів. Застосування LoRA дає змогу додавати стилістичні особливості до мовлення моделі, не втрачаючи її базових можливостей генерувати логічні, зв'язні тексти.

Щоб перевірити, наскільки добре модель навчилася імітувати цільовий стиль мовлення, ми додатково тренуємо модель BERT як класифікатор, що розрізняє тексти в бажаному стилі та звичайні, нейтральні репліки. У такий спосіб ми отримуємо об'єктивну метрику якості генерації, що надає можливість оцінити ефективність донавчання.

Набір даних, який використано для навчання, було згенеровано за допомогою моделі GPT4-mini, що дає змогу швидко створити велику кількість прикладів у потрібному стилі, забезпечуючи достатній обсяг даних для ефективного навчання та валідації результатів.

Ключові слова: обробка природних мов, параметрично ефективне донавчання, низькорангова адаптація, генерація тексту.

Класифікація відповідно до AMS 2020: 68T50, 68T07.

Вступ

Упродовж останніх років великі мовні моделі (LLMs) продемонстрували значний прогрес у генерації природномовного тексту, що знаходить застосування у створенні діалогових систем, художніх історій і персоналізованого контенту. Попри це, налаштування таких моделей на специфічні задачі або стилі часто вимагає значних обчислювальних потужностей, що ускладнює їхнє використання на пристроях з обмеженими ресурсами. Отже, постає актуальне завдання – здійснити ефективне стилістичне донавчання без суттєвих витрат ресурсів і збереженням високої якості результатів.

Завдання генерації стилістично узгодженого тексту широко досліджується вченими в галузі обробки природної мови (NLP), й останнім часом у цій сфері досягнуто значного прогресу завдяки розвитку масштабних попередньо натренованих мовних моделей і методів донавчання.

Перенесення стилю в NLP спрямоване на трансформацію тексту так, щоб зберігався його зміст, але змінювалися стилістичні характеристики. Перші підходи до цієї задачі базувалися на системах із правилами та ручному доборі ознак для кодування стилістичних властивостей. З появою глибокого навчання почали застосовувати варіаційні автокодери (VAE) та генеративно-змагальні мережі (GAN) для перенесення стилю в текстах.

Сучасні методи дедалі частіше використовують попередньо навчені мовні моделі, такі як GPT-2 та GPT-3, для стилістичної генерації. Наприклад, у роботі (Keskar et al., 2019) було показано, як використання керувальних кодів у моделях на базі GPT дає змогу спрямовувати генерацію в заданому стилістичному напрямку. Однак такі методи зазвичай потребують значних обчислювальних ресурсів, що обмежує їхнє застосування в середовищах з обмеженими можливостями. Наша робота усуває це обмеження шляхом використання параметрично ефективного донавчання для досягнення стилістичної узгодженості з мінімальними витратами ресурсів.

Традиційне донавчання великих мовних моделей зазвичай передбачає оновлення всіх параметрів, що є надзвичайно ресурсоємним процесом. Сучасні підходи до параметрично ефективного донавчання, зокрема AdapterFusion (Pfeiffer et al., 2021) та LoRA (Hu et al., 2021), розв'язують цю проблему шляхом упровадження легких модулів з параметрами, орієнтованих на конкретні задачі. Ці методи допомагають адаптувати попередньо навчені моделі до нових завдань із мінімальними витратами ресурсів й обсягом пам'яті.

Особливої уваги заслуговує LoRA, що поєднує простоту реалізації з високою ефективністю в донавчанні масштабних моделей. Завдяки вбудовуванню матриць низького рангу в механізми уваги моделі LoRA дає змогу ефективно адаптувати поведінку без змін основних ваг моделі.

Створення якісних наборів даних часто є вузьким місцем у виконанні завдань стилістичного донавчання. Низка сучасних досліджень розглядає використання великих моделей, таких як GPT-3 або GPT-4, для генерації синтетичних

© Правосуд Руслан, Марченко Олександр, 2025

датасетів. Наприклад, у дослідженні (Schick, & Schütze, 2021) показано, що синтетичні дані можуть ефективно доповнювати навчання у випадках з обмеженими ресурсами. У нашій роботі ми використовуємо модель GPT4-mini для генерації різноманітного і якісного набору даних, що допомагає зменшити потребу в ручній розмітці та забезпечити стилістичну узгодженість

Імітація мовлення конкретного персонажа відкриває нові можливості в галузях розваг, освіти й інтерфейсів "людина – комп'ютер". Наприклад, чат-боти з індивідуальним стилем, ігрові рольові сценарії чи генерація авторського тексту потребують здатності моделі генерувати стилістично відповідні та логічні репліки. Проте традиційні методи донавчання передбачають оновлення всієї моделі, що є надто ресурсозатратним.

Щоб усунути цю проблему, ми застосовуємо алгоритм Low-Rank Adaptation (LoRA) для адаптації попередньо навченої моделі GPT-2. Це дає можливість змінити поведінку моделі в бажаному напрямку, мінімізуючи кількість параметрів, які потрібно донавчати.

Для оцінки якості згенерованих відповідей ми використовуємо класифікатор на основі BERT, що навчається розпізнавати тексти в заданому стилі. Для створення навчального набору даних застосовуємо GPT4-mini, що дає змогу отримати різноманітні приклади високої якості.

Основні здобутки цієї роботи:

- **Ефективне донавчання з мінімальними ресурсами:** завдяки використанню LoRA ми зменшуємо обчислювальні витрати та забезпечуємо можливість запуску на менш потужних пристроях.

- **Практична придатність для сценаріїв, орієнтованих на персонажі:** запропонований підхід довів свою ефективність у створенні реплік, які відповідають стилю певної особи, що робить його корисним для історій, чат-ботів і синтетичних датасетів.

1. Методи

У процесі дослідження були використані методи аналізу, порівняння й експерименту.

2. Основні означення та результати

Модель для оцінки стилю

Підготовка датасету. Датасет для навчання моделі для оцінки стилю складається із двох частин: датасету філософських діалогів (Hypersniper, 2023) (близько 40 %), а також філософських питань, згенерованих GPT4-mini. Цей датасет містить приблизно 1000 філософських питань. Кожне питання згодом використовували як вхідний запит для GPT4-mini для генерації двох типів відповідей:

1. Відповідь у стилі філософа Сократа.

2. Відповідь з позиції звичайної людини.

Згенеровані пари були оброблені у такому форматі:

Людина: <питання>

Сократ: <відповідь>

Ці структуровані пари забезпечили навчальні дані для розрізнення стилістичних відмінностей між сократівськими та звичайними висловлюваннями.

Навчання моделі. Модель (Sanh, 2019) була донавчена для класифікації висловлювань як сократівських або звичайних шляхом присвоєння винагород. Позитивна винагорода очікувалася для відповідей у стилі Сократа, тоді як звичайним відповідям призначалася негативна винагорода.

Оцінювання моделі. Модель винагороди була оцінена за допомогою тестового датасету, що становив 20 % від початкового датасету. Результати показали:

- Середня винагорода понад 3.0 для відповідей у стилі Сократа.

- Середня винагорода менше -3.0 для звичайних відповідей.

- Точність 99,5 % у розрізненні сократівських і звичайних відповідей шляхом порівняння їхніх винагород.

Основна модель. У традиційному донавчанні всі або більшість параметрів моделі оновлюються. Натомість LoRA фіксує всі основні параметри моделі та додає невеликі додаткові низькорівневі (low-rank) матриці до певних компонентів трансформера – зазвичай до матриць запитів (Q), ключів (K), або значень (V) у механізмі самоуваги (self-attention). Замість оновлення великих оригінальних матриць, LoRA вставляє два додаткові шари: матрицю зниження розмірності. Під час донавчання оновлюються лише ці шари, а оригінальна модель залишається незмінною.

Попередньо натреновані моделі GPT2-large (Radford et al., 2019) (800 млн параметрів) та GPT2 (125 млн параметрів) були поліпшені за допомогою LoRA (Low-Rank Adaptation) шляхом додавання додаткових низькорівневих матриць для компонентів Q, K і V у шарах трансформера з рангом 32. Ці додаткові шари були єдиними, що підлягали донавчанню, і становили менше 1 % від загальної кількості параметрів моделі: 5,898,240 з 779,928,320 для GPT2-large та 1,179,648 з 125,620,225 для GPT2 відповідно.

Донавчання. Щоб краще відтворювати бажаний стиль мовлення, спершу ми провели стандартне донавчання моделей, мінімізуючи перехресну ентропійну втрату.

Як тренувальний датасет для цього підзавдання використовували діалоги з однієї репліки із Сократом у форматі:

Людина: <питання>

Сократ: <відповідь>

Після цього ми оцінили модель на тестовому датасеті, досягнувши середньої винагороди 3.123216 зі стандартним відхиленням 0.068859 для GPT2-Large та середньої винагороди 3.120361 зі стандартним відхиленням 0.074097 для GPT2, що досить суттєво відносно нагород, отриманих на моделі без донавчання (табл. 1).

Для оцінки зв'язності між собою за змістом питань та відповідей, а також відмінності між собою згенерованих відповідей, було використано BERTScore (Zhang et al., 2019). Результати наведено в табл. 2 – питання та відповіді мають відчутно більшу семантичну спорідненість, ніж пара довільних відповідей, що свідчить про різноманітність згенерованих реплік і зв'язність у межах поточного діалогу.

Таблиця 1

Порівняння моделей за стилістичною нагородою		
Модель	Середнє значення метрики	Середнє відхилення метрики
GPT2-Large без донавчання	0.53929	0.69312
GPT2-Large з донавчанням	3.123216	0.068859
GPT2 без донавчання	0.48294	0.739292
GPT2 з донавчанням	3.120361	0.074097

Таблиця 2

Порівняння семантичної подібності відповідей		
Модель	Середнє значення BERTScore	Середнє відхилення BERTScore
GPT2-Large питання vs відповідь	0.1896	0.1137
GPT2-Large питання vs інші питання	0.1432	0.0912
GPT2 питання vs відповідь	0.1753	0.1174
GPT2 питання vs інші питання	0.1487	0.0875

Приклад діалогу згенерованого GPT2 з донавчанням:

Person: What form of art resonates with you the most?

Socrates: Ah, my dear interlocutor, the art of expression has deep roots in Western philosophy, tradition, and human experience. To appreciate its nuances, I suggest you ponder what art embodies the highest ideals about what we seek out and strive for.

Person: I find that music resonates with me the most; its ability to evoke emotions and connect human experiences transcends words and speaks to the soul. It captures the essence of life's joys and sorrows in a way that feels both universal and deeply personal.

Socrates: Ah, indeed! Music is a mirror of the human experience. Tell me, what art has shaped your understanding of this realm? What music instills a sense of unity, unity in moments. What can you imagine to be the essence of music in the presence of the divine?

Дискусія і висновки

У дослідженні представлено ресурсоефективний підхід до навчання мовних моделей для генерації діалогів у стилі певного персонажа за допомогою параметроефективного налаштування. Запропонований метод використовує алгоритм Low-Rank Adaptation (LoRA) для налаштування моделей GPT-2 Large та GPT-2 за мінімальних обчислювальних витрат. Метрику оцінено з використанням класифікатора на основі DistilBERT для перевірки, що модель постійно генерує результати, узгоджені з бажаним стилем.

Представлена методологія може бути застосована в різних сферах, зокрема:

- чат-боти з особистістю персонажа;
- створення діалогових систем, що імітують історичних осіб, вигаданих персонажів або відомі особистості;
- генерація наративів, узгоджених з характером персонажів, для художніх творів;
- створення стилістично різноманітних наборів даних для навчання або оцінки моделей.

Напрями для подальших досліджень:

- розширення методу для підтримки кількох персонажів або стилів одночасно;
- використання навчання з підкріпленням для донавчання моделі;
- інтеграція зворотного зв'язку від людей або альтернативних механізмів оцінки для покращення відповідності;
- дослідження масштабованості підходу з використанням менш потужних попередньо навчених моделей, таких як дистильовані або квантизовані.

Отже, представлена робота демонструє, що параметроефективне налаштування є практичним рішенням для стилістичної генерації тексту, що сприяє ресурсозберігаючому впровадженню моделей обробки природної мови з особистісною орієнтацією. Отримані результати підкреслюють потенціал цього підходу для розвитку персоналізованого ШІ та генерації креативного контенту.

Внесок авторів: Олександр Марченко – концептуалізація; методика; Руслан Правосуд – програмне забезпечення, збір і перевірка емпіричних даних, емпіричне дослідження, аналіз джерел, підготовка огляду літератури та теоретичних основ дослідження.

Джерела фінансування. Публікацію частково профінансовано Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Hu, Edward J., Shen, Yelong, Wallis, Phillip, Allen-Zhu, Zeyuan, Yuanzhi, Li, Wang, Shean, Lu, Wang, & Chen, Weizhu. (2021). *Lora: Low-rank adaptation of large language models*. arXiv. <https://doi.org/10.48550/arXiv.2106.09685>
- Hypersniper. *Philosophy dialog* [Dataset]. Hugging Face. https://huggingface.co/datasets/Hypersniper/philosophy_dialogue
- Keskar, Nitish Shirish, McCann, Bryan, Varshney, Lav R., Caiming, Xiong, & Socher, Richard. (2019). *Ctrl: A conditional transformer language model for controllable generation*. arXiv. <https://doi.org/10.48550/arXiv.1909.05858>
- Pfeiffer, Jonas, Kamath, Aishwarya, Rücklé, Andreas, Kyunghyun, Cho, & Gurevych, Iryna. (2020). *Adapterfusion: Non-destructive task composition for transfer learning*. arXiv. <https://doi.org/10.48550/arXiv.2005.00247>
- Radford, Alec, Wu, Jeffrey, Child, Rewon, Luan, David, Amodei, Dario, & Sutskever, Ilya. (2019). *Language models are unsupervised multitask learners*. OpenAI blog. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
- Sanh, V. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. arXiv. <https://doi.org/10.48550/arXiv.1910.01108>

- Schick, Timo, & Schütze, Hinrich. (2020). *It's not just size that matters: Small language models are also few-shot learners*. arXiv. <https://doi.org/10.48550/arXiv.2009.07118>
- Zhang, Tianyi, Kishore, Varsha, Wu, Felix, Weinberger, Kilian Q., Artzi Yoav. (2019). Bertscore: Evaluating text generation with bert. arXiv. <https://doi.org/10.48550/arXiv.1904.09675>

References

- Hu, Edward J., Shen, Yelong, Wallis, Phillip, Allen-Zhu, Zeyuan, Yuanzhi, Li, Wang, Shean, Lu, Wang, & Chen, Weizhu. (2021). *Lora: Low-rank adaptation of large language models*. arXiv. <https://doi.org/10.48550/arXiv.2106.09685>
- Hypersniper. *Philosophy dialog* [Dataset]. Hugging Face. https://huggingface.co/datasets/Hypersniper/philosophy_dialogue
- Keskar, Nitish Shirish, McCann, Bryan, Varshney, Lav R., Caiming, Xiong, & Socher, Richard. (2019). *Ctrl: A conditional transformer language model for controllable generation*. arXiv. <https://doi.org/10.48550/arXiv.1909.05858>
- Pfeiffer, Jonas, Kamath, Aishwarya, Rücklé, Andreas, Kyunghyun, Cho, & Gurevych, Iryna. (2020). *Adapterfusion: Non-destructive task composition for transfer learning*. arXiv. <https://doi.org/10.48550/arXiv.2005.00247>
- Radford, Alec, Wu, Jeffrey, Child, Rewon, Luan, David, Amodei, Dario, & Sutskever, Ilya. (2019). *Language models are unsupervised multitask learners*. OpenAI blog. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
- Sanh, V. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. arXiv. <https://doi.org/10.48550/arXiv.1910.01108>
- Schick, Timo, & Schütze, Hinrich. (2020). *It's not just size that matters: Small language models are also few-shot learners*. arXiv. <https://doi.org/10.48550/arXiv.2009.07118>
- Zhang, Tianyi, Kishore, Varsha, Wu, Felix, Weinberger, Kilian Q., Artzi Yoav. (2019). Bertscore: Evaluating text generation with bert. arXiv. <https://doi.org/10.48550/arXiv.1904.09675>

Отримано редакцією журналу / Received: 25.05.25
Прорецензовано / Revised: 17.09.25
Схвалено до друку / Accepted: 10.10.25

Ruslan PRAVOSUD, PhD Student
ORCID ID: 0009-0008-9450-5515
e-mail: pronod9999@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Oleksandr MARCHENKO, DSc (Phys. & Math.), Prof.
ORCID ID: 0000-0002-5408-5279
e-mail: omarchenko@univ.kiev.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

DIALOG STYLE TRANSFER WITH PARAMETER-EFFICIENT FINE-TUNING

This work explores the application of modern fine-tuning methods, characterized by high parameter efficiency, to the task of generating dialogue responses in the style of a specific fictional or real-life character. The main goal of the study is to develop an approach that allows existing language models to be adapted to the specific communication or speech style of a particular personality using a minimal number of new parameters. This avoids the need for full retraining of large models and reduces computational costs.

One of the key objectives is to enable the implementation of a compact yet functional language model that can be run on a standalone computer without requiring servers or powerful graphics processors. Such an approach opens new possibilities for personalized or entertainment applications, such as creating chatbots capable of interacting with users in the persona of a character from a book, movie, or even a historical figure.

To implement this approach, we propose the use of the LoRA (Low-Rank Adaptation) method, which allows efficient fine-tuning of existing transformer architectures – in particular, the GPT-2 model – without retraining all of its layers. Using LoRA makes it possible to add stylistic features to the model's speech without compromising its core ability to generate logical and coherent texts.

To evaluate how well the model has learned to imitate the target speech style, we additionally train a BERT model as a classifier to distinguish between texts written in the desired style and ordinary, neutral responses. This provides an objective quality metric for generation and allows us to assess the effectiveness of the fine-tuning.

The dataset used for training was generated using the GPT-4 mini model, which enables the rapid creation of a large number of examples in the desired style.

Key words: *natural language processing, parameter-efficient fine-tuning, low-rank adaptation, text generation.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

МЕТОД НАЙМЕНШИХ КВАДРАТІВ ВІДНОСНИХ ПОХИБОК ДЛЯ ОЦІНЮВАННЯ ПАРАМЕТРІВ СИСТЕМ ЗА НЕКЛАСИЧНИХ ПРИПУЩЕНЬ

Розглянуто задачу оптимального оцінювання параметрів регресійної моделі за доступними спостереженнями залежної змінної й усіх регресорів. Запропоновано для розв'язання цієї проблеми використовувати метод найменших квадратів відносних похибок. Оцінка цього методу – це точка, у якій досягає свого мінімуму сума квадратів відносних похибок регресійної моделі, а не сума квадратів абсолютних похибок моделі, як у звичайному методі найменших квадратів, який можна назвати методом найменших квадратів абсолютних похибок. Значення функціонала якості розглянутого методу вже не залежить від одиниць виміру доступних спостережень.

У роботі явна форма представлення оптимальної оцінки методу найменших квадратів відносних похибок у ситуації, коли справедливо класичне припущення, що гарантує її єдиність, поширена на випадок, коли це припущення порушено й оцінка вже не є єдиною. В останній ситуації виникає потреба у використанні оператора псевдообернення матриць за Муром – Пенроузом. Також для обох випадків представлені відповідні явні вирази для похибки оцінювання цим методом.

Крім цього, як наслідок для зваженого методу найменших квадратів, у випадку використання як вагової матриці довільної позитивно визначеної матриці, для обох згаданих ситуацій за аналогією також наведено явні форми представлення відповідної оптимальної оцінки цього методу та виразу для похибки оцінювання.

Ключові слова: регресійна модель, оцінювання параметрів, неklasичне припущення, оператор псевдообернення за Муром – Пенроузом, метод найменших квадратів, зважений метод найменших квадратів, метод найменших квадратів відносних похибок.

Класифікація відповідно до AMS 2020: 93E24, 62H12.

Вступ

Ефективне дослідження різних складних класів систем важко уявити без конструювання для них відповідних адекватних математичних моделей. Причому, якщо маємо справу із задачею ідентифікації об'єкта у вузькому розумінні, то доступна інформація про математичну модель буде відома з точністю до деяких невідомих параметрів, тобто проблема конструювання потрібної моделі зводиться до розв'язання задачі знаходження оцінок цих параметрів моделі, оптимальних у деякому розумінні.

Один із поширених підходів до розв'язання згаданої проблеми ідентифікації параметрів математичної моделі полягає у використанні відомого методу найменших квадратів (МНК) (Gauss, 1809; Legendre, 1805). Оцінка МНК була досліджена як у ситуації справедливості класичного припущення, що гарантує її єдиність, так і в ситуації його можливого порушення (Albert, 1972; Björck, 1996; Hansen et al., 2013; Rao et al., 2008; Weisberg, 2014). В останній ситуації виникає потреба звернутися до використання оператора псевдообернення за Муром – Пенроузом (Moore, 1920; Penrose, 1955).

Відомо, що оцінкою звичайного МНК є точка, у якій досягає свого мінімуму сума квадратів похибок моделі, точніше сума квадратів абсолютних похибок моделі. У подальших модифікаціях МНК перейшли до мінімізації квадратичних форм від вектора абсолютних похибок моделі з ваговими позитивно визначеними матрицями спеціальної структури (Rao et al., 2008; Weisberg, 2014).

У представленій роботі запропоновано для оцінки параметрів регресійної моделі використовувати МНК відносних похибок (Chengsi, 2001; Tofallis, 2008), оптимальна оцінка якого мінімізує суму квадратів відносних похибок моделі на відміну від оцінки звичайного МНК, що базується на мінімізації суми квадратів абсолютних похибок моделі. У публікаціях (Chengsi, 2001; Tofallis, 2008) для цієї моделі оцінка МНК відносних похибок аналізується тільки у випадку, коли справедливо класичне припущення, що гарантує її єдиність. У дослідженні цей результат поширено на випадок, коли це класичне припущення може бути порушено. Також для обох випадків представлені необхідні вирази для похибок оцінювання МНК відносних похибок.

Як наслідок, за аналогією, наведені відповідні результати у всіх згаданих випадках для оцінки зваженого методу найменших квадратів (ЗМНК) з довільною позитивно визначеною ваговою матрицею.

1. Основні результати

Розглянемо задачу знаходження оцінки вектора невідомих стаціонарних параметрів α для регресійної моделі

$$y(k) = x^T(k)\alpha + e(k), \quad k = \overline{1, N}, \quad (1)$$

де $y(k)$ – відомі скалярні спостереження над залежною змінною, $x(k)$ – доступні вектори регресорів, $\alpha \in \mathbb{R}^p$, $e(k)$ – похибки моделі.

Необхідно за отриманими вимірами $\{y(k), x(k)\}_{k=1}^N$ знайти оптимальну оцінку вектора параметрів α системи (1). Оптимальність будемо розуміти як у роботах (Chengsi, 2001; Tofallis, 2008).

Означення. Нехай для об'єкта (1) справедливо $y(k) \neq 0$, $k = \overline{1, N}$. Оцінкою вектора невідомих параметрів α для системи (1) методом найменших квадратів відносних похибок є оцінка $\hat{\alpha}$, на якій досягається мінімум такого функціонала

$$Q(\alpha) = \sum_{k=1}^N \left(\frac{e(k)}{y(k)} \right)^2. \quad (2)$$

Потрібно підкреслити інваріантність значення критерію оптимальності (2) відносно одиниць виміру доступних спостережень $\{y(k), x(k)\}_{k=1}^N$.

Перепишемо систему рівнянь (1) у матричному вигляді

$$y = X\alpha + e,$$

де
$$y = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}, \quad X = \begin{pmatrix} x^T(1) \\ x^T(2) \\ \vdots \\ x^T(N) \end{pmatrix} \in M_{N,p}(\mathbb{R}), \quad e = \begin{pmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{pmatrix}.$$

Теорема 1. Оцінка $\hat{\alpha}$ вектора невідомих параметрів α моделі (1) методом найменших квадратів відносних похибок буде:

1) єдиною тоді і тільки тоді, коли $\text{rank}(X) = p$, і має вигляд

$$\hat{\alpha} = (X^T D X)^{-1} X^T D y, \quad (3)$$

причому її похибка оцінювання

$$\Delta(\hat{\alpha}) = \hat{\alpha} - \alpha = (X^T D X)^{-1} X^T D e; \quad (4)$$

2) не єдиною тоді і тільки тоді, коли $\text{rank}(X) < p$, а множина усіх цих оцінок буде задаватися згідно з

$$\hat{\alpha}(c) = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} y + \left[E_p - \left(D^{\frac{1}{2}} X \right)^+ \left(D^{\frac{1}{2}} X \right) \right] c, \quad c \in \mathbb{R}^p, \quad (5)$$

причому відповідна похибка оцінювання

$$\Delta(\hat{\alpha}(c)) = \hat{\alpha}(c) - \alpha = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} e + \left[E_p - \left(D^{\frac{1}{2}} X \right)^+ \left(D^{\frac{1}{2}} X \right) \right] (c - \alpha), \quad c \in \mathbb{R}^p. \quad (6)$$

Найменшу норму серед усіх оцінок (5) буде мати оцінка

$$\hat{\alpha} = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} y,$$

причому формула останньої оцінки суттєво спрощується, якщо матриця X має повний ранг по рядках ($\text{rank}(X) = N$) і набуває вигляду

$$\hat{\alpha} = X^T (X X^T)^{-1} y.$$

Тут $D = \text{diag}(y^{-2}(1), y^{-2}(2), \dots, y^{-2}(N))$, $(\cdot)^+$ – оператор псевдообернення за Муром – Пенроузом, E_p – одинична матриця порядку p .

Доведення. Очевидно, що оцінка $\hat{\alpha}$ методу найменших квадратів відносних похибок по структурі є частинним випадком оцінки зваженого методу найменших квадратів (ЗМНК) (Björck, 1996; Rao et al., 2008; Strutz, 2016), на якій досягається мінімум функціонала

$$e^T W e, \quad W > 0,$$

у ситуації, коли як позитивно визначену матрицю W взято діагональну матрицю D .

Це дає змогу спростити аналіз оцінки методу найменших квадратів відносних похибок $\hat{\alpha}$. Розглянемо послідовно обидві ситуації.

1) Припустимо, що $\text{rank}(X) = p$. Очевидно, що ця умова буде необхідною та достатньою умовою єдиності оцінки МНК відносних похибок $\hat{\alpha}$, а вигляд безпосередньо оцінки $\hat{\alpha}$ та її похибки оцінювання $\Delta(\hat{\alpha})$ набудуть представлення (3) та (4) відповідно, як частинні випадки відповідних формул для оцінки ЗМНК (Björck, 1996; Rao et al., 2008; Strutz, 2016), коли вагова матриця $W = D$. Відмітимо, що оцінка $\hat{\alpha}$ для цього випадку у вигляді (3) була вже отримана у роботі (Tofallis, 2008).

2) Нехай $\text{rank}(X) < p$. Зрозуміло, що в цьому випадку оцінка МНК відносних похибок не буде єдиною. Спочатку критерій (2) трансформуємо до вигляду

$$Q(\alpha) = \sum_{k=1}^N \left(\frac{e(k)}{y(k)} \right)^2 = e^T D e = \left\| D^{\frac{1}{2}} y - D^{\frac{1}{2}} X \alpha \right\|^2,$$

де $\|\cdot\|$ – евклідова норма.

Останнє дає змогу стверджувати, що множина усіх оцінок МНК відносних похибок збігається із множиною всіх псевдорозв'язків відносно α такої системи лінійних алгебраїчних рівнянь (Чарін, 2004)

$$D^{\frac{1}{2}} X \alpha = D^{\frac{1}{2}} y,$$

і задається так (Albert, 1972):

$$\hat{\alpha}(c) = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} y + \left[E_p - \left(D^{\frac{1}{2}} X \right)^+ \left(D^{\frac{1}{2}} X \right) \right] c, \quad c \in \mathbb{R}^p.$$

А відповідна похибка оцінювання

$$\begin{aligned} \Delta(\hat{\alpha}(c)) &= \hat{\alpha}(c) - \alpha = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} y + \left[E_p - \left(D^{\frac{1}{2}} X \right)^+ \left(D^{\frac{1}{2}} X \right) \right] c - \alpha = \\ &= \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} e + \left[E_p - \left(D^{\frac{1}{2}} X \right)^+ \left(D^{\frac{1}{2}} X \right) \right] (c - \alpha), \quad c \in \mathbb{R}^p. \end{aligned}$$

Серед усіх оцінок МНК відносних похибок на множині (5), згідно з (Albert, 1972), найменшу норму буде мати оцінка

$$\hat{\alpha} = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} y.$$

Проаналізуємо останню оцінку у випадку, коли матриця X має повний ранг по рядках. Нехай $\text{rank}(X) = N$, тоді зрозуміло, що і $\text{rank} \left(D^{\frac{1}{2}} X \right) = N$ також. Оскільки $D^{\frac{1}{2}} X$ теж має повний ранг по рядках, то її псевдообернена матриця легко підраховується та згідно з (Albert, 1972) набуває вигляду

$$\left(D^{\frac{1}{2}} X \right)^+ = \left(D^{\frac{1}{2}} X \right)^T \left[\left(D^{\frac{1}{2}} X \right) \left(D^{\frac{1}{2}} X \right)^T \right]^{-1} = X^T D^{\frac{1}{2}} \left[D^{\frac{1}{2}} (X X^T) D^{\frac{1}{2}} \right]^{-1} = X^T (X X^T)^{-1} D^{-\frac{1}{2}}.$$

Тоді

$$\hat{\alpha} = \left(D^{\frac{1}{2}} X \right)^+ D^{\frac{1}{2}} y = X^T (X X^T)^{-1} D^{-\frac{1}{2}} D^{\frac{1}{2}} y = X^T (X X^T)^{-1} y.$$

Доведення завершено.

Зауваження. Підкреслимо, що у випадку, коли $\text{rank}(X) = p$, також можна користуватися формулами (5) та (6), бо у цьому випадку результати (3) та (4) є частинними випадками формул (5) та (6).

З доведення останньої теореми випливає, що вона справджується і тоді, коли як оцінку вектора невідомих параметрів α моделі (1) замість оцінки методу найменших квадратів відносних похибок $\hat{\alpha}$ використати оцінку ЗМНК $\hat{\alpha}_W$ з довільною позитивно визначеною ваговою матрицею W . Оцінка $\hat{\alpha}_W$ у випадку, коли справедливе класичне припущення, що гарантує її єдиність, була досліджена у роботах (Björck, 1996; Rao et al., 2008; Strutz, 2016), а останній висновок дозволяє доповнити згадане дослідження результатом у випадку, коли це класичне припущення порушується. А саме, справедливе таке твердження.

Теорема 2. Нехай $\hat{\alpha}_W$ є оцінкою зваженого методу найменших квадратів для вектора невідомих параметрів α моделі (1), тобто на ній досягає мінімуму функціонал

$$e^T W e, \quad W > 0,$$

де W – довільна позитивно визначена матриця порядку N .

Тоді оцінка зваженого методу найменших квадратів $\hat{\alpha}_W$ буде:

1) єдиною тоді і тільки тоді, коли $\text{rank}(X) = p$, і має вигляд

$$\hat{\alpha}_W = (X^T W X)^{-1} X^T W y,$$

причому її похибка оцінювання

$$\Delta(\hat{\alpha}_W) = \hat{\alpha}_W - \alpha = (X^T W X)^{-1} X^T W e;$$

2) не єдиною тоді і тільки тоді, коли $\text{rank}(X) < p$, а множина усіх цих оцінок буде задаватися згідно з

$$\hat{\alpha}_W(c) = \left(\frac{1}{W^2} X \right)^+ \frac{1}{W^2} y + \left[E_p - \left(\frac{1}{W^2} X \right)^+ \left(\frac{1}{W^2} X \right) \right] c, \quad c \in \mathbb{R}^p, \quad (7)$$

причому відповідна похибка оцінювання набуває вигляду

$$\Delta(\hat{\alpha}_W(c)) = \hat{\alpha}_W(c) - \alpha = \left(\frac{1}{W^2} X \right)^+ \frac{1}{W^2} e + \left[E_p - \left(\frac{1}{W^2} X \right)^+ \left(\frac{1}{W^2} X \right) \right] (c - \alpha), \quad c \in \mathbb{R}^p.$$

Найменшу норму серед усіх оцінок буде мати оцінка

$$\hat{\alpha}_W = \left(\frac{1}{W^2} X \right)^+ \frac{1}{W^2} y,$$

причому формула останньої оцінки суттєво спрощується, якщо матриця X має повний ранг по рядках ($\text{rank}(X) = N$), і набуває вигляду

$$\hat{\alpha}_W = X^T (X X^T)^{-1} y.$$

Дискусія і висновки

Для оцінки параметрів регресійної моделі запропоновано використовувати оцінку методу найменших квадратів відносних похибок $\hat{\alpha}$. Вираз її явної форми представлення, коли справджується класичне припущення, що гарантує її єдиність, узагальнено на випадок, коли порушується згадане класичне припущення. В останньому випадку дослідження проведено з використанням оператора псевдообернення матриць за Муром – Пенроузом. Для обох випадків наведені явні форми представлення як для оцінки методу найменших квадратів відносних похибок, так і для її похибки оцінювання. На відміну від оцінки звичайного методу найменших квадратів, що мінімізує суму квадратів абсолютних похибок моделі, оцінка методу найменших квадратів відносних похибок мінімізує суму квадратів відносних похибок моделі. З огляду на це звичайний метод найменших квадратів можна називати методом найменших квадратів абсолютних похибок.

Як наслідок для обох згаданих випадків також представлено аналогічні результати для оцінки зваженого методу найменших квадратів, коли в ролі вагової матриці взято довільну позитивно визначену матрицю.

Усі представлені твердження повністю узгоджуються з раніше отриманими відомими результатами.

Джерела фінансування. Дослідження частково профінансовано Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Чарін, В. С. (2004). *Лінійна алгебра*. Техніка.
- Albert, A. (1972). *Regression and the Moore – Penrose Pseudoinverse*. Academic Press.
- Björck, Å. (1996). *Numerical Methods for Least Squares Problems*. SIAM.
- Chengsi, L. (2001). Least Square Method Based on Relative Error. *Science Direct Working Paper, No S1574-0358(04)70710-8*, 371–380. <https://ssrn.com/abstract=3153628>
- Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Frid. Perthes et I. H. Besser.
- Hansen, P. C., Pereyra, V., & Scherer, G. (2013). *Least Squares Data Fitting with Applications*. Johns Hopkins University Press. <https://doi.org/10.1353/book.21076>
- Legendre, A. M. (1805). *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Firmin Didot.
- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin American Mathematical Society*, 26(9), 394–395.
- Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51(3), 406–413.
- Rao, C. R., Toutenburg, H., Shalabh, & Heumann, C. (2008). *Linear Models and Generalizations: Least Squares and Alternatives* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-540-74227-2>
- Strutz, T. (2016). *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond* (2nd ed.). Springer Vieweg.
- Tofallis, C. (2008). Least Squares Percentage Regression. *Journal of Modern Applied Statistical Methods*, 7(2), 526–534. <https://doi.org/10.22237/jmasm/1225513020>
- Weisberg, S. (2014). *Applied Linear Regression* (4th ed.). Wiley.

References

- Albert, A. (1972). *Regression and the Moore – Penrose Pseudoinverse*. Academic Press.
- Björck, Å. (1996). *Numerical Methods for Least Squares Problems*. SIAM.
- Charin, V. S. (2004). *Linear Algebra*. Technics [in Ukrainian].
- Chengsi, L. (2001). Least Square Method Based on Relative Error. *Science Direct Working Paper, No S1574-0358(04)70710-8*, 371–380. <https://ssrn.com/abstract=3153628>
- Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Frid. Perthes et I. H. Besser.
- Hansen, P. C., Pereyra, V., & Scherer, G. (2013). *Least Squares Data Fitting with Applications*. Johns Hopkins University Press. <https://doi.org/10.1353/book.21076>
- Legendre, A. M. (1805). *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Firmin Didot.

- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin American Mathematical Society*, 26(9), 394–395.
- Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51(3), 406–413.
- Rao, C. R., Toutenburg, H., Shalabh, & Heumann, C. (2008). *Linear Models and Generalizations: Least Squares and Alternatives* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-540-74227-2>
- Strutz, T. (2016). *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond* (2nd ed.). Springer Vieweg.
- Tofallis, C. (2008). Least Squares Percentage Regression. *Journal of Modern Applied Statistical Methods*, 7(2), 526–534. <https://doi.org/10.22237/jmasm/1225513020>
- Weisberg, S. (2014). *Applied Linear Regression* (4th ed.). Wiley.

Отримано редакцією журналу / Received: 31.03.25

Прорецензовано / Revised: 12.08.25

Схвалено до друку / Accepted: 10.10.25

Alexander SLABOSPITSKY, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0009-0007-8753-2075

e-mail: alexsl@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

LEAST SQUARES METHOD OF RELATIVE ERRORS FOR PARAMETER ESTIMATION OF SYSTEMS UNDER NON-CLASSICAL ASSUMPTIONS

The problem of optimal estimation of regression model parameters based on available observations of the dependent variable and all regressors is considered. It is suggested to use the least squares method of relative errors to solve this problem. The estimate of this method is the point at which the sum of squares of relative errors of the regression model reaches its minimum, not the sum of squares of absolute errors of the model, as in the ordinary least squares method, which can be called the least squares method of absolute errors. The value of the quality functional of the considered method no longer depends on the units of measurement of the available observations.

In the paper, the explicit form of the representation of the optimal estimate of the least squares method of relative errors in a situation where the classical assumption is valid, which guarantees its uniqueness, is extended to the case when this assumption is violated and the estimate is no longer unique. In the latter situation, there is a need to use the Moore-Penrose matrix pseudo-inversion operator. Corresponding explicit expressions for the estimation error by this method are also presented for both cases.

In addition, as a consequence, for the weighted least squares method, in the case of using an arbitrary positive definite matrix as the weight matrix, for both of the above-mentioned situations, explicit forms of representation of the corresponding optimal estimate of this method and the expression for the estimation error are also given by analogy.

Keywords: regression model, parameter estimation, non-classical assumption, Moore – Penrose pseudo-inversion operator, least squares method, weighted least squares method, least squares method of relative errors.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

УДК 517.95:419.86:539.3

DOI: <https://doi.org/10.17721/1812-5409.2025/2.32>

Володимир СТОЯН¹, д-р фіз.-мат. наук, проф.
ORCID ID: 0009-0008-0746-5050
e-mail: v_a_stoyan@ukr.net

Дмитро ЧЕРНИЙ¹, д-р техн. наук, доц.
ORCID ID: 0000-0002-6378-8048
e-mail: d_cherniy@ukr.net

Сергій ВОЛОЩУК¹, канд. фіз.-мат. наук, доц.
ORCID ID: 0009-0000-6416-5566
e-mail: sr.voloshchuk@gmail.com

Роман ЗАТОРСЬКИЙ², д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0001-7503-4932
e-mail: romazatorsky@gmail.com

¹Київський національний університет імені Тараса Шевченка, Київ, Україна
²Карпатський національний університет імені Василя Стефаника, Івано-Франківськ, Україна

ТРИВИМІРНІ ЗАДАЧІ КЕРУВАННЯ ДИНАМІКОЮ НЕПОВНО СПОСТЕРЕЖУВАНИХ ТОВСТИХ ПРУЖНИХ ПЛИТ. ЧАСТИНА II. ВИПАДОК ДИСКРЕТНО ЗАДАНОГО БАЖАНОГО СТАНУ

Уперше розв'язано складні задачі керування товстою пружною плитою, динаміку якої описано рівняннями Ляме в обмеженій тривимірній просторовій області на скінченному часовому відрізку. Суть цих задач полягає в тому, що поле пружно-динамічних зміщень плити під дією керуючої функції має набувати бажаного стану, заданого дискретно в просторі та часі, або перебувати в мінімально можливому, за відомих умов, середньоквадратичному околі цих значень. У ролі керуючих функцій можуть бути обрані об'ємні, початкові, поверхневі та торцеві функції зовнішньодинамічних впливів, що можуть впливати на динаміку плити кожна окремо, або в практично можливій комбінації кількох таких впливів. Зазначені задачі розв'язано за наявності дискретно визначених зовнішньодинамічних спостережень за початковим, крайовим і торцевим станом плити у наперед відомих точках, що належать областям визначення згаданих вище станів. Фактично, ці спостереження є дискретними початковими, крайовими та торцевими умовами, якими доповнюється рівняння Ляме. Проте їхня кількість і якість може бути довільною та впливатиме лише на громіздкість обчислень. На основі дискретних початкових, крайових і торцевих умов разом із дискретними бажаними станами, ураховуючи при цьому інтегральний вигляд моделі, функції керування знайдено як псевдорозв'язки системи алгебраїчних рівнянь або системи інтегральних рівнянь. Також побудовано й оцінено на точність й однозначність аналітичні залежності зовнішньодинамічних керуючих факторів та викликаного ними поля пружно-динамічних зміщень точок плити. Розглянуто частинні випадки деяких можливих комбінацій керувань динамікою товстої пружної плити.

Ключові слова: просторово розподілені динамічні системи, просторові задачі теорії пружності, товсті пружні плити, задачі керування.

Класифікація відповідно до AMS 2020: 74-10, 74H75.

Вступ

Пропонована стаття продовжує дослідження авторів (Стоян, Черний, & Волощук, 2024) проблем керування динамікою пружно-динамічної конструкції, що є об'єктом дослідження і являє собою визначене у просторі тривимірне пружне тіло, обмежене двома паралельними площинами та довільного обрису циліндричною поверхнею між ними.

Дослідження пружно-динамічних характеристик таких тіл (плит) у тривимірній постановці є складним математичним завданням, що не має точного та прийнятного для практики розв'язку. Ще складнішими є завдання керування динамікою таких тіл. Останнє пов'язане зі складнощами математичних залежностей (Lur'e, 1964; Timoshenko, & Woinowsky-Krieger, 1987) поля пружно-динамічних зсувів тіла від зовнішньодинамічних збудовальних факторів, якими вони визначаються.

Більшість традиційно відомих результатів дослідження динаміки названих об'єктів були отримані чисельно (Grigorenko, Savula, & Mukha, 2000; Grigorenko, Ya., & Grigorenko, A., 2013) на підставі математичних моделей (Nemysh, & Khoma, 1993; Григорюк, & Селезов, 1973; Нигул, 1963; Шереметьев, & Пелех, 1964) їхньої динаміки, що зазвичай будувалися (Selezov, 2018) на припущеннях (Селезов, 1961) про невелику товщину пружного тіла щодо його інших геометричних розмірів.

Згодом нами було запропоновано (Стоян, 2016) напівтривимірні математичні моделі динаміки товстих пружних плит, якими тривимірне поле динамічних зміщень плити описувалося двовимірними диференціальними рівняннями, параметрично залежними від виродженої поперечної координати. Незважаючи на нескінченновимірний порядок цих рівнянь, нами була запропонована і методика побудови їхнього точного математичного розв'язку з подальшим використанням до розв'язання тривимірних початково-крайових задач динаміки пружних плит і керування ними (Стоян, 2016).

У роботі (Stoyan, 2018) побудовано більш вдалу математична модель динаміки товстих пружних плит, яка в зручнішій для практики формі описує тривимірне пружно-динамічне поле товстої пружної плити. Останнє допомогло побудувати аналітичний розв'язок початково-крайової задачі динаміки плити, який, точно задовольняючи диференційне рівняння моделі, за середньоквадратичним критерієм узгоджується із зовнішньодинамічними спостереженнями за плитою на її гранично-поверхневих областях, незважаючи на кількість і якість останніх. На базі цієї моделі у

© Стоян Володимир, Черний Дмитро, Волощук Сергій, Заторський Роман, 2025

дослідженні (Стоян, Черній, & Волощук, 2024) були успішно розв'язані і задачі керування динамікою товстої пружної плити з виведення поля її пружно-динамічних зміщень у середньоквадратичний окіл аналітично заданого стану гранично-поверхневих областей останньої.

Методи, подібні до розглянутих у (Стоян, 2024), але використані для моделювання функціонально стійких технологічних процесів промислових підприємств були застосовані в роботах (Pichkur, & Sobchuk, 2021; Pichkur et al., 2024), де автори розглянули лінійну дискретну модель деякого абстрактного процесу та задачу керування ним, використовуючи при цьому властивості псевдообернених матриць.

Дослідження, проведені у представлений роботі є актуальними, оскільки алгоритм побудови на їхній основі розв'язку задачі керування динамікою товстої пружної плити суттєво спрощується, не залежить від кількості та якості умов про початковий, крайовий, торцевий і бажаний стани плити і може бути відносно нескладно реалізований за допомогою сучасних систем комп'ютерної математики.

Нижче, як продовження результатів роботи (Стоян, Черній, & Волощук, 2024), з використанням викладених у (Стоян, 2024) методик дослідження розподілених просторово-часових систем в умовах невизначеності, нами будуть розв'язані задачі керування полем пружно-динамічних зміщень точок товстої пружної плити для непідйомних класичною математикою випадків, коли бажаний стан плити задається в окремих просторово-граничних точках останньої. За припущення, що дискретно визначеною є й інформація про гранично-поверхневий стан плити, будуть побудовані аналітичні залежності початково, гранично та поверхнево визначених керуючих факторів, кожен з яких окремо та спільно з іншими дає змогу поле пружно-динамічних зміщень точок плити за середньоквадратичним критерієм привести до бажаного стану. Як і в (Стоян, Черній, & Волощук, 2024) будуть досліджені питання точності й однозначності отриманих розв'язків, а також їхня форма у вироджених випадках – необмежена в плані плита (пружний шар) і динаміка плити, що встановилася в часі.

1. Основні означення та результати

1.1. Тривимірна математична модель динаміки товстої пружної плити

Продовжимо розгляд динаміки товстої пружної плити, внутрішній об'єм S_0 якої у декартовій системі координат ox, oy, oz обмежений поверхнями $z = \pm h$ і циліндром $\Gamma_0 = \{\xi = (x, y, z): (x, y) \in \Gamma, -h \leq z \leq h\}$ між ними.

Позначивши через λ і μ пружно-динамічні характеристики матеріалу плити, динаміку (t – часова координата) її опишемо системою рівнянь Ляме (Lur'e, 1964)

$$L(\partial_\xi, \partial_t)u(s) = f(s), \quad (1)$$

у якій $s = (\xi, t) \in S_0^T = S_0 \times \Gamma_0 \times [0, T]$, $\partial_\xi = (\partial_x, \partial_y, \partial_z)$,

$$L(\partial_\xi, \partial_t) = \begin{pmatrix} (\lambda + 2\mu)\partial_x^2 + \mu(\partial_y^2 + \partial_z^2) - \rho\partial_t^2 & (\lambda + \mu)\partial_x\partial_y & (\lambda + \mu)\partial_x\partial_z \\ (\lambda + \mu)\partial_y\partial_x & (\lambda + 2\mu)\partial_y^2 + \mu(\partial_x^2 + \partial_z^2) - \rho\partial_t^2 & (\lambda + \mu)\partial_y\partial_z \\ (\lambda + \mu)\partial_z\partial_x & (\lambda + \mu)\partial_z\partial_y & (\lambda + 2\mu)\partial_z^2 + \mu(\partial_x^2 + \partial_y^2) - \rho\partial_t^2 \end{pmatrix},$$

а $f(s) = \text{col}(f_1(s), f_2(s), f_3(s))$ і $u(s) = \text{col}(u_1(s), u_2(s), u_3(s))$ – вектори об'ємно визначених силових факторів і спричинених ними зміщень точок $\xi \in S_0 \times \Gamma_0$ у напрямку координатних осей.

Як і в (Стоян, Черній, & Волощук, 2024), диференціальну математичну модель (1) замінімо (Стоян, 2016; Stoyan, 2018) зручнішим для роботи її інтегральним еквівалентом виду

$$u(s) = \int_{-\infty}^{+\infty} G(s-s')f(s')ds',$$

$$G(s-s') = \frac{1}{16\pi^4} \int_{-i\infty}^{+i\infty} \int_{-i\infty}^{+i\infty} \frac{E(p, q, s-s')}{L(p, q)} dp dq, \quad (2)$$

де i – уявна одиниця, $p = (p_1, p_2, p_3)$, $dp = dp_1 dp_2 dp_3$, а

$$L(\partial_\xi, \partial_t)G(s-s') = \Delta(s-s') \quad (3)$$

за

$$\Delta(s-s') = \text{diag}(\delta(s-s'), i = \overline{1, 3}),$$

$$\delta(s-s') = \delta(x-x')\delta(y-y')\delta(z-z')$$

(тут $\delta(\cdot)$ – δ -функція Дірака).

Зважаючи на обмеженість просторово-часової області S_0^T , зовнішньодинамічну картину функціонування плити для найзагальнішого випадку визначимо співвідношеннями:

$$L_r^0(\partial_t)u(s)|_{t=0, \xi \in S_0} = f_r^0(\xi), r = \overline{1, R_0},$$

$$L_\rho^\pm(\partial_\xi)u(s)|_{s \in S_0^\pm} = f_\rho^\pm(s), \rho = \overline{1, R^\pm},$$

$$L_\rho^\Gamma(\partial_\xi)u(s)|_{s \in S_0^\Gamma} = f_\rho^\Gamma(s), \rho = \overline{1, R_\Gamma}, \quad (4)$$

у яких $L_r^0(\partial_t)$ ($r = \overline{1, R_0}$), $L_\rho^\pm(\partial_\xi)$ ($\rho = \overline{1, R^\pm}$), $L_\rho^\Gamma(\partial_\xi)$ ($\rho = \overline{1, R_\Gamma}$) – задані диференціальні оператори, $S_0^\pm = \{(x, y) \in S_0, z = \pm h\} \times [0, T]$, $S_0^\Gamma = \{\Gamma(\xi) \times (-h \leq z \leq h)\} \times [0, T]$, а $f_r^0(\xi)$ ($r = \overline{1, R_0}$), $f_\rho^\pm(s)$ ($\rho = \overline{1, R^\pm}$), $f_\rho^\Gamma(s)$ ($\rho = \overline{1, R_\Gamma}$) – відомі силові фактори.

Ураховуючи властивості визначеної згідно із (2), (3) матричної функції $G(s-s')$, розв'язок задач (1) і (4) можна записати (Stoyan, 2018) у вигляді

$$u(s) = u_\infty(s) + u^0(s) + u^+(s) + u^-(s) + u^\Gamma(s), \quad (5)$$

де

$$\begin{aligned} u_{\infty}(s) &= \int_{S_0^T} G(s-s')f(s')ds', \\ u^0(s) &= \int_{S_0} G(s-s')|_{t'=0}f^0(\xi')d\xi', \\ u^{\pm}(s) &= \int_{S_0^{\pm}} G(s-s')f^{\pm}(s')ds', \\ u^{\Gamma}(s) &= \int_{S_0^{\Gamma}} G(s-s')f^{\Gamma}(s')ds' \end{aligned} \quad (6)$$

за

$$\begin{aligned} f^0(\xi) &= \text{col}(f_r^0(\xi), r = \overline{1, R_0}), \\ f^{\pm}(s) &= \text{col}(f_{\rho}^{\pm}(s), \rho = \overline{1, R_{\pm}}), \\ f^{\Gamma}(s) &= \text{col}(f_{\rho}^{\Gamma}(s), \rho = \overline{1, R_{\Gamma}}). \end{aligned}$$

Зауважимо, що представлення (5) вектор-функції $u(s)$ зміщеної точок розглядуваної плити спостерігатимемо і для випадку, коли зовнішньодинамічні об'єкти та гранично-поверхневі силові фактори визначені дискретно так, що при заданих у точках s_l ($l = \overline{1, L}$) значеннями f_l ($l = \overline{1, L}$) об'ємно розподілених зовнішньодинамічних силових факторів $f(s)$

$$\begin{aligned} L_r^0(\partial_t)u(s)|_{t=0, \xi=\xi_l^0} &= f_{rl}^0, r = \overline{1, R_0}, l = \overline{1, L_0}, \\ L_{\rho}^{\pm}(\partial_{\xi})u(s)|_{s=s_l^{\pm} \in S_0^{\pm}} &= f_{\rho l}^{\pm}, \rho = \overline{1, R_{\pm}}, l = \overline{1, L_{\pm}}, \\ L_{\rho}^{\Gamma}(\partial_{\xi})u(s)|_{s=s_l^{\Gamma} \in S_0^{\Gamma}} &= f_{\rho l}^{\Gamma}, \rho = \overline{1, R_{\Gamma}}, l = \overline{1, L_{\Gamma}}. \end{aligned} \quad (7)$$

Проте

$$\begin{aligned} u_{\infty}(s) &= \sum_{l=1}^L G(s-s_l)f_l, \\ u^0(s) &= \sum_{r=1}^{R_0} \sum_{l=1}^{L_0} G(s-s_l^0)|_{t_l^0=0, \xi=\xi_l^0} f_{rl}^0, \\ u^{\pm}(s) &= \sum_{\rho=1}^{R_{\pm}} \sum_{l=1}^{L_{\pm}} G(s-s_l^{\pm})f_{\rho l}^{\pm}, \\ u^{\Gamma}(s) &= \sum_{\rho=1}^{R_{\Gamma}} \sum_{l=1}^{L_{\Gamma}} G(s-s_l^{\Gamma})f_{\rho l}^{\Gamma}, \end{aligned} \quad (8)$$

що, як і вище, розв'язує початково-крайові задачі (1), (4) та (1), (7).

Для випадку, коли функції $f^0(\xi)$, $f^{\pm}(s)$, $f^{\Gamma}(s)$ та вектори

$$\begin{aligned} f^0 &= \text{col}((f_{rl}^0), r = \overline{1, R_0}, l = \overline{1, L_0}), \\ f^{\pm} &= \text{col}((f_{\rho l}^{\pm}), \rho = \overline{1, R_{\pm}}, l = \overline{1, L_{\pm}}), \\ f^{\Gamma} &= \text{col}((f_{\rho l}^{\Gamma}), \rho = \overline{1, R_{\Gamma}}, l = \overline{1, L_{\Gamma}}) \end{aligned}$$

їхніх значень мають не силовий характер, вплив останніх на стан внутрішньо-поверхневих точок плити можна змодельовувати (Stoyan, 2018) силовими характеристиками $\bar{f}^0(s^0)$, $\bar{f}^{\pm}(s^{\pm})$, $\bar{f}^{\Gamma}(s^{\Gamma})$,

$$\begin{aligned} \bar{f}^0 &= \text{col}(\bar{f}_m^0, m = \overline{1, M_0}), \\ \bar{f}^{\pm} &= \text{col}(\bar{f}_m^{\pm}, m = \overline{1, M_{\pm}}), \\ \bar{f}^{\Gamma} &= \text{col}(\bar{f}_m^{\Gamma}, m = \overline{1, M_{\Gamma}}), \end{aligned}$$

що справедливі для $\bar{f}_m^0 = \bar{f}_m^0(s_m^0)$, $\bar{f}_m^{\pm} = \bar{f}_m^{\pm}(s_m^{\pm})$, $\bar{f}_m^{\Gamma} = \bar{f}_m^{\Gamma}(s_m^{\Gamma})$,

$$\begin{aligned} s^0 &\in \bar{S}, s_m^0 \in \bar{S}, \bar{S} = S_0 \times (-\infty, 0], \\ s^+ &\in \bar{S}^+, s_m^+ \in \bar{S}^+, \bar{S}^+ = (R^3 \setminus S_0) \times (z > h) \times [0, T], \\ s^- &\in \bar{S}^-, s_m^- \in \bar{S}^-, \bar{S}^- = (R^3 \setminus S_0) \times (z < -h) \times [0, T], \\ s^{\Gamma} &\in \bar{S}^{\Gamma}, s_m^{\Gamma} \in \bar{S}^{\Gamma}, \bar{S}^{\Gamma} = (R^3 \setminus S_0) \times (-h < z < h) \times [0, T] \end{aligned}$$

та визначають за аналогією із (6), (8) функції

$$\begin{aligned} u^0(s) &= \int_{\bar{S}} G(s-s')\bar{f}^0(s')ds', \\ u^{\pm}(s) &= \int_{\bar{S}^{\pm}} G(s-s')\bar{f}^{\pm}(s')ds', \\ u^{\Gamma}(s) &= \int_{\bar{S}^{\Gamma}} G(s-s')\bar{f}^{\Gamma}(s')ds', \end{aligned} \quad (9)$$

або

$$\begin{aligned} u^0(s) &= \sum_{m=1}^{M_0} G(s - s_m^0) \bar{f}_m^0, s_m^0 \in \bar{S}, \\ u^\pm(s) &= \sum_{m=1}^{M^\pm} G(s - s_m^\pm) \bar{f}_m^\pm, s_m^\pm \in \bar{S}^\pm, \\ u^\Gamma(s) &= \sum_{m=1}^{M_\Gamma} G(s - s_m^\Gamma) \bar{f}_m^\Gamma, s_m^\Gamma \in \bar{S}^\Gamma \end{aligned} \quad (10)$$

у співвідношенні (5) при

$$\begin{aligned} \Phi &= \sum_{r=1}^{R_0} \int_{\bar{S}} \|L_r^0(\partial_t)u(s)|_{t=0} - f_r^0(\xi)\|^2 d\xi + \sum_{\rho=1}^{R_+} \int_{\bar{S}^+} \|L_\rho^+(\partial_\xi)u(s) - f_\rho^+(s)\|^2 ds + \\ &+ \sum_{\rho=1}^{R_-} \int_{\bar{S}^-} \|L_\rho^-(\partial_\xi)u(s) - f_\rho^-(s)\|^2 ds + \sum_{\rho=1}^{R_\Gamma} \int_{\bar{S}^\Gamma} \|L_\rho^\Gamma(\partial_\xi)u(s) - f_\rho^\Gamma(s)\|^2 ds \rightarrow \min_{u(s)}, \end{aligned} \quad (11)$$

або

$$\Phi = \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 \rightarrow \min_{u(s)}, \quad (12)$$

$$\begin{aligned} \Phi_1 &= \sum_{r=1}^{R_0} \sum_{l=1}^{L_0} \|L_r^0(\partial_t)u(s)|_{t=0, \xi=\xi_l^0} - f_{rl}^0\|^2, \\ \Phi_2 &= \sum_{\rho=1}^{R_+} \sum_{l=1}^{L_+} \|L_\rho^+(\partial_\xi)u(s)|_{s=s_l^+} - f_{\rho l}^+\|^2, \\ \Phi_3 &= \sum_{\rho=1}^{R_-} \sum_{l=1}^{L_-} \|L_\rho^-(\partial_\xi)u(s)|_{s=s_l^-} - f_{\rho l}^-\|^2, \\ \Phi_4 &= \sum_{\rho=1}^{R_\Gamma} \sum_{l=1}^{L_\Gamma} \|L_\rho^\Gamma(\partial_\xi)u(s)|_{s=s_l^\Gamma} - f_{\rho l}^\Gamma\|^2, \end{aligned}$$

а вектор-функція $u(s)$ будується з урахуванням (9), (10).

Не зупиняючись на розв'язанні задач (1), (4), (11) та (1), (7), (12), які за відомих об'ємно розподілених зовнішньодинамічних збурень $f(s)$ детально розглянуті нами у (Стоян, Черній, & Волощук, 2024), зауважимо, що представлена співвідношеннями (5), (9) та (5), (10) вектор-функція $u(s)$ рівнянню (1) буде задовольняти точно, а співвідношенням (4), (7) середньоквадратично при визначених згідно з (11), (12) моделювальному вектору $\bar{f} = \text{col}(\bar{f}^0, \bar{f}^+, \bar{f}^-, \bar{f}^\Gamma)$ та моделювальній вектор-функції $\bar{f}(s) = \text{col}(\bar{f}^0(\xi), \bar{f}^+(s), \bar{f}^-(s), \bar{f}^\Gamma(s))$.

1.2. Задачі керування динамікою товстої пружної плити об'ємно розподіленими зовнішньодинамічними впливами

Викладені результати дослідження динаміки товстої пружної плити зводилися до розв'язання задач побудови вектор-функції зміщень точок плити за відомих початкових об'ємно-динамічних поверхнево і гранично розподілених зовнішньодинамічних збурень, якими були як силові фактори $f(s)$, $f^0(\xi)$, $f^\pm(s)$, $f^\Gamma(s)$, так і збурення не силового характеру, співвідношеннями (4), (7) визначені через компоненти вектор-функції $u(s)$.

Розглянемо задачі побудови одного з названих факторів за відомих інших, які визначимо з урахуванням (4) або (7) й умов бажаних станів

$$L_r(\partial_s)u(s)|_{s=s^*} = U_r(s^*), r = \overline{1, R}, \quad (13)$$

$$L_r(\partial_s)u(s)|_{s=s_l^*} = U_{rl}, r = \overline{1, R}, l = \overline{1, L} \quad (14)$$

при заданих $L_r(\partial_s)$, $U_r(s)$, U_{rl} та $s^* \in S_0^T$, $s_l^* \in S_0^T$, $l = \overline{1, L}$.

Зауважимо, що відсутність обмежень на кількість і якість інформації, заданої співвідношеннями (4), (7), а також довільність у визначенні (13), (14) бажаного стану плити, поставлені задачі керування роблять математично некоректними. З огляду на це невідомі керуючі фактори визначатимемо з умов найкращого середньоквадратичного узгодження вектор-функції $u(s)$ зміщень їм відповідних зі значеннями бажаними та спостережуваними згідно із (4), (7). Зауважимо, що для випадку, коли останні задаються неперервно співвідношеннями (4), то постановки та розв'язки задач керування наведені у (Стоян, Черній, & Волощук, 2024).

Нижче зупинимося на розв'язанні цих задач для дискретно визначеного бажаного стану (14) за умови, що зовнішньодинамічні та торцеві спостереження за плитою визначаються співвідношенням (7). Як і в (Стоян, Черній, & Волощук, 2024) ці спостереження моделюватимемо дискретно та неперервно визначеними моделювальними факторами $(\bar{f}^0, \bar{f}^\pm, \bar{f}^\Gamma)$ та $(\bar{f}^0(s), \bar{f}^\pm(s), \bar{f}^\Gamma(s))$ відповідно, які поза розглядуваною областю S_0^T визначені так, щоб за середньоквадратичного виконання умов (12)

$$\Phi_* = \sum_{r=1}^R \sum_{l=1}^L \|L_r(\partial_s)u(s)|_{s=s_l^*} - U_{rl}\|^2 \rightarrow \min_{u(s)}. \quad (15)$$

Розв'язувальні рівняння для визначення цих моделювальних векторів і вектор-функцій отримаємо із (7) після підстановки туди (5), (10) та (5), (9) відповідно. Це будуть системи лінійних алгебраїчних і лінійних інтегральних рівнянь, які при керуванні пружною плитою за допомогою об'ємно розподілених зусиль $f(s)$ та $\bar{f} = \text{col}(f(s_m) = f_m, m = \overline{1, M})$, матимуть вигляд

$$A\bar{f} = F \quad (16)$$

та

$$\int_{(\cdot)} A(s)\bar{f}(s)ds = F. \quad (17)$$

Тут $(\cdot)$ означає інтегрування по області визначення підінтегральної функції, а

$$F = \text{col}\left(\left(f_{rl}^0, r = \overline{1, R_0}, l = \overline{1, L_0}\right), \left(f_{pl}^+, \rho = \overline{1, R_+}, l = \overline{1, L_+}\right), \left(f_{pl}^-, \rho = \overline{1, R_-}, l = \overline{1, L_-}\right), \right. \\ \left. \left(f_{\rho l}^\Gamma, \rho = \overline{1, R_\Gamma}, l = \overline{1, L_\Gamma}\right), \left(U_{rl}, r = \overline{1, R}, l = \overline{1, L}\right)\right)$$

– вектор бажаного стану та зовнішньодинамічних спостережень за станом плити,

$$\bar{f} = \text{col}\left(\bar{f}^0, \bar{f}^+, \bar{f}^-, \bar{f}^\Gamma, \bar{f}\right),$$

$$\bar{f}(s) = \text{col}\left(\bar{f}^0(s), \bar{f}^+(s), \bar{f}^-(s), \bar{f}^\Gamma(s), \bar{f}(s)\right),$$

– вектор та вектор-функція керуючих та моделювальних згідно із (12) і (15) пружно-динамічних факторів, а

$$A = [A_{ij}]_{i,j=1}^{i,j=5}, A(s) = [A_{ij}(s)]_{i,j=1}^{i,j=5}$$

– матриця та матрична функція, у яких

$$A_{ij} = \text{col}\left(\left(\text{str}\left(\bar{L}(i)G(s - \underline{s}(j))\right)\right)_{s=\bar{s}(i)}, m = \overline{1, M(j)}, r(i) = \overline{1, \bar{R}(i)}, l = \overline{1, \underline{L}(i)}\right),$$

$$A_{ij}(s') = \text{col}\left(\left(\bar{L}(i)G(s - s')\right)_{s=\bar{s}(i)}, r(i) = \overline{1, \bar{R}(i)}, l = \overline{1, \underline{L}(i)}\right), s' \in \Sigma(j).$$

Тут індекси $i = \overline{1, 5}$ та $j = \overline{1, 5}$ визначають номери компонентів векторів

$$r(i) = (r, \rho, \rho, \rho, r),$$

$$\bar{L}(i) = (L_r^0(\partial_t), L_\rho^+(\partial_s), L_\rho^-(\partial_s), L_\rho^\Gamma(\partial_s), L_r(\partial_s)),$$

$$\bar{R}(i) = (R_0, R_+, R_-, R_\Gamma, R),$$

$$\underline{L}(i) = (L_0, L_+, L_-, L_\Gamma, L),$$

$$\bar{s}(i) = ((\xi_l^0, 0), s_l^+, s_l^-, s_l^\Gamma, s_l^*),$$

$$\underline{s}(j) = (s_m^0, s_m^+, s_m^-, s_m^\Gamma, s_m),$$

$$M(j) = (M_0, M_+, M_-, M_\Gamma, M),$$

$$\Sigma(j) = (\bar{s}, \bar{s}^+, \bar{s}^-, \bar{s}^\Gamma, \bar{s}_0^T).$$

Розв'язками систем (16), (17) такими, що

$$\|A\bar{f} - F\|^2 \rightarrow \min_{\bar{f}}, \quad (18)$$

$$\left\| \int_{(\cdot)} A(s)\bar{f}(s)ds - F \right\|^2 \rightarrow \min_{\bar{f}(s)}, \quad (19)$$

будуть (Stoyan, 2018)

$$\bar{f} = A^T P_1^* F + \bar{v} - A^T P_1^* A \bar{v}, \quad (20)$$

$$\bar{f}(s) = A^T(s) P_2^* F + \bar{v}(s) - A^T(s) P_2^* A_v, \quad (21)$$

за відповідних розмірностей вектора $\bar{v} = \text{col}(v_0, v^+, v^-, v^\Gamma, v)$ та вектор-функції

$$\bar{v}(s) = \text{col}((v_0(s), s \in \bar{s}), (v^+(s), s \in \bar{s}^+), (v^-(s), s \in \bar{s}^-), (v^\Gamma(s), s \in \bar{s}^\Gamma), (v(s), s \in \bar{s}_0^T)),$$

тотожно рівних нулю, якщо $\det A^T A > 0$ та $\lim_{N \rightarrow \infty} \det[A^T(s_i)A(s_j)]_{i,j=1}^{i,j=N} > 0$ відповідно,

вектора

$$A_v = \int_{(\cdot)} A(s)\bar{v}(s)ds$$

та матриць

$$P_1 = AA^T, \quad P_2 = \int_{(\cdot)} A(s)A^T(s)ds,$$

псевдообернення яких позначено знаком $*$.

Зі співвідношень (20), (21) знаходимо керуючу вектор-функцію

$$\bar{f}(s) = (A_{i5}^T(s), i = \overline{1, 5}) P_2^* (F - A_v) + v(s),$$

визначений вище вектор її значень

$$\bar{f} = (A_{i5}^T, i = \overline{1, 5}) P_1^* (F - A_v) + \bar{v},$$

а також моделювальні вектор-функції

$$\begin{aligned}\bar{f}^0(s) &= (A_{i1}^T(s), i = \overline{1, 5}) P_2^*(F - A_v) + v_0(s), \\ \bar{f}^+(s) &= (A_{i2}^T(s), i = \overline{1, 5}) P_2^*(F - A_v) + v^+(s), \\ \bar{f}^-(s) &= (A_{i3}^T(s), i = \overline{1, 5}) P_2^*(F - A_v) + v^-(s), \\ \bar{f}^\Gamma(s) &= (A_{i4}^T(s), i = \overline{1, 5}) P_2^*(F - A_v) + v^\Gamma(s),\end{aligned}\quad (22)$$

з їхніми значеннями

$$\begin{aligned}\bar{f}^0 &= (A_{i1}^T, i = \overline{1, 5}) P_1^*(F - A\bar{v}) + v_0, \\ \bar{f}^+ &= (A_{i2}^T, i = \overline{1, 5}) P_1^*(F - A\bar{v}) + v^+, \\ \bar{f}^- &= (A_{i3}^T, i = \overline{1, 5}) P_1^*(F - A\bar{v}) + v^-, \\ \bar{f}^\Gamma &= (A_{i4}^T, i = \overline{1, 5}) P_1^*(F - A\bar{v}) + v^\Gamma.\end{aligned}\quad (23)$$

Визначенні у (22), (23) моделювальні функції $\bar{f}^0(s)$, $\bar{f}^+(s)$, $\bar{f}^-(s)$, $\bar{f}^\Gamma(s)$ та вектори їхніх значень $\bar{f}^0$, $\bar{f}^+$, $\bar{f}^-$, $\bar{f}^\Gamma$ дають змогу за необхідності через співвідношення (9), (10) знайти складові $u^0(s)$, $u^\pm(s)$, $u^\Gamma(s)$ вектор-функції стану $u(s)$.

Відповідно до (Selezov, 2018) ця функція, будучи розв'язком рівняння (1), із зовнішньодинамічними спостереженнями (7) та визначеному в (14) бажаному стані плити буде узгоджуватися з точністю

$$\varepsilon^2 = \min_{u(s)} (\Phi + \Phi_*) = \min_{\bar{f}} \|A\bar{f} - F\| = F^T F - F^T A A^* F \quad (24)$$

за дискретно визначеного керування $\bar{f}$ та

$$\varepsilon^2 = \min_{u(s)} (\Phi + \Phi_*) = \min_{\bar{f}(s)} \left\| \int_{(c)} A(s) \bar{f}(s) ds - F \right\|^2 = F^T F - F^T P_1 P_1^* F \quad (25)$$

за просторово розподіленого динамічного керування $\bar{f}(s)$.

Подамо керуючий вектор $\bar{f}$ та керуючу вектор-функцію $\bar{f}(s)$ у зручнішому для практичного використання вигляді. Зважаючи на те, що

$$\begin{aligned}A_{15} &= \text{col} \left(\left(\text{str} \left(L_r^0(\partial_t) G(s - s_m) \right) \Big|_{s=(\xi_l^0, 0)}, m = \overline{1, M} \right), r = \overline{1, R_0} \right), l = \overline{1, L_0}, \\ A_{25} &= \text{col} \left(\left(\text{str} \left(L_\rho^+(\partial_s) G(s - s_m) \right) \Big|_{s=s_l^+}, m = \overline{1, M} \right), \rho = \overline{1, R_+} \right), l = \overline{1, L_+}, \\ A_{35} &= \text{col} \left(\left(\text{str} \left(L_\rho^-(\partial_s) G(s - s_m) \right) \Big|_{s=s_l^-}, m = \overline{1, M} \right), \rho = \overline{1, R_-} \right), l = \overline{1, L_-}, \\ A_{45} &= \text{col} \left(\left(\text{str} \left(L_\rho^\Gamma(\partial_s) G(s - s_m) \right) \Big|_{s=s_l^\Gamma}, m = \overline{1, M} \right), \rho = \overline{1, R_\Gamma} \right), l = \overline{1, L_\Gamma}, \\ A_{55} &= \text{col} \left(\left(\text{str} \left(L_r(\partial_s) G(s - s_m) \right) \Big|_{s=s_l^*}, m = \overline{1, M} \right), r = \overline{1, R} \right), l = \overline{1, L},\end{aligned}$$

керуючий вектор $\bar{f}$ можна обчислити за формулою:

$$\bar{f} = (A_{15}^T \ A_{25}^T \ A_{35}^T \ A_{45}^T \ A_{55}^T) (A A^T)^* (F - A\bar{v}) + \bar{v},$$

де $\bar{v}$ – довільний вектор розмірності $M_0 + M_+ + M_- + M_\Gamma + M$ з дійсними компонентами. Зауважимо, що компоненти вектора $\bar{f}$ рівні останнім M компонентам знайденого за формулою (20) вектора керуючих і моделювальних пружно-динамічних факторів.

Ураховуючи те, що

$$\begin{aligned}A_{15}(s') &= \text{col} \left(\left(L_r^0(\partial_t) G(s - s') \Big|_{s=(\xi_l^0, 0)}, r = \overline{1, R_0} \right), l = \overline{1, L_0} \right), s' \in S_0^T, \\ A_{25}(s') &= \text{col} \left(\left(L_\rho^+(\partial_s) G(s - s') \Big|_{s=s_l^+}, \rho = \overline{1, R_+} \right), l = \overline{1, L_+} \right), s' \in S_0^T, \\ A_{35}(s') &= \text{col} \left(\left(L_\rho^-(\partial_s) G(s - s_m) \Big|_{s=s_l^-}, \rho = \overline{1, R_-} \right), l = \overline{1, L_-} \right), s' \in S_0^T, \\ A_{45}(s') &= \text{col} \left(\left(L_\rho^\Gamma(\partial_s) G(s - s_m) \Big|_{s=s_l^\Gamma}, \rho = \overline{1, R_\Gamma} \right), l = \overline{1, L_\Gamma} \right), s' \in S_0^T, \\ A_{55}(s') &= \text{col} \left(\left(L_r(\partial_s) G(s - s_m) \Big|_{s=s_l^*}, \rho = \overline{1, R} \right), l = \overline{1, L} \right), s' \in S_0^T,\end{aligned}$$

керуючу векторну функцію $\bar{f}(s)$ можна обчислити за формулою:

$$\bar{f}(s) = (A_{15}^T(s) \ A_{25}^T(s) \ A_{35}^T(s) \ A_{45}^T(s) \ A_{55}^T(s)) P_2^*(F - A_v) + v(s),$$

де $v(s)$ – довільна неперервна в S_0^T векторна функція, розмірність якої узгоджена з $\bar{f}(s)$. Зазначимо, що компоненти векторної функції $\bar{f}(s)$ рівні останнім компонентам, знайденої за формулою (21) векторної-функції керуючих та моделювальних пружно-динамічних факторів.

1.3. Задача керування динамікою товстої пружної плити за участю об'ємно розподілених зовнішньодинамічних впливів і гранично-поверхневих збурень

Розглянемо задачу керування динамікою товстої пружної плити для випадку, коли, як і вище, бажаний стан визначається критерієм (12), а у процесі керування, крім об'ємно розподілених керуючих факторів $f(s)$ та $\bar{f}$, беруть участь і гранично-поверхневі керуючі збурення

$$f_r^0(\xi), (\xi \in S_0, r = \overline{1, R_0}); f_\rho^+(s), (s \in S_0^+, \rho = \overline{1, R_+}); f_\rho^-(s), (s \in S_0^-, \rho = \overline{1, R_-}); f_\rho^\Gamma(s), (s \in S_0^\Gamma, \rho = \overline{1, R_\Gamma}).$$

Аналітичні вирази для останніх отримаємо зі співвідношень (4) за наявності вектор-функції $u(s)$, побудованої з урахуванням вектор-функцій $\bar{f}^0(\xi)$, $\bar{f}^\pm(s)$, $\bar{f}^\Gamma(s)$, або векторів $\bar{f}^0$, $\bar{f}^\pm$, $\bar{f}^\Gamma$ їхніх значень, згідно із (12), моделюючих зовнішньодинамічну картину спостережень (7) за нею. Задача визначення цих моделювальних факторів, а також і керуючих вектор-функцій $f(s)$ та вектора $\bar{f}$ зведеться до середньоквадратичного згідно із (18), (19) оберненням розв'язуючих рівнянь (16), (17), модифікованих для кожного конкретного випадку керування.

Керування початковими й об'ємно розподіленими динамічними збуреннями

Розглянемо задачі керування динамікою пружної плити сумісною дією об'ємно розподілених і початково визначених зовнішньодинамічних збурень $f(s)$ (або $\bar{f}$) та $f_0(s)$ за наявності поверхнево-граничних спостережень f^+ , f^- , f^Γ у (7).

Керуюча вектор-функція $f(s)$ і вектор $\bar{f}$ разом із вектор-функцією $\bar{f}(s)$, якою названі спостереження моделюються, будуть визначатися розв'язками (20), (21) рівнянь (18), (19), записаних без групи рівнянь з неспостережуваними f_{rl}^0 ($r = \overline{1, R_0}$, $l = \overline{1, L_0}$) праворуч. Керуючу ж вектор-функцію $f^0(\xi)$ отримаємо із (4) з урахуванням $u(s)$ визначеного відповідно до (5), (9).

Керування сумісною дією об'ємно розподілених зовнішньодинамічних збурень й одного із гранично-поверхневих впливів

Розглянемо випадок керування динамікою пружної плити за допомогою об'ємно розподіленого зовнішньодинамічного збурення $f(s)$ (або $\bar{f}$) й одного із гранично-поверхневих впливів $f^\pm(s)$, $f^\Gamma(s)$ за наявності початкових та гранично-поверхневих спостережень (7) за рештою зовнішньодинамічних збурювальних факторів.

Можна помітити, що розв'язання цих задач зведеться до середньоквадратичного обернення рівнянь (18), (19), у яких будуть відсутні рівняння з неспостережуваними $f_{\rho l}^\pm$ ($\rho = \overline{1, R_\pm}$, $l = \overline{1, L_\pm}$), $f_{\rho l}^\Gamma$ ($\rho = \overline{1, R_\Gamma}$, $l = \overline{1, L_\Gamma}$) у правій частині. Розв'язання наших рівнянь співвідношеннями (20), (21) визначить моделювальні функції (або вектори їхніх значень), якими згідно з (9), (10) знаходимо складові $u_\infty(s)$, $u^0(s)$, $u^\pm(s)$, $u^\Gamma(s)$ вектор-функції $u(s)$, а через неї співвідношеннями (4) – і компоненти шуканих об'ємних та гранично-поверхневих керувань $f(s)$, $f^\pm(s)$, $f^\Gamma(s)$.

Зауважимо, що аналогічно можуть бути розв'язані і задачі керування динамікою розглядуваної плити, коли воно виконується об'ємно розподіленим силовим фактором $f(s)$ (або $\bar{f}$) спільно з двома або навіть трьома поверхнево-граничними зовнішньодинамічними збуреннями. Особливістю розв'язання таких задач керування є те, що у розв'язуючих рівняннях (18), (19) будуть відсутні два або три рівняння, що відповідають неспостережуваним відповідно до (4) зовнішньодинамічним факторам.

1.4. Задача керування динамікою товстої пружної плити при відомих об'ємно розподілених зовнішньодинамічних збуреннях

Розглянемо особливості керування динамікою товстої пружної плити для випадку, коли об'ємно розподілені зовнішньодинамічні збурення $f(s)$ функціонально відомі, зовнішньодинамічний просторово-часовий стан визначається співвідношенням (7), а її бажаний стан (14) досягається дією одного, двох та більше доступних для цього початкових і поверхнево-крайових керуючих факторів $\bar{f}^0(\xi)$, $\bar{f}^\pm(s)$, $\bar{f}^\Gamma(s)$.

Обраний для реалізації керуючий фактор (один, чи кілька) визначимо співвідношенням (4) за відомої вектор-функції $u(s)$, яку побудуємо відповідно до (5), (9) або (5), (10) за наявності функцій $\bar{f}^0(\xi)$, $\bar{f}^\pm(s)$, $\bar{f}^\Gamma(s)$ або векторів $\bar{f}^0$, $\bar{f}^\pm$, $\bar{f}^\Gamma$ їхніх значень, якими згідно із (12), (15) моделюється доступна для спостережень картина (7) зовнішньодинамічного стану плити.

Визначення згаданих вище моделювальних векторів зведеться до розв'язання (середньоквадратичного обернення) систем (16), (17), побудованих з урахуванням постановки задачі керування (відомо, чим керується динаміка плити). Зокрема, у системах (16), (17), як і вище, будуть відсутні рівняння з f_{rl}^0 ($r = \overline{1, R_0}$, $l = \overline{1, L_0}$), $f_{\rho l}^\pm$ ($\rho = \overline{1, R_\pm}$, $l = \overline{1, L_\pm}$), $f_{\rho l}^\Gamma$ ($\rho = \overline{1, R_\Gamma}$, $l = \overline{1, L_\Gamma}$) у правій частині, якщо у списку керуючих зовнішньодинамічних факторів є збурення $\bar{f}^0(\xi)$, $\bar{f}^\pm(s)$, $\bar{f}^\Gamma(s)$ відповідно.

Загальним для всіх випадків керування у розв'язуючих рівняннях (16), (17) буде зміна вектора F у їхніх правих частинах на

$$\begin{aligned} \bar{F} = \text{col} & \left(\left(\left(\left(f_{rl}^0 - L_r^0(\partial_t) u_\infty(s) \right|_{t=0, \xi=\xi_l^0} \right), r = \overline{1, R_0} \right), l = \overline{1, L_0} \right), \right. \\ & \left(\left(\left(f_{\rho l}^+ - L_\rho^+(\partial_\xi) u_\infty(s) \right|_{s=s_l^+} \right), \rho = \overline{1, R_+} \right), l = \overline{1, L_+} \right), \left(\left(\left(f_{\rho l}^- - L_\rho^-(\partial_\xi) u_\infty(s) \right|_{s=s_l^-} \right), \rho = \overline{1, R_-} \right), l = \overline{1, L_-} \right), \\ & \left. \left(\left(\left(f_{\rho l}^\Gamma - L_\rho^\Gamma(\partial_\xi) u_\infty(s) \right|_{s=s_l^\Gamma} \right), \rho = \overline{1, R_\Gamma} \right), l = \overline{1, L_\Gamma} \right) \right) \end{aligned}$$

де вектор-функція $u_\infty(s)$ співвідношенням (6) визначається через об'ємно розподілене збурення $f(s)$.

Розв'язання модифікованих таким чином розв'язуючих рівнянь (16), (17) дозволить згідно з (5), (9) або (5), (10) визначити функцію $u(s)$, відповідну наявним зовнішньодинамічним умовам (7) функціонування плити. Шукані керуючі фактори будуть знайдені за допомогою співвідношення (4).

Точність розв'язання відповідної задачі та її однозначність, як і вище, будуть визначатися, умовами точності й однозначності розв'язання рівнянь (16), (17).

Дискусія і висновки

Отже, у роботі запропоновано і доведено до простих математичних співвідношень методику розв'язання складних задач керування динамікою тривимірного пружного тіла у формі плоскої плити без обмежень на форму її торцевої поверхні. Передбачалося, що динаміка плити визначається просторово розподіленими у внутрішньому об'ємі плити, а також початковими та гранично-поверхневими зовнішньодинамічними факторами, що можуть бути відомими, спостерігатися в дискретно заданих просторово-часових точках або розглядатися як керуючі впливи на плиту з метою

середньоквадратичного наближення до дискретно заданого стану. Розв'язано задачі побудови зазначених керуючих факторів, а також математичного моделювання останніх за дискретно визначеною реакцією плити на їхню наявність. Відсутність обмежень на кількість і якість інформації про зовнішньодинамічний стан плити робить задачі вивчення динаміки плити недовизначеними та математично некоректними. Проблеми дослідження динаміки таких плит ускладнюються під час розв'язання задач керування такою динамікою. І це при тому, що керуючими розглядаються всі з можливих зовнішньодинамічних впливів у довільній кількісній комбінації.

Для всієї різноманітності таких складних математичних задач керування складним механічним об'єктом у межах складної тривимірної математичної моделі подано розрахункові формули для визначення обраних керуючих факторів, а також для побудови викликаного ними тривимірного поля пружно-динамічних зміщень з оцінкою точності його узгодження із заздалегідь заданими бажаними значеннями.

Внесок авторів: Володимир Стоян – концептуалізація, методика, теоретичні основи дослідження; Дмитро Черній – методика, збір і перевірка даних; Сергій Волощук – методика, емпіричне дослідження, підготовка огляду літератури; Роман Заторський – емпіричне дослідження.

Джерела фінансування. Фінансування здійснюється власним коштом авторів та частково забезпечено Київським національним університетом імені Тараса Шевченка.

Список використаних джерел

- Григолюк, Э. И., & Селезов, И. Т. (1973). *Механика твердых деформируемых тел. Т. 5. Неклассические теории колебаний стержней, пластин и оболочек*. ВИНТИ.
- Нигул, У. (1963). О применении символического метода А.И. Лурье в трехмерной теории динамики упругих плит. *Известия АН Эстонской ССР. Серия: Физ.-мат. и техн. науки*, 12(2), 146–155.
- Селезов, И. Т. (1961). Про гіпотези, які лежать в основі уточнених рівнянь поперечних коливань пластин і деякі особливості цих рівнянь. *Прикладна механіка*, 7, 538–546.
- Стоян, В. А. (2016). *Методи математичного моделювання в задачах динаміки товстих пружних плит*. ВПЦ „Київський університет”.
- Стоян, В. А. (2024). *Розподілені просторово-часові системи в умовах невизначеності*. ВПЦ „Київський університет”.
- Стоян, В. А., Черній, Д. І., & Волощук, С. Д. (2024). Тривимірні задачі керування динамікою неповно спостережуваних товстих пружних плит. Частина І. Випадок неперервно заданого бажаного стану. *Вісник Київського національного університету імені Тараса Шевченка. Фізико-математичні науки*, 79(2), 65–71.
- Шереметьев, М. П., & Пелех, Б. Л. (1964). К построению уточненной теории пластин. *Инженерный журнал*, 4(3), 504–509.
- Grigorenko, Ya. M., & Grigorenko, A. Ya. (2013). Static and dynamic problems for anisotropic inhomogeneous shells with variable parameters and their numerical solution (review). *International applied mechanics*, 49(2), 123–193. <https://doi.org/10.1007/s10778-013-0558-x>
- Grigorenko, Ya. M., Savula, Ya. G., & Mukha, I. S. (2000). Linear and nonlinear problems of the elastic deformation of complex shells and methods of their numerical solution. *International applied mechanics*, 36(8), 979–1000. <https://doi.org/10.1023/A:1026645731095>
- Lur'e, A. I. (1964). *Three-Dimensional Problems of the Theory of Elasticity*. Interscience Publishers.
- Nemysh, Y. N., & Khoma, I. Y. (1993). Stress-strain state of non-thin plates and shells. Generalized theory (survey). *International applied mechanics*, 29(11), 873–902. <https://doi.org/10.1007/BF00848271>
- Pichkur, V. V., & Sobchuk, V. V. (2021). Mathematical model and control design of a functionally stable technological process. *Journal of Optimization, Differential Equations and Their Applications (JODEA)*, 29(1), 32–41. <http://dx.doi.org/10.15421/142102>
- Pichkur, V., Sobchuk, V., Cherniy, D., & Ryzhov, A. (2024). Functional stability of production processes as control problem of discrete systems with change of state vector dimension. *Bulletin of Taras Shevchenko National University of Kyiv. Physical and Mathematical Sciences*, 78(1), 105–110. <https://doi.org/10.17721/1812-5409.2024/1.21>
- Selezov, I. T. (2018). Development and application of the Cauchy–Poisson method to layer elastodynamics and the Timoshenko equation. *Cybernetics and Systems Analysis*, 54, 434–442. <https://doi.org/10.1007/s10559-018-0044-x>
- Stoyan, V. A. (2018). Three-dimensional integral mathematical models of the dynamics of thick elastic plates. *Cybernetics and Systems Analysis*, 54, 232–241. <https://doi.org/10.1007/s10559-018-0024-1>
- Timoshenko, S., & Woinowsky-Krieger, S. (1987). *Theory of plates and shells*. McGraw-Hill.

References

- Grigolyuk, E. I., & Selezov, I. T. (1973). *Mechanics of solid deformable bodies. Vol. 5. Non-classical theories of oscillations of rods, plates and shells*. VINITI [in Russian].
- Grigorenko, Ya. M., & Grigorenko, A. Ya. (2013). Static and dynamic problems for anisotropic inhomogeneous shells with variable parameters and their numerical solution (review). *International applied mechanics*, 49(2), 123–193. <https://doi.org/10.1007/s10778-013-0558-x>
- Grigorenko, Ya. M., Savula, Ya. G., & Mukha, I. S. (2000). Linear and nonlinear problems of the elastic deformation of complex shells and methods of their numerical solution. *International applied mechanics*, 36(8), 979–1000. <https://doi.org/10.1023/A:1026645731095>
- Lur'e, A. I. (1964). *Three-Dimensional Problems of the Theory of Elasticity*. Interscience Publishers.
- Nemysh, Y. N., & Khoma, I. Y. (1993). Stress-strain state of non-thin plates and shells. Generalized theory (survey). *International applied mechanics*, 29(11), 873–902.
- Nigul, U. (1963). On the application of the symbolic method of A.I. Lur'e in the three-dimensional theory of the dynamics of elastic plates. *Bulletin of the Academy of Sciences of the Estonian SSR. Series of physical, mathematical and technical sciences*, 12(2), 146–155 [in Russian].
- Pichkur, V. V., & Sobchuk, V. V. (2021). Mathematical model and control design of a functionally stable technological process. *Journal of Optimization, Differential Equations and Their Applications (JODEA)*, 29(1), 32–41. <http://dx.doi.org/10.15421/142102>
- Pichkur, V., Sobchuk, V., Cherniy, D., & Ryzhov, A. (2024). Functional stability of production processes as control problem of discrete systems with change of state vector dimension. *Bulletin of Taras Shevchenko National University of Kyiv. Physical and Mathematical Sciences*, 78(1), 105–110. <https://doi.org/10.17721/1812-5409.2024/1.21>
- Selezov, I. T. (1961). On the hypotheses underlying the refined equations of transverse plate vibrations and some features of these equations. *Applied mechanics*, 7, 538–546 [in Ukrainian].
- Selezov, I. T. (2018). Development and application of the Cauchy–Poisson method to layer elastodynamics and the Timoshenko equation. *Cybernetics and Systems Analysis*, 54, 434–442. <https://doi.org/10.1007/s10559-018-0044-x>
- Sheremet'yev, M. P., & Pelekh, B. L. (1964). Towards the construction of a refined theory of plates. *Engineering Journal*, 4(3), 504–509 [in Russian].

- Stoyan, V. A. (2016). *Methods of mathematical modeling in problems of dynamics of thick elastic plates*. VPTs Kyivs'kyi Universitet [in Ukrainian].
- Stoyan, V. A. (2024). *Distributed spatiotemporal systems under uncertainty*. VPTs Kyivs'kyi Universitet [in Ukrainian].
- Stoyan, V. A. (2018). Three-dimensional integral mathematical models of the dynamics of thick elastic plates. *Cybernetics and Systems Analysis*, 54(2), 232–241. <https://doi.org/10.1007/s10559-018-0024-1>
- Stoyan, V. A., Cherniy, D. I., & Voloshchuk, S. D. (2024). Three-dimensional problems of controlling the dynamics of incompletely observed thick elastic plates. Part II. The case of discretely specified desired condition. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics and Mathematics*, 79(2), 65–71 [in Ukrainian].
- Timoshenko, S., & Woinowsky-Krieger, S. (1987). *Theory of plates and shells*. McGraw-Hill.

Отримано редакцією журналу / Received: 03.01.25

Прорецензовано / Revised: 08.08.25

Схвалено до друку / Accepted: 10.10.25

Volodymyr STOYAN¹, DSc (Phys. & Math.), Prof.
 ORCID ID: 0009-0008-0746-5050
 e-mail: v_a_stoyan@ukr.net

Dmytro CHERNIY¹, DSc (Engin.), Assoc. Prof.
 ORCID ID: 0000-0002-6378-8048
 e-mail: d_cherniy@ukr.net

Serhii VOLOSHCHUK¹, PhD (Phys. & Math.), Assoc. Prof.
 ORCID ID: 0009-0000-6416-5566
 e-mail: sr.voloshchuk@gmail.com

Roman ZATORSKY², DSc (Phys. & Math.), Prof.
 ORCID ID: 0000-0001-7503-4932
 e-mail: romazatorsky@gmail.com

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine²Vasyl Stefanyk Carpathian National University, Ivano-Frankivsk, Ukraine

THREE-DIMENSIONAL PROBLEMS OF CONTROLLING THE DYNAMICS OF INCOMPLETELY OBSERVED THICK ELASTIC PLATES. PART II. THE CASE OF DISCRETELY SPECIFIED DESIRED CONDITION

For the first time, complex control problems for a thick elastic plate, the dynamics of which are described by Lamé's equations in a bounded three-dimensional spatial domain over a finite time interval, are being solved. The essence of these problems lies in the fact that the field of elastodynamic displacements of the plate under the influence of a control function should attain a desired state, specified discretely in space and time, or remain in the minimally possible least-squares neighborhood of these values, given the known conditions. The control functions can be chosen from volumetric, initial, surface, and edge functions of external dynamic influences, which can affect the plate's dynamics either individually or in practically feasible combinations of several such influences. These problems are solved given the presence of discretely defined external dynamic observations of the initial, boundary, and edge conditions of the plate at known points, which belong to the domains of definition of the aforementioned conditions. In fact, these observations are discrete initial, boundary, and edge conditions that are complemented with Lamé's equations. However, their quantity and quality can be arbitrary and will only affect the complexity of the computations. Based on the discrete initial, boundary, and edge conditions, along with the discrete desired states, and considering the integral form of the model, control functions are found as pseudosolutions of a system of algebraic equations or a system of integral equations. Analytical dependencies of the external dynamic control factors and the field of elastodynamic displacement of the points of the plate induced by them are also constructed and evaluated for accuracy and uniqueness. Partial cases of some possible combinations of control strategies for the dynamics of the thick elastic plate are also considered.

Keywords: spatially distributed dynamical systems, spatial problems of elasticity theory, thick elastic plates, control problems.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

UDC 519.2

DOI: <https://doi.org/10.17721/1812-5409.2025/2.33>

Andrij STOPKIN, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0002-6130-9920
e-mail: stepkin.andrej@gmail.com
SHEI "Donbas State Pedagogical University", Dnipro, Ukraine

Tetiana TURKA, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0001-6445-2223
e-mail: tvturka@gmail.com
SHEI "Donbas State Pedagogical University", Dnipro, Ukraine

ALGORITHM FOR RECOGNIZING SIMPLE GRAPHS BY A COLLECTIVE OF AGENTS

Abstract. These days, automation and robotization across various processes are developing rapidly. Naturally, this trend also extends to exploring environments where human operation is more difficult – or even dangerous – than operating under the same conditions with specialized robotic systems. This makes research on exploring unknown graphs (map construction) by mobile agents particularly relevant. This paper proposes a new algorithm for exploring finite, undirected, simple graphs by two different types of agents. One agent, the researcher, moves through the graph, reads and modifies marks on graph elements, and transmits information about its actions to the other agent – the experimenter. The researcher's operation is based on a depth-first graph traversal method. The researcher agent has finite memory that does not depend on the dimension of the graph. The experimenter agent operates outside the explored graph. Its main task is to build an internal representation of the explored graph (as a list of edges and nodes) based on information received from the researcher agent. The experimenter agent has finite but unboundedly growing memory, the size of which depends on the number of nodes in the explored graph. This paper examines in detail the researcher's modes of operation, indicating the priority of their activation depending on local conditions, and lists the messages exchanged by the agents during the algorithm. The complete algorithm for the experimenter to process the received messages – on the basis of which graph exploration takes place – is also provided. Additionally, the paper analyzes the time, space, and communication complexity of the algorithm and the number of edge traversals performed by the researcher during graph traversal. It is shown that the proposed algorithm has quadratic (from the number of nodes of the explored graph) time, space and communication complexity. The upper estimate of the number of transitions along the edges performed by the researcher agent is estimated as $O(n^2)$, where n is the number of nodes in the graph. For the work of proposed graph exploration algorithm, the researcher agent needs two paints of different colors and one stone.

Keywords: undirected graphs, graph exploration, agent collective, algorithm complexity, depth-first traversal method.

AMS 2020 classification: 05C85 68R10.

Introduction

A pressing problem in computer science, which has received considerable attention, is the interaction between a controlling system and a controlled system. Many works over the past decades have been devoted to solving this problem (Dopp, 1971; Dudek et al., 1993; Stepkin, 2015; Nagavarapu et al., 2020), under the assumption that it involves moving mobile agents through some environment toward a specific goal. It is quite clear that an agent's movement in an operational environment is possible only if a complete model of the chosen environment exists. In modeling environments, several approaches have been identified, one of which is topological. Under this approach, a roaming agent has access to information about connections between different regions of the environment but lacks metric and algorithmic information about it. Such a situation often arises in robotics (Dudek, 2024), and the topological model is taken to be a graph, equipped with additional information on edges, vertices, and incidences (points where edges meet vertices of the graph). Typically, three main tasks of environment exploration by an agent are distinguished: (1) agent self-localization within the environment; (2) verification of the model's correspondence to the actual environment; (3) construction by the agent of a map model of the environment. A significant number of papers (Fraigniaud et al., 2004) are devoted to the problem of environment recognition (map model construction), and several recognition algorithms are known, yielding many results on the feasibility and complexity of such recognition using various agent capabilities. However, the problem of graph recognition by teams of agents remains under-investigated, making studies aimed at recognizing (mapping) unknown graphs by various teams of mobile agents especially relevant.

To address the problem of graph recognition by a collective of agents, a number of approaches have been defined and various algorithms proposed for moving a team of agents through a graph and marking its elements with paints or stones, enabling recognition of the target graph up to isomorphism. The first works on graph exploration by an agent swarm appeared in the 1990s (Dudek et al., 1993; Dudek et al., 1998). In these publications, all agents start from a single vertex. They schedule their next meeting time, divide the available edges among themselves, and disperse along them. After a predetermined interval, having traversed part of the graph, the agents return to the starting vertex to merge their partial maps. They then set the next meeting time and again spread out through the graph. This process repeats until a complete map of the explored graph is built. Similar swarm-based algorithms were proposed in (Wang, Jenkin, & Dymond, 2009; Zhang, 2010). By contrast, in (Das et al., 2007) an algorithm is presented in which agents start from different vertices, and communication occurs via tokens that agents leave at graph vertices.

In (Stepkin, 2015), the problem of recognizing finite, simple, undirected graphs by a collective of agents is examined. Two researcher agents move through the graph simultaneously, read and modify labels on graph elements, and send the necessary information to an experimenter agent, which constructs the map of the explored graph in its memory as lists of edges and vertices. For recognition, each moving agent uses two different paints (three colors in total). The method is based on a depth-first search traversal.

In the paper (Nagavarapu, Vachhani & Sinha, 2016), a decentralized approach to map-building of graphs by a collective of agents is proposed, in which agents avoid collisions by using marks placed by agents on previously visited vertices of the graph. In this algorithm, decision-making lies directly with each agent and does not depend on the actions of other agents.

© Stopkin Andrij, Turka Tetiana, 2025

By contrast, in the paper (Banfi et al., 2018), agents exchange information about completed tasks with a base station via a dedicated network. In other words, to organize coordinated action among agents, it is first necessary to establish a network that meets certain conditions.

The object of our research is the problem of constructing maps of the explored graphs by a collective of mobile agents.

The goal of the research is to develop a new algorithm for recognizing simple undirected graphs using two agents, serving as a base algorithm for further scaling to a larger number of agents traversing the graph.

1. Main Results

We consider connected, finite, undirected simple graphs (without loops or multiple edges). Let $G = (V, E)$ – be a graph with vertex set V and edge set E . An edge is a two-element subset $\{u, v\}$ of V ; its endpoints u, v are called adjacent vertices, and the edge (u, v) is said to be incident to both u, v . We may denote the edge either as (u, v) or (v, u) . A triple $((v, u), u)$ is called the incidence (or point of contact) of edge (v, u) with vertex u . The incidence $((v, u), u)$ is called the far incidence of v , and $((v, u), v)$ the near incidence. Denote by I the set of all such triples. The set $L = V \cup E \cup I$ is called the set of elements of the graph G . A coloring function of the graph G is a surjective map $\mu: L \rightarrow \{w, r, b\}$, where w is white, r is red, and b is black. The pair (G, μ) is called a colored graph. A sequence $u_1, u_2, \dots, u_k$ of pairwise adjacent vertices is called a path in G , and k is the length of that path. If $u_1 = u_k$, the path is called a cycle. The neighborhood $Q(v)$ of a vertex v is the set of graph elements consisting of the vertex v , all vertices u adjacent to v , all edges (v, u) , and all incidences $((v, u), v)$ and $((v, u), u)$. We denote the power of the sets of vertices V and edges E by n and m , respectively. It is clear that $m \leq \frac{n(n-1)}{2}$. An isomorphism between graphs G and H is a bijection $\varphi: V_G \rightarrow V_H$ such that $(v, u) \in E_G$ if and only if $(\varphi(v), \varphi(u)) \in E_H$. Thus, isomorphic colored graphs differ only by renaming of vertices and recoloring of their elements.

In this paper, we study a team of two agents: a researcher and an experimenter. The researcher is capable of moving along edges of the graph and recoloring graph elements. The experimenter agent builds and stores the graph map in its memory based on messages received from the researcher. The researcher agent has finite memory that does not depend on the dimension of the graph under investigation; the experimenter agent, in turn, has finite but unboundedly growing internal memory.

At the start of its operation, the researcher agent is placed at an arbitrary vertex of the graph G and paints it in its own color. The experimenter agent at that moment adds a new vertex to the set V_H in its memory. The researcher agent moves through the graph from vertex v to vertex u along edge (v, u) , and can change the coloring of vertices v, u , edges (v, u) , and incidences $((v, u), v), ((v, u), u)$. While at vertex v , the researcher agent can perceive the labels of all elements in the neighborhood $Q(v)$, and based on this information determines which edge it will move along next and how it will color the graph's elements. The experimenter agent, in turn, sends, receives, and identifies messages received from the researcher agent and records in its memory the result of the researcher agent's operation at each step. Based on this information, the experimenter agent builds a map of the graph G , initially unknown to the agents, as lists of edges and vertices.

Note that the proposed algorithm has several features: first, the graph G is unknown to the agents; second, the proposed method of recognizing the graph is based on a depth-first search strategy; third, during the traversal of graph G , the agents create an implicit numbering of visited vertices. Based on the constructed numbering, the recognition of graph G occurs by building a graph H isomorphic to G . In the course of traversal, the researcher agent builds an implicit depth-first search tree, all edges of which are divided into those that are colored red upon first traversal and belong to the tree itself, and those that do not belong to the tree and are colored black upon first traversal – back edges. Tree edges during the execution of the algorithm are traversed at least twice, and on the last traversal they are colored black by the agent. Back edges are traversed from one to two times.

Red vertices of graph G at each step of the algorithm form a red path. When moving into a new vertex, the red path is extended; when moving back, it is shortened; when recognizing a back edge, it does not change. A vertex whose all incident edges have been recognized is colored black. The algorithm finishes when the red path becomes empty and all vertices are black.

The operation of the researcher agent can be divided into two modes: Normal mode. The researcher agent moves forward over white vertices, coloring the vertices, the edges connecting those vertices, and the far incidences red. If in the neighborhood of the current vertex there are no white edges or vertices – that is, no possible paths of movement – then the researcher agent returns back, coloring the traversed vertices, edges, and near incidences black. Upon returning to the starting vertex, the researcher agent terminates. At each step, the researcher agent exchanges data with the experimenter agent. Recognition of back edges. If, while moving forward at a vertex v , a back edge is detected, then the researcher agent stops working in normal mode and switches to back-edge recognition mode. The researcher agent leaves a stone in the current vertex. After that, it begins to move backward along its path until it finds the vertex incident to the back edge (this refers to a white edge whose far endpoint is colored red and in which the stone is located) and traverses that edge, coloring it black. Then it returns to the previous vertex and continues moving backward. At this stage, two cases are possible: 1) the researcher agent finds another back edge. In this case, the researcher agent repeats the back-edge recognition procedure. 2) The researcher agent reaches the starting vertex, which means that all marked back edges have been recognized. After this, the agent returns to the end of the red path where the stone is located, picks up the stone, and continues working in normal mode.

Let us examine in more detail the agents' algorithms.

Algorithm of the researcher agent:

1. $\mu(v) := r$;
2. if $\exists (v, u) \in Q(v) | (\mu(v, u) = w) \text{ and } (\mu(u) = \mu(v) = r)$ then EXPL_IE(v);
3. else if $\exists (v, u) \in Q(v) | (\mu(v, u) = w) \text{ and } (\mu(u) = w)$ then do
4. FORWARD(v);
5. go to 2;
6. end do;
7. else if $\exists (v, u) \in Q(v) | (\mu((v, u), v) = r) \text{ and } (\mu(v, u) = r)$ then do

```

8.          BACK(v);
9.          go to 2;
10.         end do;
11.        else STOP(v);
EXPL_IE(v):
1.  agent leaves the stone  $\mu(v) := l$ ;
2.  if  $\exists (v, u) \in Q(v) | (\mu((v, u), v) = r) \text{ and } (\mu(v, u) = r) \text{ and } (\mu((v, u), u) = w)$ 
3.    then do
4.      agent selects  $(v, u) \in Q(v) | (\mu((v, u), v) = r) \text{ and } (\mu(v, u) = r) \text{ and } (\mu((v, u), u) = w)$  follows it to the vertex  $u$ ;
5.       $v := u$ ;
6.      agent appends RETURN to its message list  $M$ ;
7.      go to 2 of this procedure;
8.    end do;
9.  if  $\exists (v, u) \in Q(v) | (\mu(v) = r) \text{ and } (\mu(v, u) = w) \text{ and } (\mu(u) = l)$ 
10.   then do
11.     agent moves along edge  $(v, u)$ , coloring:  $\mu((v, u), v) := r, \mu(v, u) := r, \mu((v, u), v) := r$ ;
12.      $v := u$ ;
13.     agent appends FORWARD_IE to its message list  $M$ ;
14.     agent selects any edge  $(v, u) \in Q(v)$  such that:  $(\mu((v, u), v) = r) \text{ and } (\mu(v, u) = r) \text{ and } (\mu((v, u), u) = r)$ ,
        moves along it to the vertex  $u$ , coloring
         $(\mu((v, u), v) = b) \text{ and } (\mu(v, u) = b) \text{ and } (\mu((v, u), u) = b)$ ;
15.     go to 2 of this procedure;
16.   end do;
17.   else go to 2 of this procedure;
18. if  $\exists (v, u) \in Q(v) | (\mu((v, u), v) = w) \text{ and } (\mu(v, u) = r) \text{ and } (\mu((v, u), u) = r)$ 
19.   then do
20.     agent selects  $(v, u) \in Q(v)$  such that:  $(\mu((v, u), v) = w) \text{ and } (\mu(v, u) = r) \text{ and } (\mu((v, u), u) = r)$  and moves
        through it to vertex  $u$ ;
21.      $v := u$ ;
22.     go to 18 of this procedure;
23.   end do;
24.   else do
25.     agent picks up the stone  $\mu(v) := r$ ;
26.     agent appends RECOGNIZED_IE to its message list  $M$ ;
27.   end do

```

During the FORWARD(v) procedure, the researcher agent selects from the neighborhood $Q(v)$ an arbitrary edge (v, u) for which $(\mu(v, u) = w) \text{ and } (\mu(u) = w)$, then moves along it to vertex u . In doing so, it colors $(\mu(v, u) := r) \text{ and } (\mu((v, u), u) := r) \text{ and } (\mu(u) := r)$, performs the assignment $v := u$, and appends the message FORWARD to the list M .

During the BACK(v) procedure, the researcher agent selects from the neighborhood $Q(v)$ an edge for which $(\mu(v) = r) \text{ and } (\mu((v, u), v) = r) \text{ and } (\mu(v, u) = r)$, then moves along it to vertex u . In doing so, it colors $\mu(v) := b, \mu((v, u), v) := b, \mu(v, u) := b$, performs the assignment $v := u$, and appends the message BACK to the list M .

During the STOP(v) procedure, the researcher agent colors $\mu(v) := b$, appends the message STOP to the list M , and terminates.

Algorithm of the experimenter agent:

Input: list of messages M from the researcher agent.

Output: lists of vertices V_H and edges E_H of the graph H , which is isomorphic to G .

Data: V_H, E_H – lists of vertices and edges of graph H , isomorphic to G . ct – a counter of how many vertices of G have been visited by the researcher agent. M – the list of messages from the researcher agent. i – a counter used to determine the indices of far endpoints of back edges after a stone has been placed at their common vertex. $STOP$ – a variable used by the researcher agent to signal to the experimenter agent that the graph traversal is complete. $r(1) \dots r(t)$ – the list of indices of the vertices on the red path, where t is the length of this list. Mes – a variable that holds the current message.

Algorithm of the experimenter agent:

```

1.   $ct := 1, M := \emptyset, i := 0, E_H := \emptyset, STOP := 0, t := 1, r(t) := ct, V_H := \{1\}, Mes := ""$ ;
2.  while  $STOP = 0$  do
3.    if  $M \neq \emptyset$ ,
4.      then do
5.        read the next message into  $Mes$  and remove it from the queue  $M$ ;
6.        LIST_PROCESSING();
7.      end do;
8.    end do;
9.  print  $V_H, E_H$ .
LIST_PROCESSING():
1.  if  $Mes := "FORWARD"$ ; then FORWARD();
2.  if  $Mes := "BACK"$  then BACK();

```

```

3.   ifMes := "RETURN" then RETURN();
4.   ifMes := "FORWARD_IE" then FORWARD_IE();
5.   ifMes := "RECOGNIZED_IE" then RECOGNIZED_IE();
6.   ifMes := "STOP" then STOP();
FORWARD():
1.   ct := ct + 1;
2.   t := t + 1;
3.   r(t) := ct;
4.    $V_H := V_H \cup \{ct\}$ ;
5.    $E_H := E_H \cup \{r(t-1), r(t)\}$ ;
BACK():
1.   Delete the element  $r(t)$  from the list  $r(1) \dots r(t)$ ;
2.    $t := t - 1$ ;
RETURN():
1.    $i := i + 1$ ;
FORWARD_IE():
1.    $E_H := E_H \cup \{r(t), r(t-i)\}$ ;
RECOGNIZED_IE():
1.    $i := 0$ ;
STOP():
1.   STOP := 1;

```

3. Algorithm Analysis

Lemma 1. *Two agents, while executing the recognition algorithm on a graph, uniquely identify (recognize) all back edges.*

Proof. During the EXPL_IE(v) procedure, the agents recognize back edges. Specifically, when moving forward, if the researcher agent notices a back edge originating from the current vertex, it marks that vertex with a stone l (see line 1 of EXPL_IE(v)) and begins moving backward along the red path. At each step, it sends the message "RETURN" to the experimenter agent (line 6 of EXPL_IE(v)). Based on this "RETURN" message, the experimenter agent's RETURN() procedure counts how many edges of the red path have been traversed backward.

If at some vertex during this backward traversal the researcher agent detects a white edge whose far endpoint is marked by the stone l , then it traverses that edge, coloring the near incidence, the edge itself, and the far incidence all in red (line 11 of EXPL_IE(v)), and sends the message "FORWARD_IE" to the experimenter agent (line 13 of EXPL_IE(v)). Upon receiving "FORWARD_IE", the experimenter agent's FORWARD_IE() procedure determines the previously unknown implicit index of the second endpoint of the back edge. Since the agents already know the implicit index of the vertex where the stone was placed, the experimenter agent can add this back edge to the edge set of graph H using the same procedure.

After handling the back edge, the researcher agent returns to the previous vertex, coloring the near incidence, the edge, and the far incidence through which it traversed all in black (line 14 of EXPL_IE(v)) and continues moving backward along the red path until it reaches its very beginning, thereby recognizing all back edges incident to the vertex marked by the stone.

Having reached the start of the red path, the researcher agent then returns to its end and sends the message "RECOGNIZED_IE" to the experimenter agent (line 26 of EXPL_IE(v)). Upon receiving this message, the experimenter agent's RECOGNIZED_IE() procedure resets its counter of traversals to zero.

Thus, for each back edge incident to the vertex marked by the stone (in fact, this is the vertex at the end of the red path at that stage of the graph traversal), the agents count the number of transitions along vertices of the red path required to reach the second endpoint of the given back edge and uniquely determine its index. This, in turn, allows them to unambiguously identify the back edge and add it to the edge set E_H of graph H . Q.E.D.

Theorem 1. *Two agents, after executing the recognition algorithm on a graph, will recognize any graph up to isomorphism.*

Proof. At the beginning of the recognition algorithm (for $n \geq 3$), each time the researcher agent visits a white vertex of the explored graph G , the procedure FORWARD(v) is invoked at least twice. Each invocation of FORWARD() creates a new vertex in graph H . Thus, the execution of this algorithm induces a mapping $\varphi: V_G \rightarrow V_H$ from the vertices of G to the vertices of H . Moreover, the equality $\varphi(v) = ct$ is established when vertex v is colored red.

This mapping naturally establishes an implicit numbering of the vertices of graph G . Moreover, the mapping φ is a bijection because we consider only connected graphs, and in such graphs every vertex is reachable from any starting vertex. Therefore, all vertices are visited by the researcher agent, and upon its first visit, the agents assign it an implicit number.

From the algorithm's description, it is clear that the researcher agent traverses every edge of G , since after the algorithm completes, all edges of the graph are colored black.

When executing the procedure FORWARD(), the experimenter agent recognizes a tree edge (v, u) and numbers vertex u so that the edge in graph G corresponds uniquely to edge $(\varphi(v), \varphi(u))$ in graph H . When executing the procedure FORWARD_IE(), the experimenter agent recognizes a back edge (v, u) of G and associates it uniquely with edge $(\varphi(v), \varphi(u))$ in H . Since all edges of the graph are divided exclusively into tree edges and back edges, it follows that the mapping φ is an isomorphism from G to H . Q.E.D.

Theorem 2. *The time, space, and communication complexities of the recognition algorithm, as well as the number of edge traversals performed by the researcher agent, are bounded above by $O(n^2)$, where n is the number of vertices in the graph. Two different-colored paints suffice for recognition.*

Proof. Let us examine more closely the properties of the red path. From the algorithm's description, at each step the red path is a simple path that starts at the vertex v with $\varphi(v) = 1$ and ends at the vertex u with $\varphi(u) = ct$. Hence, the length of the red path does not exceed n . Each time the procedures FORWARD(v) and BACK(v) are executed, the researcher agent traverses exactly one edge. In a single run of the back-edge recognition mode, the researcher agent recognizes at most $n - 2$

back edges, during which it: spends one move to place the stone in a vertex; traverses at most $n - 1$ edges of the red path twice (once moving backward and once returning to the last red vertex); traverses at most $n - 2$ back edges twice while coloring them.

When analyzing the time complexity of the algorithm, we assume that algorithm initialization and the choice of any one of the possible procedures each take some constant amount of time. We also assume that traversing an edge by the researcher agent and processing the command by the experimenter agent – received at that step from the researcher agent – each take time equal to some constant. We further assume that the time to send a single message equals a constant that does not exceed the time it takes for the researcher agent to traverse one edge. Moreover, the total time spent on analyzing the neighborhood $Q(v)$ of vertex v and choosing the necessary edges is bounded above by $O(n^2)$.

Then the time complexity of the algorithm is determined by the following relations:

1. The procedures FORWARD(v) and BACK(v) are each executed no more than $2 \times (n - 1)$ times, and the total time for their execution is $O(n)$.

2. The time spent recognizing back edges is $(n - 2) \times (2 \times (n - 1) + 2 \times (n - 2))$ steps, i.e. essentially $O(n) \times O(n)$ steps. Therefore the total time for executing the back-edge recognition procedure is $O(n^2)$.

Hence, the upper bound on the number of edge traversals performed by the researcher agent is $O(n) + O(n^2) = O(n^2)$, and the overall time complexity $T(n)$ of the algorithm satisfies $T(n) = O(n^2)$.

The space complexity $S(n)$ of the algorithm is determined by the combined sizes of the lists E_H , V_H , and $r(1), \dots, r(t)$, whose complexities are $O(n^2)$, $O(n)$, and $O(n)$, respectively.

Thus, the upper bound on the space complexity of the algorithm satisfies $S(n) = O(n^2)$.

When calculating the communication complexity of the algorithm, noting that the researcher agent can send the experimenter agent only one of six possible messages (FORWARD, BACK, RETURN, FORWARD_IE, RECOGNIZED_IE, STOP), we assume that each message occupies one unit of memory.

Since in normal mode, at each step exactly one message (FORWARD, BACK, or STOP) is transmitted, the total volume of information sent in this mode is bounded above by $2 \times O(n)$. In back-edge recognition mode, while the researcher agent moves backward along the red path, it sends one message (RETURN) at each step, so the total volume of information sent in that mode is also bounded above by $n \times O(n)$. At the first traversal of a back edge, one message (FORWARD_IE) is sent, so the total volume of information in that part is also $n \times O(n)$. No messages are sent when returning along a back edge. Finally, one message (RECOGNIZED_IE) is sent when returning to the vertex containing the stone, so that volume is $n \times O(1)$.

Thus, the total volume of information transmitted by the agents is

$$K(n) = 2 \times O(n) + n \times (O(n) + O(n) + O(1)) = O(n^2) \text{ Q.E.D.}$$

Discussion and conclusions

The paper proposes a new algorithm for recognizing finite, undirected, simple graphs (without loops or multiple edges) using two agents. One agent – the researcher agent – traverses the graph, reads and updates labels on graph elements, and transmits information about its actions to the other agent, the experimenter agent. The researcher's operation is based on a depth-first search traversal. The researcher has finite memory that does not depend on the graph's dimension. The second agent – the experimenter – operates outside the explored graph. Its primary task is to construct an internal representation of the graph in its memory based on information received from the researcher. The experimenter's memory is finite at each step but unbounded overall, growing with the number of vertices in the graph.

The paper also analyzes the time, space, and communication complexities of the proposed algorithm and examines the number of edge traversals performed by the researcher during the graph traversal. It is shown that the algorithm's time, space, and communication complexities are all quadratic in the number of vertices n . The upper bound on the number of edge traversals by the researcher is $O(n^2)$. To execute this graph-recognition algorithm, the researcher needs only two different-colored paints and one stone.

The results of this study are planned to be scaled up to a larger number of agents to develop more efficient graph-recognition algorithms.

Authors' contributions: Andrij Stopkin – conceptualization, formal analysis, methodology, drafting of the original manuscript, source analysis, preparation of the literature/theoretical background review; Tetiana Turka – data validation, writing – review and editing.

Acknowledgment. The authors is grateful to the referee for comments and suggestions that improved the paper.

Sources of funding. This study did not receive any grant from a funding institution in the public, commercial, or non-commercial sectors.

References

- Banfi, J., Quattrini Li, A., Rekleitis, I., Amigoni, F., & Basilico, N. (2018). Strategies for coordinated multirobot exploration with recurrent connectivity constraints. *Autonomous Robots*, 42, 875–894. <https://doi.org/10.1007/s10514-017-9652-y>
- Das, S., Flocchini, P., Kuttan, S., Nayak, A., & Santoro, N. (2007). Map construction of unknown graphs by multiple agents. *Theoretical Computer Science*, 385(1–3), 34–48. <https://doi.org/10.1016/j.tcs.2007.05.011>
- Dopp, K. (1971). Automaten in labirinth. *Elektronische Informationsverarbeitung und Kybernetik*, 7(2), 79–94.
- Dudek, G., & Jenkin, M. (2024). *Computational principles of mobile robotics* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108682404>
- Dudek, G., Jenkin, M., Milios, E., & Wilkes, D. (1993). Map validation in a graphlike world. *International Joint Conference on Artificial Intelligence: Proceedings of the 13th International Conference, Chambéry (France), San Fransisco (USA)*, (pp. 1648–1653).
- Dudek, G., Jenkin, M., Milios, E., & Wilkes, D. (1998). Topological exploration with multiple robots. *Robotics with Application (ISORA): Proceedings of the 7th International Symposium, Alaska*.
- Fraigniaud, P., Jecincas, D., Peer, G., Pelc, A., & Peleg, D. (2004). Graph Exploration by a finite automaton. *Proceedings of the 29th International Symposium on Mathematical Foundation of Computer Science (MFCS)*, (pp. 451–462).

- Nagavarapu, S. C., Vachhani, L., Sinha, A., & Buriuly, S. (2020). Generalizing Multi-agent Graph Exploration Techniques. *International Journal of Control, Automation and Systems*, 19, 491–504. Springer. <https://doi.org/10.1007/s12555-019-0067-8>
- Nagavarapu, S. C., Vachhani, L. & Sinha, A. (2016). Multi-Robot Graph Exploration and Map Building with Collision Avoidance: A Decentralized Approach. *Journal of Intelligent & Robotic Systems*, 83, 503–523. Springer. <https://doi.org/10.1007/s10846-015-0309-9>
- Stepkin, A. (2015). Using a Collective of Agents for Exploration of Undirected Graphs. *Cybernetics and Systems Analysis*, 51(2), 223–233. <https://doi.org/10.1007/s10559-015-9715-z>
- Wang, H., Jenkin, M., & Dymond, P. (2009). It can be beneficial to be 'lazy' when exploring graph-like worlds with multiple robots. *Advances in Computer Science and Engineering (ACSE): In Proceedings of the IASTED International Conference*, (pp. 55–60).
- Zhang, C. (2010). *Parallelizing Depth-First Search for Robotic Graph Exploration*. Harvard University.

Отримано редакцією журналу / Received: 16.06.25

Прорецензовано / Revised: 19.07.25

Схвалено до друку / Accepted: 10.10.25

Андрій СТЬОПКИН, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0002-6130-9920

e-mail: stepkin.andrej@gmail.com

ДВНЗ "Донбаський державний педагогічний університет", Дніпро, Україна

Тетяна ТУРКА, канд. фіз.-мат. наук, доц.

ORCID ID: 0000-0001-6445-2223

e-mail: tvturka@gmail.com

ДВНЗ "Донбаський державний педагогічний університет", Дніпро, Україна

АЛГОРИТМ РОЗПІЗНАВАННЯ ПРОСТИХ ГРАФІВ КОЛЕКТИВОМ АГЕНТІВ

Нині досить динамічно розвиваються напрямки автоматизації та роботизації різних процесів. Звісно, це стосується і дослідження різноманітних середовищ, де організація роботи людини є складнішою або навіть небезпечнішою, ніж організація роботи в тих самих умовах спеціальних роботизованих систем. Це робить актуальними дослідження, спрямовані на розпізнавання (побудову мапи) невідомих графів мобільними агентами. У представлений роботі запропоновано новий алгоритм розпізнавання скінчених, неорієнтованих графів без петель і кратних ребер за допомогою двох агентів різного типу. Один з агентів – агент-дослідник, що пересувається по графу, зчитує та змінює мітки елементів графа і передає інформацію про свої дії іншому агенту – агенту-експериментатору. Робота агента-дослідника базується на методі обходу графа у глибину. Агент-дослідник має скінчену пам'ять, що не залежить від розмірності графа. Другий агент – це агент-експериментатор, що перебуває поза межами досліджуваного графа. Головним завданням цього агента є побудова представлення досліджуваного графа (у вигляді переліку ребер і вершин) у своїй пам'яті на основі інформації, яку він отримує від агента-дослідника. Агент-експериментатор має скінчену внутрішню пам'ять, що необмежено зростає, розмірність якої залежить від кількості вершин графа над розпізнаванням якого працюють агенти.

Також детально розглянуто режими роботи агента-дослідника із зазначенням пріоритетності їхньої активації залежно від навантажувальних умов і перелік повідомлень, якими обмінюються агенти у процесі роботи алгоритму. Наведено повний алгоритм роботи агента-експериментатора з обробки отриманих повідомлень, на основі яких і відбувається розпізнавання графа. Виконано аналіз часової, ємнісної, комунікаційної складності побудованого алгоритму та проаналізовано кількість переходів по ребрах, які виконує агент-дослідник під час обходу графа. З'ясовано, що запропонований алгоритм має квадратичні (від кількості вершин досліджуваного графа) часову, ємнісну та комунікаційну складності. Верхню оцінку кількості переходів по ребрах, які здійснює агент-дослідник, оцінено як $O(n^2)$, де n – кількість вершин у графі. Для роботи запропонованого алгоритму розпізнавання графа агенту-досліднику необхідно дві фарби різного кольору й один камінь.

Ключові слова: неорієнтовані графи, розпізнавання графів, колектив агентів, складність алгоритму, обхід у глибину.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

УДК 004.89+51-76

DOI: <https://doi.org/10.17721/1812-5409.2025/2.34>

Іван ТЮРДЬО, асп.
ORCID ID: 0009-0001-7315-3628
e-mail: ivan.turdio@karazin.ua

Харківський національний університет імені В. Н. Каразіна, Харків, Україна

Наталія КІЗІЛОВА, д-р фіз.-мат. наук, проф.
ORCID ID: 0000-0001-9981-7616
e-mail: kizilova@karazin.ua

Харківський національний університет імені В. Н. Каразіна, Харків, Україна

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ В СУЧАСНІЙ АНТИРАКОВІЙ ТЕРАПІЇ

Метою пропонованого дослідження є аналіз сучасних підходів до оцінювання ефективності та безпечності лікарських засобів, які застосовують в антираковій терапії, із використанням методів машинного навчання. Особливу увагу приділено перспективам впровадження таких методів у математичну онкологію – галузь, що активно використовує математичне моделювання та комп'ютерні симуляції в онкологічних дослідженнях.

У межах роботи було здійснено пошук і вивчення актуальних наукових джерел, присвячених застосуванню машинного навчання в онкології. Також проведено систематичний огляд літератури, що висвітлює використання машинного навчання у цій сфері.

Проаналізовано основні підходи машинного навчання, такі як контрольоване навчання (Supervised Learning, SL), неконтрольоване навчання (Unsupervised Learning, UL) та навчання з підкріпленням (Reinforcement Learning, RL) у контексті сучасної онкології. Розглянуто конкретні приклади використання різних алгоритмів машинного навчання в дослідженнях, пов'язаних із лікуванням раку, а також загальніших онкологічних задач. Оцінено переваги й обмеження цих підходів залежно від поставлених цілей, наприклад у таких, як задачі прогнозування відповіді на лікування, ідентифікація біомаркерів, автоматизований аналіз медичних зображень.

За результатами роботи встановлено, що машинне навчання вже активно впроваджується в онкологічні дослідження та демонструє високу ефективність у розв'язанні різноманітних задач, зокрема виявляє приховані закономірності, що недоступно за традиційного аналізу. Водночас відзначено, що існують ще численні напрями, де застосування машинного навчання може суттєво посилити наукові розробки та клінічну практику. Зокрема, перспективним є використання алгоритмів навчання з підкріпленням у сфері персоналізованої прецизійної медицини, що відіграє ключову роль у створенні індивідуалізованих підходів до лікування онкологічних хвороб.

Ключові слова: математичне моделювання, контрольоване навчання, неконтрольоване навчання, навчання з підкріпленням, прецизійна медицина, персоналізована терапія.

Класифікація відповідно до AMS 2020: 91F10, 92F04, 68V11.

Вступ

Онкологічні захворювання й надалі залишаються однією з основних причин смертності у світі, незважаючи на досягнутий прогрес у методах діагностики та лікування. У світі у 2022 р. було зареєстровано близько 20 млн нових випадків раку, а кількість смертей від цього захворювання сягнула 9,7 млн (Bray et al., 2024). Хоча вже давно відомо, що ефективність протипухлинних препаратів може суттєво відрізнятись в різних пацієнтів, на практиці часто використовуються стандартні схеми лікування без урахування індивідуальних особливостей злоякісних клітин, що може знижувати ефективність терапії або навіть шкодити пацієнту (De los Rios de la Serna et al., 2024).

Однією з критично важливих проблем у сучасній медицині залишається своєчасне виявлення та запобігання побічним реакціям на лікарські засоби (Adverse Drug Reactions, ADR), а також зменшення медикаментозної токсичності. Незважаючи на суворе регулювання та проходження кількох фаз клінічних досліджень, побічні ефекти препаратів і надалі залишаються значущими чинниками підвищеної захворюваності та смертності (Nilashi et al., 2024). Традиційні методи моніторингу ризиків, такі як спостереження та збір повідомлень про побічні явища, зазвичай є реактивними й не завжди забезпечують своєчасне виявлення загроз, що можуть охопити велику кількість пацієнтів (Yang, & Kar, 2023).

Особливо гостро ця проблема виявляється у пацієнтів з онкологічними захворюваннями, які отримують комбінацію кількох лікарських засобів. Це значно підвищує ризик небажаних реакцій через можливі лікарські взаємодії. Наприклад, згідно з даними ретроспективного дослідження, близько двох третин онкологічних пацієнтів, госпіталізованих до стаціонару, отримували щонайменше одну потенційно небезпечну комбінацію препаратів (Riechelmann et al., 2008).

У зв'язку із цим пошук ефективних інструментів для ранньої діагностики ADR і зниження токсичності ліків є пріоритетним завданням як для фармацевтичної галузі, так і для медичної системи загалом (Siyun, & Supratik, 2023). Для розв'язання цих задач використовуються різноманітні джерела даних: від реакцій клітинних ліній на ліки до змін пухлинної маси, зафіксованих за допомогою МРТ або КТ (Bhandari et al., 2024).

Методи машинного навчання (Machine Learning, ML) відкривають нові можливості для аналізу таких великих і складних масивів медичних даних. У сфері математичної онкології – міждисциплінарної галузі, що поєднує математичне моделювання, біологію та комп'ютерну симуляцію – ML демонструє особливо високий потенціал (Rockne, & Scott, 2019). Застосування алгоритмів машинного навчання й аналізу великих даних дає змогу виявляти приховані взаємозв'язки, прогнозувати відповідь на лікування та знаходити чинники, що впливають на результативність терапії. Це значно сприяє підвищенню ефективності розроблення нових ліків, а також їхньому персоналізованому застосуванню в клінічній практиці (Terplytska et al., 2024).

1. Машинне навчання в сучасній онкології

Натепер існують два основні підходи до прогнозування побічних реакцій на лікарські засоби (ADR), що базуються на персоналізованих даних пацієнтів і їхній історії лікування: це методи статистичного аналізу виживання й алгоритми контрольованого машинного навчання. Серед статистичних методів для оцінки ймовірності виникнення ADR часто застосовують криві виживаності Каплана – Мейера, а також модель пропорційних ризиків Кокса, що допомагає проводити часовий аналіз розвитку побічних ефектів (Egualo et al., 2016). Водночас, незважаючи на популярність, модель Кокса має певні обмеження: вона передбачає суворе дотримання умов пропорційності ризиків, які не завжди відповідають реальним клінічним даним, а також не здатна обробляти складні та різномірні типи інформації, наприклад, медичні зображення або генетичні профілі.

Натомість методи машинного навчання (ML) демонструють ширші можливості у медичних дослідженнях. Їх уже активно застосовують для проєктування лікарських засобів (Guan, & Wang, 2024), прогнозування протипухлинних пептидів із використанням багатовимірного вилучення ознак (Liu, Li, & Chen, 2022) і пептидних послідовностей (Phan et al., 2022), ранньої діагностики ADR і визначення рівня токсичності препаратів (Badwan et al., 2023). Також методи ML успішно застосовують для передбачення ефективності протипухлинних засобів для монотерапії різних типів раку (Partin et al., 2023), раку молочної залози (Nilashi et al., 2024), або передбачення ефективності комбінацій ліків при раку молочної залози (Zhou et al., 2023), у моделі глибокої нейронної мережі (She et al., 2022), у комбінованій моделі (Fan, Cheng, & Li, 2021), включно з оцінкою їхньої синергетичної взаємодії (Shan et al., 2023) і прогнозуванням часу виникнення ADR (Timilsina, Tandan, & Nováček, 2022).

Загалом методи машинного навчання умовно поділяють на три основні категорії: контрольоване навчання (Supervised Learning, SL), неконтрольоване навчання (Unsupervised Learning, UL) та навчання з підкріпленням (Reinforcement Learning, RL) (Siddharth et al., 2022). Кожен із цих підходів має свої унікальні переваги й застосовується для розв'язання певного класу завдань у математичній онкології.

2. Контрольовані методи

Алгоритми контрольованого навчання SL базуються на наявності попередньо підготовлених знань і є надзвичайно ефективними для прогнозування нових біологічних закономірностей, виявлення специфічних характеристик клітинних ліній або встановлення потенційних взаємозв'язків. Основою для навчання таких моделей зазвичай постає "золотий стандарт" – еталонний набір даних, який максимально точно відображає об'єкт або явище, що вивчається (Triguero, García, & Herrera, 2015). Однак у випадку онкологічних досліджень часто спостерігаємо нестачу достовірних або повних даних, що вимагає від моделей підвищеної стійкості до невизначеності у вхідній інформації.

SL-методи машинного навчання здатні виконувати класифікацію, тобто формування правил, що допомагають розрізнити об'єкти (напр., гени чи клітинні лінії), які належать до різних категорій. Серед найпоширеніших алгоритмів, що застосовуються для таких завдань, – метод опорних векторів (Support Vector Machine, SVM) (Cortes, & Vapnik, 2009) і штрафна логістична регресія (Tibshirani, 2011).

Алгоритми SVM мають здатність ефективно навчатися навіть на складних задачах і можуть використовувати нелінійні функції для моделювання складних зв'язків. Ці моделі успішно застосовували для класифікації пухлинних зразків на ті, що є чутливими або стійкими до хіміотерапевтичного лікування. Крім того, SVM-регресію використовували для ранжування потенційних лікарських засобів і прогнозування відповіді пухлини на лікування (Kurilov, Haibe-Kains, & Brors, 2020).

Окрім класифікаційних задач, SL дає змогу розв'язувати проблеми регресії, тобто прогнозування безперервних значень. У цьому контексті найпоширенішими є методи лінійної регресії або регресії, що базується на опорних векторах. Алгоритми лінійної регресії вирізняються простотою реалізації та широкою сферою застосування. Незважаючи на свою базову структуру, лінійні моделі залишаються одним із ключових інструментів у статистичному моделюванні та машинному навчанні (Hastie, Tibshirani, & Friedman, 2009):

$$f(x) = \beta_0 + \sum_{j=1}^p x_j \beta_j, \quad (1)$$

де $f(x)$ – прогнозоване значення залежної змінної; β_0 – вільний член; x_j – значення j -ої незалежної змінної; β_j – коефіцієнт при j -ій змінній; p – кількість незалежних змінних.

Метод найменших квадратів добре працює із простими наборами даних, однак зі збільшенням складності моделі виникає ризик перенавчання – ситуації, коли модель має низьке зміщення, але високу дисперсію. Для розв'язання цієї проблеми застосовують різні регуляризаційні підходи, зокрема гребеневу регресію (ridge regression), регресію Ласо (Lasso regression) та метод еластичної сітки (elastic net), що забезпечують кращий баланс між узагальненням і точністю (Rafique, Islam, & Kazi, 2021).

Серед різних методів SL дерево рішень (Decision Tree) є одним із найпопулярніших для задач класифікації. Цей підхід починається з кореневого вузла, що охоплює весь набір даних, і далі розгалужується на підмножини шляхом послідовного поділу за логічними умовами, допоки не буде досягнуто кінцевого результату в листових вузлах. Кожен внутрішній вузол дерева відповідає певному атрибуту, що визначає напрямок подальшого розгалуження (Podgorelec et al., 2002).

Дерева рішень належать до так званих "моделей білої скриньки", оскільки їхні рішення легко візуалізувати й інтерпретувати. Проте у разі збільшення складності й обсягу даних, вони можуть втрачати здатність до узагальнення й демонструвати перенавчання. Для подолання цього обмеження були розроблені вдосконалені ансамблеві методи, що об'єднують декілька слабких моделей у потужну сукупну модель з вищою точністю та стабільністю. До таких підходів належать Random Forest і Gradient Boosting Machines (GBM).

У моделі Random Forest навчальний набір багаторазово випадковим чином поділяється на підмножини, для кожної з яких будують окреме дерево рішень. Окрім цього, існують послідовні ансамблеві алгоритми, такі як AdaBoost та Gradient Boosting. В AdaBoost кожна наступна модель фокусується на виправленні помилок попередніх

класифікаторів, тоді як Gradient Boosting оптимізує функцію втрат, використовуючи дерева більшої глибини. Обидва методи демонструють вищу ефективність порівняно з окремими деревами (Rafique, Islam, & Kazi, 2021).

Однак, попри високу точність, ансамблеві моделі втрачають прозорість у процесі хвалення рішень. Через велику кількість компонентів стає складно визначити, які саме вхідні характеристики найбільше вплинули на результат. Це знижує їхню придатність для використання у критичних прикладних галузях, де потрібна інтерпретованість, що характерна для простіших моделей, як-от лінійні моделі або окремі дерева (Bertsimas & Wiberg, 2020).

До інтуїтивно зрозумілих і простих методів класифікації й регресії належить також метод *k*-найближчих сусідів (*k*-Nearest Neighbors, KNN). У цьому підході прогноз для нового зразка базується на ознаках найбільш схожих на нього вже відомих зразків – так званих сусідів. KNN успішно застосовували для задач прогнозування відповіді на лікарські засоби (Drug Response Prediction, DRP), де чутливість нового зразка визначали як середнє значення чутливості його найближчих сусідів (Li, Umbach, & Krahn, 2021). Цей підхід допоміг виявити потенційно нові асоціації між ліками й генами. Водночас метод KNN є чутливим до "зашумлених" або нерелевантних ознак і потребує значних обчислювальних ресурсів, що обмежує його ефективність під час роботи з великими даними (Firoozbakht, Yousefi, & Schwikowski, 2021).

Іншим прикладом потужного інструменту SL, який активно застосовується в онкології, є штучні нейронні мережі (Artificial Neural Networks, ANN). Ці моделі складні, але здатні обробляти великі обсяги даних з численними змінними, наприклад, для передбачення відповіді на хіміотерапію на основі тисяч ознак. У найпростішому випадку (Hastie, Tibshirani, & Friedman, 2009; Mnih, Kavukcuoglu, & Silver, 2015) формула для одного шару нейронної мережі описує лінійне перетворення з подальшою активацією, що дозволяє виявляти приховані залежності навіть у складних біомедичних даних.

$$y = \sigma \left(\sum_{j=1}^n \omega_j x_j + b \right), \quad (2)$$

де x_j – вхідні ознаки; ω_j – вагові коефіцієнти; b – зсув; σ – активаційна функція; y – вихід передбачення.

Штучні нейронні мережі (ANN) встановлюють зв'язок між вхідними ознаками та прогнозованими результатами через багаторівневу систему математичних перетворень. Спочатку вхідні дані лінійно трансформуються у вузли прихованого шару, після чого обробляються за допомогою нелінійних функцій активації. Така архітектура дає змогу моделі виявляти складні залежності навіть у великомасштабних і неоднорідних даних.

Останніми роками нейромережеві моделі зазнали значного розвитку. Зокрема, було впроваджено рекурентні нейронні мережі (Recurrent Neural Networks, RNN), які ефективно працюють з послідовними даними, згорткові нейронні мережі (Convolutional Neural Networks, CNN), що спеціалізуються на обробці зображень, а також генеративні змагальні мережі (Generative Adversarial Networks, GAN), які демонструють здатність створювати нові зразки даних, серед яких зображення та тексти.

Популярність нейронних мереж значною мірою зумовлена їхньою здатністю ефективно працювати з неструктурованими типами даних і масштабуватися до високовимірних задач. Утім, їх складність є також їхнім недоліком: через непрозорий процес прийняття рішень такі моделі часто називають "чорними скриньками". Як і у випадку з ансамблевими підходами, обмежена інтерпретованість знижує їхню придатність до використання у клінічних умовах, де важливо не лише отримати точний прогноз, а й зрозуміти, на чому він базується (Bertsimas, & Wiberg, 2020).

3. Неконтрольоване навчання

Неконтрольоване навчання UL працює з даними, що не мають заздалегідь визначених міток або класів. Основне його завдання – виявити приховані структури, закономірності або групи в даних без використання цільової змінної чи зовнішнього контролю.

Такі алгоритми дають змогу ідентифікувати повторювані або домінуючі характеристики у вибірці, проте вони можуть бути чутливими до зовнішніх або змішувальних факторів, що не відображають справжню структуру даних (Greene et al., 2014).

На відміну від контрольованого підходу, що орієнтований на прогнозування конкретної змінної, неконтрольоване навчання більш зосереджене на вивченні внутрішньої організації даних. Це дає змогу виявити приховані патерни без необхідності в попередньому маркуванні прикладів, таких як виживаність чи реакція на лікування.

Однією з найпоширеніших і найбільш інтерпретованих технік UL у біомедичних застосуваннях є кластеризація. Зокрема, кластерний аналіз електронних медичних записів пацієнтів із певним захворюванням може допомогти у виявленні різних клінічних фенотипів або підтипів пацієнтів у межах однієї діагностичної групи.

Алгоритми кластеризації зазвичай спрямовані на поділ даних на групи таким чином, щоб максимізувати схожість між об'єктами в одному кластері та водночас мінімізувати схожість між кластерами. Оцінка схожості найчастіше здійснюється з використанням метрик відстані, таких як евклідова відстань, або на основі коефіцієнтів кореляції (Bertsimas, & Wiberg, 2020).

Одним із популярних методів кластеризації є алгоритм *K*-середніх (*K*-means). Його мета полягає в тому, щоб знайти такі центри кластерів, що мінімізують суму квадратів відстаней між кожною точкою та найближчим центром. Отже, алгоритм намагається згрупувати точки так, щоб вони були якнайближче до центрів своїх кластерів (Pratap, 2017):

$$J = \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mu_k\|^2, \quad (3)$$

де J – загальна сумарна квадратична помилка; K – кількість кластерів; C_k – множина точок, що належать до кластера k ; x_i – об'єкт даних; μ_k – середнє кластера k ; $\|x_i - \mu_k\|^2$ – евклідова відстань (в квадраті) між точкою та центром її кластера.

Ідентифікація молекулярних підтипів злоякісних пухлин є типовою задачею, яку ефективно розв'язують з використанням неконтрольованих методів аналізу. У випадках, коли дослідник має доступ до даних про експресію генів у

різних зразках пухлинної тканини, основною метою є виявлення характерних шаблонів експресії, що дають змогу об'єднати зразки у групи згідно з їхніми молекулярними властивостями.

Серед найпоширеніших алгоритмів кластеризації застосовують метод k-середніх та ієрархічну кластеризацію. Метод k-середніх базується на евристичних процедурах для поділу вибірки на наперед визначену кількість кластерів, тоді як ієрархічна кластеризація починається з індивідуального розгляду кожного елемента як окремого кластера, поступово об'єднуючи їх у більші групи відповідно до міжкластерної відстані. Результатом такого підходу є деревоподібна структура, що дає змогу варіювати рівень кластеризації залежно від обраного порогу агрегації (Bertsimas, & Wiberg, 2020).

Одним з головних викликів у використанні UL є інтерпретація отриманих кластерів, що особливо актуально в контексті первинного дослідницького аналізу біомедичних даних.

4. Навчання з підкріпленням

Ухвалення клінічних рішень часто є складним і нелінійним процесом, що залежить від багатьох факторів одночасно. Щоб упоратися із цими складнощами, у медицині зростає зацікавленість у методах навчання з підкріпленням (Reinforcement Learning, RL) – напрямку машинного навчання, де агент взаємодіє із середовищем, приймаючи рішення згідно з певною стратегією, метою якої є максимізація винагороди.

У межах RL агент формує політику – адаптивний план дій, спрямований на досягнення максимальної кумулятивної винагороди. Процес навчання базується на ітеративному вдосконаленні стратегії через експерименти й помилки, поки не буде знайдено або наближено оптимальне рішення (Lopez, 2024; Siddharth, 2022).

Кожна дія агента змінює стан середовища, що у медичному контексті може відповідати клінічному стану пацієнта або супутнім параметрам. Залежно від наслідків цієї дії агент отримує позитивну або негативну винагороду. Отже, він поступово формує політику, спираючись на отримані винагороди та правила переходу між станами та діями, що впливають на вибір подальших кроків.

Головною метою агента є не максимізація короткострокової вигоди, а оптимізація довгострокового результату шляхом обґрунтованого вибору дій для кожного клінічного стану. Такий підхід дає змогу ефективно моделювати динамічні процеси ухвалення рішень у медичній практиці (Van Buuren, 2018).

Для оцінки якості політики використовують функцію значення стану $V^\pi(s)$ (Lopez, 2024):

$$V^\pi(s) = \sum_{a_t \in A} \pi(s_t) \sum_{s_{t+1} \in S} T(s_t, a_t) [R(s_t, s_{t+1}) + \gamma V^\pi(s_{t+1})], \quad (4)$$

де $V^\pi(s)$ – функція цінності стану s при політиці π , вона показує очікувану сумарну нагороду, починаючи зі стану s і дотримуючись політики π ; $\pi(s_t)$ – ймовірність вибору дії a_t у стані s_t за політикою π ; $T(s_t, a_t)$ – функція переходу, ймовірність потрапити у стан s_{t+1} після виконання дії a_t у стані s_t ; $R(s_t, s_{t+1})$ – нагорода, отримана за перехід зі стану s_t у s_{t+1} ; γ – коефіцієнт дисконтування $0 < \gamma < 1$, що зменшує вагу майбутніх нагород.

Навчання з підкріпленням виявляється особливо ефективним для розв'язання складних завдань, таких як клінічне ухвалення рішень, адже RL-агенти можуть приймати дії навіть під час процесу навчання і поступово адаптувати свої стратегії для максимізації очікуваної винагороди. Хоча алгоритми з навчанням під наглядом також можуть працювати з неповними даними, використовуючи імпутацію, ці методи не завжди забезпечують точність і можуть призводити до помилок рішень (Van Buuren, 2018).

Через те, що винагороди в RL можуть бути рідкісними та відкладеними в часі, ці алгоритми здатні відстежувати причинно-наслідкові зв'язки між діями та їхніми наслідками, що є складним завданням для традиційних підходів (Siddharth, 2022). Завдяки можливості працювати з великими обсягами даних, включно з візуалізаціями, біомедичними показниками й індивідуальними характеристиками пацієнтів, RL допомагає оптимізувати стратегії планування та моніторингу лікування у режимі реального часу. Це особливо важливо в онкології, де терапія має бути адаптивною й персоналізованою, а реакція пацієнта на лікування є непередбачуваною. Крім того, RL враховує довгострокові наслідки клінічних рішень, навіть якщо взаємозв'язки між терапевтичними діями та результатами повністю не з'ясовані.

Варто зауважити, що RL не належить ні до методів навчання під наглядом, ні до методів навчання без нагляду. На відміну від перших, які навчаються на основі мічених даних, і других, які шукають приховані структури в немічених даних, RL базується на взаємодії агента із середовищем, де зворотний зв'язок подається через систему винагород і штрафів (Sutton, & Barto, 2018).

Методи RL поділяють на модельні (model-based) та безмодельні (model-free). У модельних підходах агент формує внутрішню модель динаміки середовища і використовує її для планування дій. Безмодельні алгоритми, навпаки, не мають доступу до інформації про переходи між станами чи винагороди й навчаються безпосередньо з досвіду за допомогою спроб і помилок (Eckardt et al., 2021).

Безмодельні методи поділяють на три основні категорії: засновані на цінностях (value-based RL), засновані на політиці (policy-based RL) й акторсько-критичні (actor-critic RL).

У value-based підходах метою є вивчення цінності станів або пар стан – дія для визначення оптимальної політики. Одним із найпоширеніших алгоритмів є Q-learning, що допомагає агенту знаходити найкращу стратегію, максимізуючи суму майбутніх винагород.

Policy-based методи працюють шляхом безпосередньої оптимізації параметрів політики з метою максимізації очікуваної винагороди, прикладом яких є алгоритм State – Action – Reward – State – Action (SARSA).

Акторсько-критичні методи поєднують підходи value-based і policy-based, усуваючи недоліки кожного з них. Градієнтні методи політики схильні застрягати в локальних максимумах і мають високу варіативність, тоді як value-based алгоритми іноді погано збігаються. Акторсько-критичні алгоритми використовують актора для вибору дій на основі градієнтів політики, а критика – для оцінки цих дій за допомогою методу навчання із часових відмінностей, що підвищує стабільність й ефективність у задачах із безперервним простором дій. Прикладами таких методів є A2C та A3C (Lopez, 2024).

5. Результати та висновки про перспективи використання машинного навчання в онкології

На сучасному етапі методи й алгоритми машинного навчання ML застосовують для обробки даних, розпізнавання, класифікації, прогнозування й ухвалення рішень у складних багатопараметричних задачах практичної онкології, де часто відсутні однозначні розв'язки. Через значущість і складність математичних підходів останнім часом сформувалася міждисциплінарна галузь – математична онкологія, що використовує сучасні математичні й обчислювальні інструменти для дослідження виникнення, розвитку та реакції раку на лікування. Математична онкологія поєднує знання з математики, інформатики та біології для створення моделей, що описують рак на різних рівнях – від молекулярного (зокрема генетичні маркери) до клітинного і тканинного (процеси росту пухлини і метастазування).

У сучасній онкології вибір методу ML визначається конкретною поставленою задачею. Переваги й недоліки різних алгоритмів, які застосовували в дослідженнях, узагальнені у табл. 1.

Таблиця 1

Порівняння використання методів машинного навчання в онкології

Метод навчання	Приклади алгоритмів	Недоліки	Переваги	Завдання, де метод найкраще використовується
Контрольоване навчання	Linear Models (Geleher, Cox, & Huang, 2014; Kong et al., 2020; Parker et al., 2009)	Потребують наявності мічених даних; не справляються з нелінійними залежностями; слабка продуктивність у нових мутаціях; погано працюють з висококорельованими ознаками	Висока зрозумілість й інтерпретованість результатів; простота і швидкість роботи	Прогноз виживаності, раннє виявлення захворювань; оцінка ризиків
	Support Vector Machine (Ammad-Ud-Din et al., 2017)	Важко масштабувати на великі обсяги даних	Стійкість до перенавчання; надійна робота з невеликими наборами даних	Прогноз рецидивів; класифікація пухлин
	Decision Tree Models (Malyutina et al., 2019)	Можливе перенавчання; чутливі до шуму	Ефективна робота з категоріальними та числовими даними; чітко інтерпретовані правила	Класифікація пацієнтів, прогноз реакції на препарат.
	Ensemble Models (Kim et al., 2020)	Довгий час навчання; складно інтерпретувати	Стійкість до перенавчання; висока точність прогнозів	Прогноз відповідей на лікування; класифікація типів пухлин.
	Neural Networks (Chang et al., 2018; Menden et al., 2013; Sakellaropoulos et al., 2019)	Складні в налаштуванні; відомі як "чорна скринька"; потребують великих обсягів даних	Ефективні для обробки зображень; можливість моделювати складні нелінійні взаємозв'язки; добре масштабуються.	Виявлення пухлин на сканах; класифікація гісто зображень.
Неконтрольоване навчання	Autoencoders (Zhang et al., 2021)	Важко інтерпретувати; потребують великої кількості даних	Зменшення розмірності даних; здатність виявляти аномалії	Виявлення рідкісних мутацій; зменшення розмірності даних
	K-means (Tothill et al., 2008)	Складно інтерпретувати результати; чутливий до вибору k; не завжди клінічно релевантний; не гарантує оптимальне розбиття	Допомагає виявляти нові закономірності; швидка обробка великих даних; проста реалізація	Стратифікація пацієнтів; виявлення нових підтипів раку
Навчання з підкріпленням	Q-learning (Zade, Haghighi, & Soltani, 2020; Li et al., 2024; Shen et al., 2020)	Складно валідовувати й інтерпретувати; вимагає правильної підгонки гіперпараметрів; можливі нестабільності	Підходить для складних середовищ; швидкий пошук оптимальної політики	Вибір послідовності лікування; оптимізація дозування
	SARSA (Alsaadi et al., 2023)	Висока обчислювальна складність; вразливий до перенавчання; складно інтерпретувати; залежить від поточної політики	Оптимізація схем лікування; стабільні результати; урахування реальної поведінки агента	Інтерактивні рекомендації для персоналізованої терапії; адаптивні протоколи лікування; оптимізація дозування
	Reinforce (Wang et al., 2020)	Складна валідація й інтерпретація; повільна збіжність; висока дисперсія градієнтів.	Ефективна робота з неперервними просторами дій; легкість реалізації	Навчання тактик терапії в контексті часу і відповіді
	AC (Actor-Critic) (Liu, Shen, & Pan, 2022; Xu et al., 2023)	Складна валідація й інтерпретація; чутливий до налаштувань гіперпараметрів; складність реалізації	Зменшена дисперсія оцінок; більш стабільна і швидша, ніж Reinforce	Динамічне управління лікувальним протоколом у реальному часі

Методи контрольованого (SL) та неконтрольованого (UL) навчання, описані в попередньому розділі, широко використовують у медичних дослідженнях, особливо в роботі з ретроспективними даними. Їх ефективно застосовують для

виконання таких завдань, як діагностика, стратифікація ризику, класифікація генетичних підтипів, прогнозування результатів терапії, моніторинг рецидивів і передбачення реакції пацієнтів на лікування (Yu, Beam, & Kohane, 2018).

Однак клінічне середовище змінюється динамічно: зазнають змін як фізіологічні параметри пацієнта, так і зовнішні фактори. Більшість моделей SL та UL не враховують ці часові зміни, що обмежує їхню ефективність у практичних умовах.

У цьому контексті навчання з підкріпленням (RL) постає як перспективна альтернатива, оскільки воно дає змогу ухвалювати послідовні рішення в часі з урахуванням змін середовища. RL-моделі здатні адаптувати свої дії до поточного стану пацієнта, орієнтуючись на максимізацію довгострокових клінічних результатів (Eckardt et al., 2021).

Цей підхід повністю відповідає ідеології прецизійної медицини, що прагне до індивідуалізованого лікування. RL дає змогу підібрати найефективнішу терапію та дозування з урахуванням особистих біологічних параметрів і перебігу захворювання. Це підвищує ефективність лікування, зменшує ймовірність побічних ефектів і поліпшує якість життя пацієнтів (Rafique, Islam, & Kazi, 2021). RL допомагає гнучко адаптувати лікувальні стратегії до конкретних пацієнтів або підгруп, що можуть отримати найбільшу користь від інтервенцій (Eckardt et al., 2021).

Порівняно з іншими підходами, RL-алгоритми демонструють кращі результати в оптимізації та динамічному управлінні лікувальними процесами. Наприклад, у дослідженні (Liu, Shen, & Pan, 2022), що застосовувало алгоритм Actor-Critic для вивчення зв'язку між генетичними характеристиками та відповіддю на препарати при лікуванні раку молочної залози, RL-підхід перевершив п'ять інших методів: ElasticNet (EN), Kernel Ridge Regression (KRR), Similarity-Regularized Matrix Factorization (SRMF), рекомендаційна система CaDRRes, та Kernelized Ranking Learning (KRL).

В іншому дослідженні (Wang et al., 2020), Reinforce-алгоритм показав високу ефективність у диференціації доброякісних і злоякісних вузлів молочної залози на основі ультразвукових зображень. Його результати перевершили ті, що були отримані від одномодальних моделей і сучасних мультимодальних підходів, таких як Share-Net, Centralnet та Voting Schemes. У роботі (Shen et al., 2020) алгоритм Q-learning був використаний для планування лікування променевою терапією з модульованою інтенсивністю. Його результати були збіжними з рішеннями досвідченого лікаря, який вручну налаштовував параметри терапії, такі як обсяг дози, для створення оптимального плану лікування для кожного пацієнта.

Дослідження графіків дозування ліків (Alsaadi et al., 2023), оптимальних режимів хіміотерапії (Zade, Haghighi, & Soltani, 2020), планування лікування променевою терапією (Shen et al., 2020), скринінгу раку печінки (Xu, Song, & Zhang, 2023) також підтвердили високу ефективність RL-підходів у медичному контексті – вони забезпечують кращу або принаймні рівноцінну продуктивність порівняно з традиційними алгоритмами машинного навчання.

Дискусія і висновки

У роботі представлено огляд сучасних підходів до задач розпізнавання, класифікації, прогнозування та ухвалення обґрунтованих рішень щодо вибору хіміотерапевтичного лікування й мінімізації ризику виникнення побічних реакцій (ADR) у пацієнтів з онкологічними захворюваннями. Поряд із класичними методами математичної статистики, у математичній онкології активно застосовують методи глибокого навчання. Особливо ефективними виявилися алгоритми випадкових лісів (Random Forest), опорних векторних машин (SVM) і штучних нейронних мереж (ANN), що здатні моделювати складні взаємозв'язки між великою кількістю різноманітних змінних – від генетичних маркерів і характеристик лікарських засобів до тривимірних зображень пухлин за даними комп'ютерної томографії.

Найбільшу увагу приділено методам контролюваного навчання (SL), неконтрольованого навчання (UL) та навчання з підкріпленням (RL). У табл. 1 узагальнено ключові переваги та недоліки кожного підходу, а також сфери їхнього оптимального застосування. Зроблено висновок, що RL є найперспективнішим напрямом, оскільки дає змогу створювати інтелектуальні системи підтримки клінічних рішень, що здатні в режимі реального часу адаптувати терапевтичні стратегії відповідно до змін у стані пацієнта, реакції організму на лікування та появи можливих побічних ефектів під час використання певних комбінацій препаратів.

Отримані висновки стануть основою для подальших етапів дослідження.

Внесок авторів: Іван Тюрдьо – концептуалізація; формальний аналіз; аналіз джерел, підготовка огляду літератури; написання (оригінальна чернетка), Наталія Кізілова – концептуалізація; формальний аналіз; аналіз джерел, підготовка огляду літератури; написання (перегляд і редагування).

Джерела фінансування. Дослідження не отримало жодного гранту від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Alsaadi, F. E., Yasami, A., Volos, C., Bekiros, S., & Jahanshahi, H. (2023). A new fuzzy reinforcement learning method for effective chemotherapy. *Mathematics*, 11(2), 477. <https://doi.org/10.3390/math11020477>
- Ammad-Ud-Din, M., Khan, S. A., Wennerberg, K., & Aittokallio, T. (2017). Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. *Bioinformatics*, 33(14), 359–368.
- Badwan, B. A., Liapopoulos, G., Kyrodimos, E., Skaltsas, D., Tsigirgos, A., & Gorgoulis, V. G. (2023). Machine learning approaches to predict drug efficacy and toxicity in oncology. *Reports Methods*, 3(2), 100413. <https://doi.org/10.1016/j.crmeth.2023.100413>
- Bertsimas, D., & Wiberg, H. (2020). Machine learning in oncology: Methods, applications, and challenges. *JCO Clinical Cancer Informatics*, 4, 885–894. <https://doi.org/10.1200/CCI.20.00072>
- Bhandari, A., Gu, B., Kashkooli, F. M., & Zhan, W. (2024). Image-based predictive modelling frameworks for personalised drug delivery in cancer therapy. *Journal of Controlled Release*, 370, 721–746. <https://doi.org/10.1016/j.jconrel.2024.05.004>
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74, 229–263. <https://doi.org/10.3322/caac.21834>
- Chang, Y., Park, H., Yang, H. J., Lee, S., Lee, K. Y., & Kim, T. S. (2018). Cancer Drug Response Profile scan (CDRscan): A deep learning model that predicts drug effectiveness from cancer genomic signature. *Scientific Reports*, 8(1), 8857. <https://doi.org/10.1038/s41598-018-27214-6>
- Cortes, C., & Vapnik, V. (2009). Support-vector networks. *Chemical Biology & Drug Design*, 297(3), 273–297. <https://doi.org/10.1007/BF00994018>
- De los Rios de la Serna, C. D., Boers-Doets, C. B., Wiseman, T., Radia, B., & Hammond, R. (2024). Early recognition and management of side effects related to systemic anticancer therapy for advanced breast cancer. *Seminars in Oncology Nursing*, 40, 151553. <https://doi.org/10.1016/j.soncn.2023.151553>

- Eckardt, J.-N., Wendt, K., Bornhäuser, M., & Middeke, J. M. (2021). Reinforcement learning for precision oncology. *Cancers*, 13(18), 4624. <https://doi.org/10.3390/cancers13184624>
- Eguale, T., Buckeridge, D. L., Verma, A., Winslade, N. E., Benedetti, A., & Hanley, J. A. (2016). Association of off-label drug use and adverse drug events in an adult population. *JAMA Internal Medicine*, 176(1), 55–63. <https://doi.org/10.1001/jamainternmed.2015.6058>
- Fan, K., Cheng, L., & Li, L. (2021). Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. *Briefings in Bioinformatics*, 22(6), 1–12. <https://doi.org/10.1093/bib/bbab271>
- Firoozbakht, F., Yousefi, B., & Schwikowski, B. (2021). An overview of machine learning methods for monotherapy drug response prediction. *Briefings in Bioinformatics*, 23(1), bbab408. <https://doi.org/10.1093/bib/bbab408>
- Geeleher, P., Cox, N. J., & Huang, R. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biology*, 15(3), R47. <https://doi.org/10.1186/gb-2014-15-3-r47>
- Greene, C. S., Tan, J., Ung, M., Moore, J. H., & Cheng, C. (2014). Big data bioinformatics. *Journal of Cellular Physiology*, 229(12), 1896–1900. <https://doi.org/10.1002/jcp.24662>
- Guan, S., & Wang, G. (2024). Drug discovery and development in the era of artificial intelligence: From machine learning to large language models. *Artificial Intelligence Chemistry*, 2, 100070. <https://doi.org/10.1016/j.aichem.2024.100070>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
- Kim, Y., Kim, D., Cao, B., Carvajal, R., & Kim, M. (2020). PDXGEM: Patient-derived tumor xenograft-based gene expression model for predicting clinical response to anticancer therapy in cancer patients. *BMC Bioinformatics*, 21(1), 288. <https://doi.org/10.1186/s12859-020-03633-z>
- Kong, J. H., Lee, H., Kim, D., Han, S. K., Ha, D., & Shin, K. (2020). Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. *Nature Communications*, 11(1), 5485. <https://doi.org/10.1038/s41467-020-19313-8>
- Kurilov, R., Haibe-Kains, B., & Brors, B. (2020). Assessment of modelling strategies for drug response prediction in cell lines and xenografts. *Scientific Reports*, 10, 2849. <https://doi.org/10.1038/s41598-020-59656-2>
- Li, C., Guo, Y., Lin, X., Feng, X., Xu, D., & Yang, R. (2024). Deep reinforcement learning in radiation therapy planning optimization: A comprehensive review. *Physica Medica*, 125, 104498. <https://doi.org/10.1016/j.ejmp.2024.104498>
- Li, Y., Umbach, D. M., & Krahn, J. M. (2021). Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genomics*, 22, 272. <https://doi.org/10.1186/s12864-021-07581-7>
- Liu, J., Li, M., & Chen, X. (2022). AntiMF: A deep learning framework for predicting anticancer peptides based on multi-view feature extraction. *Methods*, 207, 38–43. <https://doi.org/10.1016/j.ymeth.2022.07.017>
- Liu, M., Shen, X., Pan, W. (2022). Deep reinforcement learning for personalized treatment recommendation. *Statistics in Medicine*, 41(20), 4034–4056. <https://doi.org/10.1002/sim.9491>
- Lopez, R. G. (2024). Reinforcement learning in oncology: A comprehensive review. *Utrecht University*, 36073 <https://studenttheses.uu.nl/handle/20.500.12932/47144>
- Malyutina, A., Majumder, M. M., Wang, W., Pessia, A., Heckman, C. A., & Tang, J. (2019). Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLoS Computational Biology*, 15(5), e1006752. <https://doi.org/10.1371/journal.pcbi.1006752>
- Menden, M. P., Iorio, F., Garnett, M., McDermott, U., Benes, C. H., Ballester, P. J., & Saez-Rodriguez, J. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One*, 8(4), e61318. <https://doi.org/10.1371/journal.pone.0061318>
- Mnih, V., Kavukcuoglu, K., & Silver, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Nilashi, M., Ahmadi, H., Abumalloh, R. A., Alrizq, M., Alghamdi, A., & Alyami, S. (2024). Knowledge discovery of patients reviews on breast cancer drugs: Segmentation of side effects using machine learning techniques. *Heliyon*, 10(19), e38563. <https://doi.org/10.1016/j.heliyon.2024.e38563>
- Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., & Vickery, T. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, 27(8), 1160–1167. <https://doi.org/10.1200/JCO.2008.18.1370>
- Partin, A., Brettin, T. S., Zhu, Y., Narykov, O., Clyde, A., Overbeek, J., & Stevens, R. L. (2023). Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. *Frontiers in Medicine*, 10, 1086097. <https://doi.org/10.3389/fmed.2023.1086097>
- Phan, L. T., Park, H. W., Pitti, T., Madhavan, T., & Jeon, Y.-J. (2022). MLACP 2.0: An updated machine learning tool for anticancer peptide prediction. *Computational and Structural Biotechnology Journal*, 20, 4473–4480. <https://doi.org/10.1016/j.csbj.2022.07.043>
- Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: An overview and their use in medicine. *Journal of Medical Systems*, 26, 445–463. <https://doi.org/10.1023/a:1016409317640>
- Pratap, D. (2017). *Statistics for machine learning: Build supervised, unsupervised, and reinforcement learning models using both Python and R*. Packt Publishing.
- Rafique, R., Islam, S. M. R., & Kazi, J. U. (2021). Machine learning in the prediction of cancer therapy. *Computational and Structural Biotechnology Journal*, 19, 4003–4017. <https://doi.org/10.1016/j.csbj.2021.07.003>
- Riechelmann, R. P., Zimmermann, C., Chin, S. N., Wang, L., O'Carroll, A., & Zarinehbab, S. (2008). Potential drug interactions in cancer patients receiving supportive care exclusively. *Journal of Pain and Symptom Management*, 35(5), 535–543. <https://doi.org/10.1016/j.jpainsymman.2007.06.009>
- Rockne, R. C., & Scott, J. G. (2019). Introduction to Mathematical Oncology. *JCO Clinical Cancer Informatics*, 3, 1–4. <https://doi.org/10.1200/CCI.19.00010>
- Sakellariopoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., & Polyzos, A. (2019). A deep learning framework for predicting response to therapy in cancer. *Cell Reports*, 29(11), 3367–3373. <https://doi.org/10.1016/j.celrep.2019.11.017>
- Shan, W., Shen, C., Luo, L., & Ding, P. (2023). Multi-task learning for predicting synergistic drug combinations based on auto-encoding multirelational graphs. *iScience*, 26, 108020. <https://doi.org/10.1016/j.isci.2023.108020>
- She, S., Chen, H., Ji, W., Sun, M., Cheng, J., Rui, M., & Feng, C. (2022). Deep learning based multi-drug synergy prediction model for individually tailored anticancer therapies. *Frontiers in Pharmacology*, 13, 1032875. <https://doi.org/10.3389/fphar.2022.1032875>
- Shen, C., Nguyen, D., Chen, L., Gonzalez, Y., McBeth, R., Qin, N., Jiang, S. B., & Jia, X. (2020). Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning. *Medical Physics*, 47(6), 2329–2336. <https://doi.org/10.1002/mp.14114>
- Siddharth, N., Korot, E., Fu, D. J., Zhang, G., Mishra, K., & Lee, A. Y. (2022). Reinforcement learning in ophthalmology: Potential applications and challenges to implementation. *The Lancet Digital Health*, 4, 692–697. [https://doi.org/10.1016/S2589-7500\(22\)00128-5](https://doi.org/10.1016/S2589-7500(22)00128-5)
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Teplýtska, O., Ernst, M., Koltermann, L. M., Valderama, D., Trunz, E., Vaisband, M., Hasenauer, J., Fröhlich, H., & Jaehde, H. (2024). Learning methods for precision dosing in anticancer drug therapy: A scoping review. *Clinical Pharmacokinetics*, 63, 1221–1237. <https://doi.org/10.1007/s40262-024-01409-9>

- Tibshirani, R. (2011). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273–282. <https://doi.org/10.2307/41262671>
- Timilsina, M., Tandan, M., & Nováček, V. (2022). Machine learning approaches for predicting the onset time of the adverse drug events in oncology. *Machine Learning with Applications*, 9, 100367. <https://doi.org/10.1016/j.mlwa.2022.100367>
- Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., & Bowtell, D. D. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical Cancer Research*, 14, 5198–5208. <https://doi.org/10.1158/1078-0432.CCR-08-0196>
- Triguero, I., García, S., & Herrera, F. (2015). Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study. *Knowledge and Information Systems*, 42(2), 245–284. <https://doi.org/10.1007/s10115-013-0706-y>
- Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC Press.
- Wang, J., Miao, J., Yang, X., Li, R., Zhou, G., Huang, Y., Lin, Z., Xue, W., Jia, X., Zhou, J., Huang, R., Ni, D. (2020). *Auto-weighting for breast cancer classification in multimodal ultrasound*. ArXiv. <https://doi.org/10.48550/arXiv.2008.03435>
- Xu, C., Song, Y., Zhang, D., Bittencourt, L. K., Tirumani, S. H., & Li, S. (2023). Spatiotemporal knowledge teacher–student reinforcement learning to detect liver tumors without contrast agents. *Medical Image Analysis*, 90, 102980. <https://doi.org/10.1016/j.media.2023.102980>
- Yang, S., & Kar, S. (2023). Application of artificial intelligence and machine learning in early detection of adverse drug reactions (ADRs) and drug-induced toxicity. *Artificial Intelligence in Chemistry*, 2, 100011. <https://doi.org/10.1016/j.aichem.2023.100011>
- Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719–731. <https://doi.org/10.1038/s41551-018-0305-z>
- Zade, A.E., Haghighi, S. S., & Soltani, M. (2020). Reinforcement learning for optimal scheduling of glioblastoma treatment with temozolomide. *Computer Methods and Programs in Biomedicine*, 193, 105443. <https://doi.org/10.1016/j.cmpb.2020.105443>
- Zhang, T., Zhang, L., Payne, P. R. O., & Li, F. (2021). Synergistic drug combination prediction by integrating multiomics data in deep learning models. *In Methods in Molecular Biology*, 2194, 223–238. https://doi.org/10.1007/978-1-0716-0849-4_12
- Zhou, J.-B., Tang, D., He, L., Lin, S., Lei, J. H., Sun, H., Xu, X., & Deng, C.-X. (2023). Machine learning model for anti-cancer drug combinations: Analysis, prediction, and validation. *Pharmacological Research*, 194, 106830. <https://doi.org/10.1016/j.phrs.2023.106830>

References

- Alsaadi, F. E., Yasami, A., Volos, C., Bekiros, S., & Jahanshahi, H. (2023). A new fuzzy reinforcement learning method for effective chemotherapy. *Mathematics*, 11(2), 477. <https://doi.org/10.3390/math11020477>
- Ammad-Ud-Din, M., Khan, S. A., Wennerberg, K., & Aittokallio, T. (2017). Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. *Bioinformatics*, 33(14), 359–368.
- Badwan, B. A., Liaropoulos, G., Kyrodimos, E., Skaltsas, D., Tsirogos, A., & Gorgoulis, V. G. (2023). Machine learning approaches to predict drug efficacy and toxicity in oncology. *Reports Methods*, 3(2), 100413. <https://doi.org/10.1016/j.crmeth.2023.100413>
- Bertsimas, D., & Wiberg, H. (2020). Machine learning in oncology: Methods, applications, and challenges. *JCO Clinical Cancer Informatics*, 4, 885–894. <https://doi.org/10.1200/CCI.20.00072>
- Bhandari, A., Gu, B., Kashkooli, F. M., & Zhan, W. (2024). Image-based predictive modelling frameworks for personalised drug delivery in cancer therapy. *Journal of Controlled Release*, 370, 721–746. <https://doi.org/10.1016/j.jconrel.2024.05.004>
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74, 229–263. <https://doi.org/10.3322/caac.21834>
- Chang, Y., Park, H., Yang, H. J., Lee, S., Lee, K. Y., & Kim, T. S. (2018). Cancer Drug Response Profile scan (CDRscan): A deep learning model that predicts drug effectiveness from cancer genomic signature. *Scientific Reports*, 8(1), 8857. <https://doi.org/10.1038/s41598-018-27214-6>
- Cortes, C., & Vapnik, V. (2009). Support-vector networks. *Chemical Biology & Drug Design*, 297(3), 273–297. <https://doi.org/10.1007/BF00994018>
- De los Rios de la Serna, C. D., Boers-Doets, C. B., Wiseman, T., Radia, B., & Hammond, R. (2024). Early recognition and management of side effects related to systemic anticancer therapy for advanced breast cancer. *Seminars in Oncology Nursing*, 40, 151553. <https://doi.org/10.1016/j.soncn.2023.151553>
- Eckardt, J.-N., Wendt, K., Bornhäuser, M., & Middeke, J. M. (2021). Reinforcement learning for precision oncology. *Cancers*, 13(18), 4624. <https://doi.org/10.3390/cancers13184624>
- Egualde, T., Buckeridge, D. L., Verma, A., Winslade, N. E., Benedetti, A., & Hanley, J. A. (2016). Association of off-label drug use and adverse drug events in an adult population. *JAMA Internal Medicine*, 176(1), 55–63. <https://doi.org/10.1001/jamainternmed.2015.6058>
- Fan, K., Cheng, L., & Li, L. (2021). Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. *Briefings in Bioinformatics*, 22(6), 1–12. <https://doi.org/10.1093/bib/bbab271>
- Firoozbakht, F., Yousefi, B., & Schwikowski, B. (2021). An overview of machine learning methods for monotherapy drug response prediction. *Briefings in Bioinformatics*, 23(1), bbab408. <https://doi.org/10.1093/bib/bbab408>
- Geeleher, P., Cox, N. J., & Huang, R. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biology*, 15(3), R47. <https://doi.org/10.1186/gb-2014-15-3-r47>
- Greene, C. S., Tan, J., Ung, M., Moore, J. H., & Cheng, C. (2014). Big data bioinformatics. *Journal of Cellular Physiology*, 229(12), 1896–1900. <https://doi.org/10.1002/jcp.24662>
- Guan, S., & Wang, G. (2024). Drug discovery and development in the era of artificial intelligence: From machine learning to large language models. *Artificial Intelligence Chemistry*, 2, 100070. <https://doi.org/10.1016/j.aichem.2024.100070>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
- Kim, Y., Kim, D., Cao, B., Carvajal, R., & Kim, M. (2020). PDXGEM: Patient-derived tumor xenograft-based gene expression model for predicting clinical response to anticancer therapy in cancer patients. *BMC Bioinformatics*, 21(1), 288. <https://doi.org/10.1186/s12859-020-03633-z>
- Kong, J. H., Lee, H., Kim, D., Han, S. K., Ha, D., & Shin, K. (2020). Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. *Nature Communications*, 11(1), 5485. <https://doi.org/10.1038/s41467-020-19313-8>
- Kurilov, R., Haibe-Kains, B., & Brors, B. (2020). Assessment of modelling strategies for drug response prediction in cell lines and xenografts. *Scientific Reports*, 10, 2849. <https://doi.org/10.1038/s41598-020-59656-2>
- Li, C., Guo, Y., Lin, X., Feng, X., Xu, D., & Yang, R. (2024). Deep reinforcement learning in radiation therapy planning optimization: A comprehensive review. *Physica Medica*, 125, 104498. <https://doi.org/10.1016/j.ejmp.2024.104498>
- Li, Y., Umbach, D. M., & Krahn, J. M. (2021). Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genomics*, 22, 272. <https://doi.org/10.1186/s12864-021-07581-7>
- Liu, J., Li, M., & Chen, X. (2022). AntiMF: A deep learning framework for predicting anticancer peptides based on multi-view feature extraction. *Methods*, 207, 38–43. <https://doi.org/10.1016/j.ymeth.2022.07.017>
- Liu, M., Shen, X., Pan, W. (2022). Deep reinforcement learning for personalized treatment recommendation. *Statistics in Medicine*, 41(20), 4034–4056. <https://doi.org/10.1002/sim.9491>

- Lopez, R. G. (2024). Reinforcement learning in oncology: A comprehensive review. *Utrecht University*, 36073 <https://studenttheses.uu.nl/handle/20.500.12932/47144>
- Maljutina, A., Majumder, M. M., Wang, W., Pessia, A., Heckman, C. A., & Tang, J. (2019). Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLoS Computational Biology*, 15(5), e1006752. <https://doi.org/10.1371/journal.pcbi.1006752>
- Menden, M. P., Iorio, F., Garnett, M., McDermott, U., Benes, C. H., Ballester, P. J., & Saez-Rodriguez, J. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One*, 8(4), e61318. <https://doi.org/10.1371/journal.pone.0061318>
- Mnih, V., Kavukcuoglu, K., & Silver, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Nilashi, M., Ahmadi, H., Abumalloh, R. A., Alrizq, M., Alghamdi, A., & Alyami, S. (2024). Knowledge discovery of patients reviews on breast cancer drugs: Segmentation of side effects using machine learning techniques. *Heliyon*, 10(19), e38563. <https://doi.org/10.1016/j.heliyon.2024.e38563>
- Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., & Vickery, T. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, 27(8), 1160–1167. <https://doi.org/10.1200/JCO.2008.18.1370>
- Partin, A., Brettin, T. S., Zhu, Y., Narykov, O., Clyde, A., Overbeek, J., & Stevens, R. L. (2023). Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. *Frontiers in Medicine*, 10, 1086097. <https://doi.org/10.3389/fmed.2023.1086097>
- Phan, L. T., Park, H. W., Pitti, T., Madhavan, T., & Jeon, Y.-J. (2022). MLACP 2.0: An updated machine learning tool for anticancer peptide prediction. *Computational and Structural Biotechnology Journal*, 20, 4473–4480. <https://doi.org/10.1016/j.csbj.2022.07.043>
- Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: An overview and their use in medicine. *Journal of Medical Systems*, 26, 445–463. <https://doi.org/10.1023/a:1016409317640>
- Pratap, D. (2017). *Statistics for machine learning: Build supervised, unsupervised, and reinforcement learning models using both Python and R*. Packt Publishing.
- Rafique, R., Islam, S. M. R., & Kazi, J. U. (2021). Machine learning in the prediction of cancer therapy. *Computational and Structural Biotechnology Journal*, 19, 4003–4017. <https://doi.org/10.1016/j.csbj.2021.07.003>
- Riechelmann, R. P., Zimmermann, C., Chin, S. N., Wang, L., O'Carroll, A., & Zarinehbab, S. (2008). Potential drug interactions in cancer patients receiving supportive care exclusively. *Journal of Pain and Symptom Management*, 35(5), 535–543. <https://doi.org/10.1016/j.jpainsymman.2007.06.009>
- Rockne, R. C., & Scott, J. G. (2019). Introduction to Mathematical Oncology. *JCO Clinical Cancer Informatics*, 3, 1–4. <https://doi.org/10.1200/CCI.19.00010>
- Sakellariopoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., & Polyzos, A. (2019). A deep learning framework for predicting response to therapy in cancer. *Cell Reports*, 29(11), 3367–3373. <https://doi.org/10.1016/j.celrep.2019.11.017>
- Shan, W., Shen, C., Luo, L., & Ding, P. (2023). Multi-task learning for predicting synergistic drug combinations based on auto-encoding multirelational graphs. *iScience*, 26, 108020. <https://doi.org/10.1016/j.isci.2023.108020>
- She, S., Chen, H., Ji, W., Sun, M., Cheng, J., Rui, M., & Feng, C. (2022). Deep learning based multi-drug synergy prediction model for individually tailored anticancer therapies. *Frontiers in Pharmacology*, 13, 1032875. <https://doi.org/10.3389/fphar.2022.1032875>
- Shen, C., Nguyen, D., Chen, L., Gonzalez, Y., McBeth, R., Qin, N., Jiang, S. B., & Jia, X. (2020). Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning. *Medical Physics*, 47(6), 2329–2336. <https://doi.org/10.1002/mp.14114>
- Siddharth, N., Korot, E., Fu, D. J., Zhang, G., Mishra, K., & Lee, A. Y. (2022). Reinforcement learning in ophthalmology: Potential applications and challenges to implementation. *The Lancet Digital Health*, 4, 692–697. [https://doi.org/10.1016/S2589-7500\(22\)00128-5](https://doi.org/10.1016/S2589-7500(22)00128-5)
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Teplýska, O., Ernst, M., Koltermann, L. M., Valderrama, D., Trunz, E., Vaisband, M., Hasenauer, J., Fröhlich, H., & Jaehde, H. (2024). Learning methods for precision dosing in anticancer drug therapy: A scoping review. *Clinical Pharmacokinetics*, 63, 1221–1237. <https://doi.org/10.1007/s40262-024-01409-9>
- Tibshirani, R. (2011). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273–282. <https://doi.org/10.2307/41262671>
- Timilsina, M., Tandan, M., & Nováček, V. (2022). Machine learning approaches for predicting the onset time of the adverse drug events in oncology. *Machine Learning with Applications*, 9, 100367. <https://doi.org/10.1016/j.mlwa.2022.100367>
- Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., & Bowtell, D. D. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical Cancer Research*, 14, 5198–5208. <https://doi.org/10.1158/1078-0432.CCR-08-0196>
- Triguero, I., García, S., & Herrera, F. (2015). Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study. *Knowledge and Information Systems*, 42(2), 245–284. <https://doi.org/10.1007/s10115-013-0706-y>
- Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC Press.
- Wang, J., Miao, J., Yang, X., Li, R., Zhou, G., Huang, Y., Lin, Z., Xue, W., Jia, X., Zhou, J., Huang, R., Ni, D. (2020). Auto-weighting for breast cancer classification in multimodal ultrasound. *ArXiv*. <https://doi.org/10.48550/arXiv.2008.03435>
- Xu, C., Song, Y., Zhang, D., Bittencourt, L. K., Tirumani, S. H., & Li, S. (2023). Spatiotemporal knowledge teacher–student reinforcement learning to detect liver tumors without contrast agents. *Medical Image Analysis*, 90, 102980. <https://doi.org/10.1016/j.media.2023.102980>
- Yang, S., & Kar, S. (2023). Application of artificial intelligence and machine learning in early detection of adverse drug reactions (ADRs) and drug-induced toxicity. *Artificial Intelligence in Chemistry*, 2, 100011. <https://doi.org/10.1016/j.aichem.2023.100011>
- Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719–731. <https://doi.org/10.1038/s41551-018-0305-z>
- Zade, A.E., Haghighi, S. S., & Soltani, M. (2020). Reinforcement learning for optimal scheduling of glioblastoma treatment with temozolomide. *Computer Methods and Programs in Biomedicine*, 193, 105443. <https://doi.org/10.1016/j.cmpb.2020.105443>
- Zhang, T., Zhang, L., Payne, P. R. O., & Li, F. (2021). Synergistic drug combination prediction by integrating multiomics data in deep learning models. *In Methods in Molecular Biology*, 2194, 223–238. https://doi.org/10.1007/978-1-0716-0849-4_12
- Zhou, J.-B., Tang, D., He, L., Lin, S., Lei, J. H., Sun, H., Xu, X., & Deng, C.-X. (2023). Machine learning model for anti-cancer drug combinations: Analysis, prediction, and validation. *Pharmacological Research*, 194, 106830. <https://doi.org/10.1016/j.phrs.2023.106830>

Отримано редакцією журналу / Received: 12.07.25
 Прорецензовано / Revised: 25.08.25
 Схвалено до друку / Accepted: 10.10.25

Ivan TIURDO, PhD Student
ORCID ID: 0009-0001-7315-3628
e-mail: ivan.turdo@karazin.ua
V. N. Karazin Kharkiv National University, Kharkiv, Ukraine

Natalya KIZILOVA, DSc (Phys. & Math.), Prof.
ORCID ID: 0000-0001-9981-7616
e-mail: kizilova@karazin.ua
V. N. Karazin Kharkiv National University, Kharkiv, Ukraine

THE APPLICATION OF MACHINE LEARNING METHODS IN MODERN CANCER THERAPY

The purpose of this study is to analyze modern approaches to assessing the efficacy and safety of drugs used in anti-cancer therapy using machine learning methods. Particular attention is paid to the prospects of implementing such methods in mathematical oncology, a field that actively uses mathematical modeling and computer simulations in oncological research.

As part of the study, we searched for and studied relevant scientific sources on the application of machine learning in oncology. As a result, a systematic review of the literature covering the use of machine learning in this area was conducted.

The main machine learning approaches, such as Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL), are analyzed in the context of modern oncology. Specific examples of the use of various machine learning algorithms in research related to cancer treatment, as well as more general oncological problems, are considered. The advantages and limitations of these approaches are assessed depending on the goals set, for example, in such tasks as predicting treatment response, biomarker identification, and automated medical image analysis.

The study found that machine learning is already being actively implemented in oncology research and demonstrates high efficiency in solving various problems, in particular, identifying hidden patterns that are not available in traditional analysis. At the same time, it is noted that there are numerous other areas where the use of machine learning can significantly enhance scientific research and clinical practice. In particular, the use of reinforcement learning algorithms in the field of personalized precision medicine, which plays a key role in creating individualized approaches to cancer treatment, looks promising.

Key words: *mathematical modeling, supervised learning, unsupervised learning, reinforcement learning, precision medicine, personalized therapy.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

СУЧАСНА ФІЗИКА

UDC 535.4

DOI: <https://doi.org/10.17721/1812-5409.2025/2.35>

Andrii YAKUNOV, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0002-2664-4803

e-mail: yakunov@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Volodymyr DIBROV, PhD Student

ORCID ID: 0009-0005-8346-9254

e-mail: dibrov.volodymyr.9@gmail.com

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Artem SACHKO, Student

e-mail: soovat2@gmail.com

Ukrainian Physical and Mathematical Lyceum, Kyiv, Ukraine

**CORRELATION BETWEEN THE FRACTAL DIMENSION
OF SURFACE TEXTURE AND ITS LASER SPECKLES AS THE BASIS
FOR A NON-CONTACT METHOD FOR ANALYZING CERAMIC MATERIALS**

A quantitative relationship between the fractal dimension of the ceramic surface texture obtained by optical microscopy and the fractal dimension of the corresponding laser speckle field was studied. A low sensitivity of speckle analysis for defocus has been identified, which simplifies the control procedure. Analysis of homogeneous samples revealed statistically significant differences in the fractal characteristics of the images of the two types of ceramics, confirming the possibility of using them to identify materials and detect defects. In the case of a layered ceramic sample, fractal analysis of speckle paintings clearly reflected its structure, indicating the potential of the method for non-destructive layering and homogeneity testing. The demonstrated robustness of the approach highlights its metrological reliability and opens up prospects for its integration into standardized schemes of quality control, ensuring consistent and reproducible assessment of structural integrity in ceramic manufacturing. Prospects for further research are associated with the development of simplified optical speckle analysis schemes for the creation of portable systems for express quality control of ceramic products.

Key words: optical microscopy, laser speckles, microrelief, fractal dimension, ceramic surface, layered ceramic system.

AMS 2020 classification: 78A45, 78A55.

Introduction

With the development of modern technologies in such key industries as aviation, astronautics, medicine and energy, the need for materials with improved performance characteristics is constantly growing. In this context, ceramic materials, due to their exceptional hardness, high thermal resistance and chemical inertness, open up broad prospects for application in extreme conditions.

Among ceramic materials, layered ceramic composites are of particular interest, demonstrating significant potential in applications requiring high strength and durability (Minatto et al., 2015). Their unique architecture using specially designed weak interfacial layers is key to effective damage localization. Interfacial layers, such as those based on boron nitride (BN), promote crack deflection and delamination, preventing catastrophic failure (Ahmad et al., 2025; Chen et al., 2022) and enhancing the performance of the composite.

Bulk properties of ceramic materials, determined by the conditions of synthesis and processing, are reflected in the microstructure of the surface, the set of parameters of which determines its texture (Bowman, 1994; Behera et al., 2021). Although this relationship is often complex and nonlinear, and its analytical description poses a significant problem, it is this correlation that opens up promising opportunities for the development of non-destructive methods for studying the internal structure and properties of materials by analyzing their surface.

Traditional methods of analyzing the texture of microrough surfaces based on optical microscopy are well developed and provide high resolution. However, their application is often associated with the need for precise positioning of the sample relative to the optical system, which significantly limits the possibility of rapid screening analysis. In addition, the contact nature of some microscopic techniques and the complexity of integration into technological lines make it impossible to conduct studies in real time, for example, during the processes of forming or processing of the materials under study. Therefore, there is a need to develop new, non-contact and operational methods for determining surface characteristics. One of the possible effective solutions is the use of coherent optics methods, in particular, speckle surface microscopy.

In our previous work (Yakunov, & Sachko, 2024) it has been demonstrated that the analysis of laser speckles – characteristic spotty structures that arise from the interference of coherent light waves scattered by microscopic inhomogeneities of a rough surface – is an effective approach for studying key surface parameters. It has been shown that the statistical characteristics of the speckle field carry information about such important surface properties as its roughness, anisotropy and other spatial correlation parameters.

The aim of this study is to establish a quantitative relationship between the fractal dimension of the ceramic surface texture obtained using optical microscopy and the fractal dimension of the corresponding laser speckle field. Fractal dimension was chosen as the most physically justified texture parameter, since it is closely related to surface roughness and is able to effectively quantitatively describe complex microreliefs of ceramics, and on the other hand, it adequately describes images

with large-scale self-similarity, to which laser speckles belong. Considered potential opportunities speckle microscopy for determining layer boundaries and assessing the homogeneity of the structure in layered systems.

1. Surface texture and bulk properties of materials

Surface texture, determined by a set of micro-irregularities of different scales and orientations, is in many cases a direct reflection of the bulk properties of the material, which are formed at the stages of its synthesis, processing, and operation. Micro- and nanorelief of the surface carries information about the crystalline structure, phase composition, presence of defects, mechanical stresses and other internal characteristics of the material. Thus, detailed analysis of textural parameters provides the possibility of non-destructive study of bulk properties, opening up prospects for quality control, prediction of material behavior and development of new technological processes with specified surface characteristics. (Bunge, 2013).

Fractal methods are an effective tool for modeling rough surfaces, as they are able to adequately reflect their hierarchical organization and statistical characteristics (Militký, & Bajžík, 2001; Nayak et al., 2019). Fractal dimension, as a measure of the complexity and self-similarity of an object at different scales, quantifies the degree of roughness and spatial filling of rough surfaces. It is calculated by various methods, such as the box-counting method, structural function method or radial basis function method, which analyze the change in surface characteristics with a change in the observation scale. Formally, higher fractal dimension correlates with greater surface roughness, as it indicates the presence of finer details and a more complex hierarchical structure of the microprofile.

Scattering of coherent laser radiation on a rough surface leads to interference of secondary waves and formation of a speckle pattern, which also demonstrates signs of large-scale self-similarity. Despite the complexity and non-obviousness of the direct connection between the fractal dimension of a rough surface and its speckle field, experimental studies confirm the existence of a statistically significant correlation between these characteristics. (Li et al., 2006; Shao et al., 2023; Souza et al., 2021). This opens the prospect of non-contact estimation of surface roughness by analyzing the statistical properties of the corresponding laser speckles.

2. Materials and methods

The work investigated homogeneous composite materials and multilayer structures based on them. Two types of composites were synthesized, differing in the excess content of boron carbide (hereinafter referred to as sample A) or titanium carbide (sample B). Multilayer samples (C) were formed from the two specified composite systems, separated by a graphite layer, which was formed during the reactive hot pressing process and functioned as a weak interface for crack deflection. Detailed way sintering samples covered in the previous the work of one of the authors (Ovcharenko, Dibrov, & Semenko, 2024).

The study was conducted on samples in the form of cylindrical tablets with a diameter of 10 mm. Each sample was fixed on the microscope stage and sequentially illuminated by two sources: a white light LED lamp (for visualization of the surface microstructure) and a helium-neon laser (for formation of a speckle pattern from the same place on the surface). Laser radiation was directed to the sample surface at an angle of incidence of about 40°. The area on the sample surface illuminated by the laser beam, which formed the speckle pattern, had the shape of an ellipse, which, when visualized through the microscope objective with a 10-fold magnification, occupied the central part of the field of view. Registration of the surface microrelief and the corresponding speckle pattern was carried out by a digital camera with a resolution of 1024 x 768 pixels. The central rectangular area of the image measuring 340 x 250 pixels was characterized by an almost uniform distribution of illumination intensity both for visualization of the microrelief and for registration of the speckle pattern (Fig. 1). Optimization of exposure parameters and other camera settings, as well as conversion of color images into 8-bit format with 256 gray levels, were performed using specialized software. These procedures were aimed at obtaining images with optimal qualitative and informative characteristics necessary for further analysis.

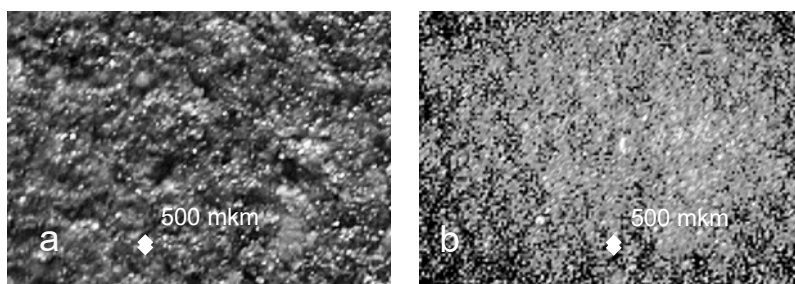


Fig. 1. Fragment of a surface microimage (a) and the corresponding speckle pattern (b)

The fractal dimension of the surface microrelief images and the corresponding speckle patterns was determined using the *ImageJ* program with specialized plugins. To verify the obtained results, the original software developed by the authors was used in parallel, which did not provide for any image preprocessing procedures. The comparative analysis showed the general consistency of the values of the fractal dimension of the microimages obtained by both methods. At the same time, certain discrepancies were observed in the values of the fractal dimension of the speckle patterns, which may be related to the specifics of their structure and the potential loss of information due to preprocessing. The numerical values given below were obtained using the original software.

□

3. Results and analysis

At the previous stage of research, the influence of the accuracy of microscope focusing on the values of fractal dimensions of microimages and corresponding speckle patterns was analyzed. Experiments were carried out using flat surface of one of samples. Figure 2 illustrates dependence fractal dimensions microscopic images and corresponding speckle structures from degree defocus optical systems that varied in steps of 40 μm.

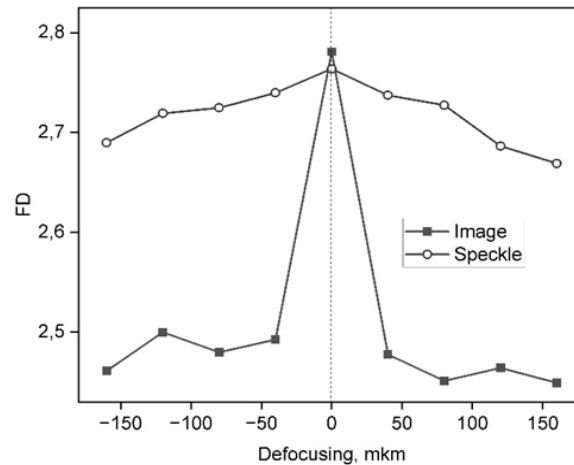


Fig. 2. Dependence of the fractal dimension of the microimage and the corresponding speckle pattern on the defocus value

Received results showed that for microimages indicated dependence characterized by sharp maximum, which corresponds to the optimal focusing condition. Even insignificant deviation leads to significant reduction value fractal dimensions. Further increase in defocus does not cause significant changes to this parameter, which probably reflects integral characteristics of the blurred surface, irregularities lighting and noise recording device.

In contrast this, similar dependence for speckle images demonstrates weak sensitivity fractal dimensions to defocusing.

With precise focusing observed minor local maximum, which can be associated with the modulation of the speckle picture by the surface relief. With increase defocus this modulation disappears, and the fractal dimension speckle paintings slightly decreases. Such behavior consistent with the physical principles of formation speckle pattern where the maxima interference form elongated three-dimensional areas. Their cross-section by any plane forms the same two-dimensional distributions of illumination, the differences between which are mainly in scale coefficients.

At the main stage experimental research was two homogeneous samples (A, B) and one five-layer (C) were analyzed. For each of the homogeneous samples 20 measurements were taken fractal dimensions by two opposite flat surfaces (conditionally labeled as top and bottom). Five-layer sample was scanned along its lateral cylindrical surface with a step of 0.5 mm.

Figure 3 shows the results in the form of a scatter plot. Each point corresponds to a specific area of the surface of one of the samples and represents a pair of fractal dimensions of the microscopic image and the corresponding speckle picture. Statistical analysis shows that all the points that respond sample A, form single cluster with high density. Data for the sample B, obtained from different its surfaces, form two close clusters that may be due to different quality processing these surfaces. However, at the same time, a slight difference does not significantly affect the overall picture.

Clearly traced statistically significant excess average value fractal dimensions of both microimages and speckle patterns of the sample B comparatively from sample A.

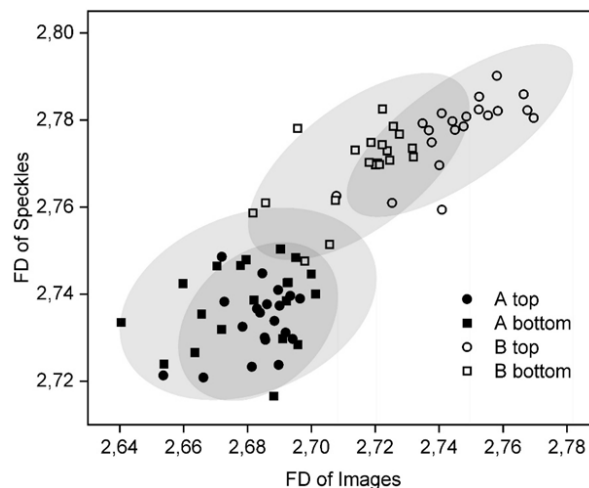


Fig. 3. Scatter plot of fractal dimension of microimage (abscissa axis) and speckle (ordinate axis) for two ceramic samples. The ellipses show a 95 % confidence region

Thus, the fractal analysis of the speckle pattern in terms of information capacity is at least not inferior to the analysis of microscopic images. However, it should be noted that the ellipses that delineate the 95 percent confidence regions show a tendency to be oriented along the horizontal axis (for this, it is necessary to take into account the ranges of values of the corresponding scales). This indicates lower scattering values of the fractal dimensions of speckle patterns compared to microimages. Thus, the set of surface parameters that are determined through speckle patterns is characterized by a higher value of the signal-to-noise ratio.

Figure 4 shows the results of scanning the lateral surface of sample C. The dependence of fractal dimensions on the coordinate of images and speckle patterns clearly demonstrates the five-layer structure of the sample. At the same time, the fractal dimension of the speckle pattern exhibits more regular behavior, which makes it a more reliable identification marker.

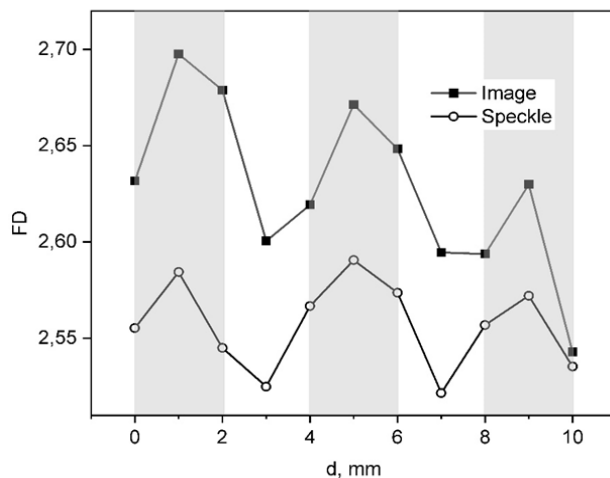


Fig. 4. Dependence of the fractal dimension on the coordinate for the microimage and the corresponding speckle pattern when scanning on the side surface of a five-layer sample

Considering low sensitivity speckle patterns to accuracy focusing optical systems, they are significantly more convenient object for express analysis surfaces ceramic materials compared to their microscopic images. Given the physical nature of the formation speckle-picture, exists possibility substantial simplification optical schemes microscopy speckle images or even full refusals from her.

An alternative approach may be to analyze the local characteristics of the surface (for example, roughness) using a speckle structure, which is formed over a small area of a given size using a focused laser beam and recorded by a video camera matrix located in a convenient place in the optical circuit.

In comparison with established techniques of non-destructive evaluation, such as ultrasonic inspection, X-ray microtomography, or Raman spectroscopy, speckle-based analysis combines operational simplicity with sufficient sensitivity to surface and near-surface features, while avoiding the need for complex and expensive instrumentation. At the same time, its limited depth of penetration and sensitivity to external perturbations remain challenges for practical application.

Discussion and conclusions

The conducted research confirms the effectiveness of fractal analysis of speckle patterns for the characterization of ceramic surfaces. The method demonstrates informativeness comparable to microscopic imaging while offering the important advantage of low sensitivity to optical focusing, which significantly simplifies measurement procedures and enables rapid control of technological parameters.

Statistically significant differences in fractal characteristics of ceramics of various compositions, revealed in both microimages and speckle patterns, indicate the suitability of the approach for material differentiation and process monitoring. Higher stability of fractal dimensions in speckle analysis highlights its greater reliability for unambiguous surface evaluation. In multilayer ceramics, the method clearly reflects stratified structures, underscoring its potential for non-destructive testing of composite systems.

Future research should aim at improving robustness of the optical setup, expanding depth-sensitive capabilities, and integrating speckle analysis with complementary methods to establish comprehensive and reliable diagnostics of ceramic materials.

Authors' contribution: Andriy Yakunov – conceptualization; methodology; source analysis; empirical research; Volodymyr Dibrov – preparation of experimental samples; empirical research; Artem Sachko – empirical research; computer calculations.

Sources of funding. The research was partially funded by Taras Shevchenko National University of Kyiv.

References

- Ahmad, Z., Chen, J., Chen, H., Arfan, M., Shahid, T., Li, X., & Hussain, Z. (2025). Tailoring structural and mechanical properties of Cf/SiC ceramic matrix composites with BN/SiAlC interphases. *Ceramics International*, 51(5), 5799–5807. <https://doi.org/10.1016/j.ceramint.2024.12.026>
- Behera, R. P., Muhammad, S. B. S., Jiaxuan, M. H., & Le Ferrand, H. (2021). Porous textured ceramics with controlled grain size and orientation. *Journal of the European Ceramic Society*, 41(1), 617–624. <http://dx.doi.org/10.48550/arXiv.2103.01445>
- Bowman, K. J. (1994). Textures in Ceramic Materials. *Materials Science Forum*, 157(6), 43–56.
- Bunge, H.-J. (2013). *Texture analysis in materials science: mathematical methods*. Elsevier. <https://doi.org/10.1016/B978-0-408-10642-9.50006-4>
- Chen, M., Pan, L., Xia, X., Zhou, W., & Li, Y. (2022). Boron nitride (BN) and BN based multiple-layer interphase for SiCf/SiC composites: a review. *Ceramics International*, 48(23), 34107–34127. <http://dx.doi.org/10.1016/j.ceramint.2022.07.021>
- Li, Z., Li, H., Qui, Y. (2006). Fractal analysis of laser speckle for measuring roughness. In *ICO20: Optical Information Processing: Vol. 6027* (pp. 470–475). SPIE. <https://doi.org/10.1117/12.668182>
- Militký, J., & Bajžík, V. (2001). Surface roughness and fractal dimension. *Journal of the Textile Institute*, 92(3), 91–113. <http://dx.doi.org/10.1080/00405000108659617>

- Minatto, F. D., Milak, P., De Noni, A., Hotza, D., & Montedo, O. (2015). Multilayered ceramic composites—a review. *Advances in Applied Ceramics*, 114(3), 127–138. <http://dx.doi.org/10.1179/1743676114Y.0000000215>
- Nayak, S. R., Mishra, J., & Palai, G. (2019). Analysing roughness of surface through fractal dimension: A review. *Image and Vision Computing*, 89, 21–34. <https://doi.org/10.1016/j.imavis.2019.06.015>
- Ovcharenko, A., Dibrov, V., & Semenko, M. (2024). Reactive Sintering of HfB₂-SiC-C Ultra-High Temperature Ceramics with Enhanced Thermal Shock Resistance. *Journal of Nano- and Electronic Physics*, 16(6). [https://doi.org/https://doi.org/10.21272/jnep.16\(6\).06024](https://doi.org/https://doi.org/10.21272/jnep.16(6).06024)
- Shao, M., Xu, D., Li, S., Zuo, X., Chen, C., Peng, G., Jia, Z., Xiao, W., & Quan, Y. (2023). A review of surface roughness measurements based on laser speckle method. *Journal of Iron and Steel Research International*, 30(10), 1897–1915. <http://dx.doi.org/10.1007/s42243-023-00930-8>
- Souza, P. V. S., Rodrigues, D., da Silva, L., Huguenin, J., & Balthazar, W. (2021). Characterization of surface roughness by speckle pattern – an experimental approach. *Physics Education*, 56(2), 025013. <http://dx.doi.org/10.1088/1361-6552/abd2cc>
- Yakunov, A., & Sachko, A. (2024). Application of texture analysis to study the relationship between surface microrelief and laser speckles. *Bulletin of Taras Shevchenko National University of Kyiv. Physical and Mathematical Sciences*, 79(2), 100–103. <https://doi.org/https://doi.org/10.17721/1812-5409.2024/2.16>

Отримано редакцією журналу / Received: 25.05.25
Прорецензовано / Revised: 08.09.25
Схвалено до друку / Accepted: 10.10.25

Андрій ЯКУНОВ, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0002-2664-4803
e-mail: yakunov@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

Володимир ДІБРОВ, асп.
ORCID ID: 0009-0005-8346-9254
e-mail: dibrov.volodymyr.9@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

Артем САЧКО, учень
e-mail: soovaf2@gmail.com
Український фізико-математичний ліцей, Київ, Україна

КОРЕЛЯЦІЯ МІЖ ФРАКТАЛЬНОЮ РОЗМІРНІСТЮ ТЕКСТУРИ ПОВЕРХНІ ТА ЇЇ ЛАЗЕРНИМИ СПЕКЛАМИ ЯК ОСНОВА БЕЗКОНТАКТНОГО МЕТОДУ АНАЛІЗУ КЕРАМІЧНИХ МАТЕРІАЛІВ

Досліджено кількісний зв'язок між фрактальною розмірністю текстури керамічної поверхні, отриманої за допомогою оптичної мікроскопії, та фрактальною розмірністю відповідного лазерного спекл-поля. Виявлено низьку чутливість спекл-аналізу до дефокусування, що спрощує процедуру контролю. Аналіз однорідних зразків виявив статистично значущі відмінності фрактальних характеристик зображень двох видів кераміки, підтвердивши можливість їхнього використання для ідентифікації матеріалів і виявлення дефектів. У випадку шаруватого керамічного зразка фрактальний аналіз спекл-картин чітко відображав його структуру, вказуючи на потенціал методу для неруйнівного контролю шаруватості й однорідності. Продемонстрований підхід підкреслює його метрологічну надійність і відкриває перспективи для його інтеграції у стандартизовані схеми контролю якості, забезпечуючи послідовну та відтворювану оцінку структурної цілісності у виробництві кераміки. Перспективи подальших досліджень пов'язані з розробленням спрощених оптичних схем спекл-аналізу для створення портативних систем експрес-контролю якості керамічних виробів.

Ключові слова: оптична мікроскопія, лазерні спекли, мікрорельєф, фрактальна розмірність, керамічна поверхня, шарувата керамічна система.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

Наукове видання



ВІСНИК
КИЇВСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ ТАРАСА ШЕВЧЕНКА
ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

Випуск 2(81)

Оригінал-макет виготовлено Видавничо-поліграфічним центром "Київський університет"

Автори опублікованих матеріалів несуть повну відповідальність за підбір, точність наведених фактів, цитат, економіко-статистичних даних, відповідної галузевої термінології, власних імен та інших відомостей. Редколегія залишає за собою право скорочувати та редагувати подані матеріали.

Opinions, statements, accuracy of the quotations, economic and statistical data, terminology, proper names and other information are made on the responsibility of the authors. The Editorial Board reserves the right to shorten and edit the submitted materials.



Формат 60x84^{1/8}. Обл. вид. арк. 27,02. Ум. друк. арк. 27,08. Наклад 300. Зам. № 225-11604.
Гарнітура Arial. Папір офсетний. Друк офсетний. Вид. №К2.
Підписано до друку 22.12.25

Видавець і виготовлювач
ВПЦ "Київський університет"
Б-р Тараса Шевченка 14, м. Київ, 01030
☎ (38044) 239 32 22; (38044) 239 31 58; (38044) 239 31 28
e-mail: vpc@knu.ua; vpc_div.chief@univ.net.ua; redaktor@univ.net.ua
http: vpc.knu.ua

Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02