ISSN 2304–6201

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

ВІСНИК

Харківського національного університету імені В.Н. Каразіна

№ 1131

Серія
«Математичне моделювання.
Інформаційні технології.
Автоматизовані системи управління»

Випуск 25 Серія заснована 2003 р.

Харків
2014

Статті містять дослідження у галузі математичного моделювання та обчислювальних методів, інформаційних технологій, захисту інформації. Висвітлюються нові математичні методи дослідження та керування фізичними, технічними та інформаційними процесами, дослідження з програмування та комп'ютерного моделювання в наукоємних технологіях.

Для викладачів, наукових працівників, аспірантів, працюючих у відповідних або суміжних напрямках.

Затверджено до друку рішенням Вченої ради Харківського національного університету імені В. Н. Каразіна (протокол № 10 від 03.11.2014 р.)

Адреса редакційної колегії: 61022, м. Харків, майдан Свободи, 6,
ХНУ імені В. Н. Каразіна, к. 538.
Тел. +380 (57) 705-42-81, Email: Victor.O.Mischenko@univer.kharkov.ua .

Статті прорецензовано.

Свідоцтво про державну реєстрацію КВ № 11825-696 ПР від 04.10.2006.

ЗМІСТ

УДК: 004.712 519.172.4

Методи покращення ефективності для систем високошвидкісної класифікації пакетів

Ю. В. Бойко, К. С. Дєєв
Київський національний університет імені Тараса Шевченка

В статті розглянуті методи та підходи в реалізації класифікаторів мережевого трафіку. Вказані інструменти використовуються як детектори аномальної мережевої активності, які засновані на імплементації алгоритму Aho-Corasick. Основна частина роботи присвячена огляду шляхів підвищення ефективності роботи класифікатора та мінімізації часу обробки мережевого пакету потрібного для визначення його належності до окремого класу трафіку. Висновки, отримані в роботі, можуть бути використані для створення розподіленої системи класифікації пакетів з оптимальною архітектурою.

Ключові слова: *Мережевий моніторинг, аналіз трафіку, класифікація пакетів, виявлення вторгнень.*

В статье рассмотрены методы и подходы в реализации классификаторов трафика. Указанные инструменты используются как детекторы аномальной сетевой активности, которые основаны на имплементации алгоритма Aho-Corasick. Основная часть работы посвящена обзору путей повышения эффективности работы классификатора и минимизации времени обработки сетевого пакета, требующегося для определения его принадлежности к определенному классу трафика. Выводы, полученные в работе, могут быть использованы для создания распределённой системы классификации пакетов с оптимальной архитектурой.

Ключевые слова: *Сетевой мониторинг, анализ трафика, классификация пакетов, обнаружение вторжений.*

The article describes methods and approaches chosen for development of network traffic classifiers. These tools usable as detectors of abnormal activities are based on Aho-Corasick algorithm implementation. The main part of the work is devoted to the overview of ways to improve the classifiers efficiency as well as to minimization of processing time needed to detect traffic class the particular network packet belongs to. Obtained conclusions can be used to create distributed classification system with optimal architecture.

Keywords: *Network monitoring, traffic analysis, packet classifying, intrusion detection system.*

Вступ

За останні роки можна спостерігати велике підвищення загроз безпеці критичних компонентів сучасних мережевих архітектур. Ці системи працюють на основі політик послідовної перевірки ряду наперед сформованих правил та відповідних процедур, які необхідно виконати при співпадінні того чи іншого правила для окремого мережевого пакета чи композиції таких пакетів (режим перевірки сесій). Будемо вважати, що існує набір правил: $\langle R_1, ..., R_n \rangle$, класифікатор ідентифікує набір правил, які мають співпасти для позитивного спрацювання. Таким чином, класифікація пакетів це механізм що встановлює відповідність приналежності пакету до окремого правила шляхом перевірки мережевого пакету чи окремих його заголовків та встановлення як ця відповідність має бути в подальшому відпрацьована для аналізу. Пакетні

класифікатори таких систем мають працювати в режимі реального часу на високошвидкісних інтерфейсах мережевих плат та одночасно мінімізувати чи зовсім відкидати можливість помилкових спрацювань чи втрати пакетів[1]. Більш-того, розвиток технологій вимагає від вказаних систем змін які стосуються процедур обробки пакетних заголовків, а саме, якщо раніше достатньо було перевіряти відповідність IP-адрес та пізніше номерів портів то тепер необхідно аналізувати більш широкий набір параметрів: TCP-ідентифікатори з'єднань, TTL, поля TOS/DSCP. Спроможна здатність є одним з основних показників кількісних характеристик комплексу. Можливість аналізувати корисне навантаження в режимі реального часу потребує мережевих карт з збільшеними пакетними буферами.

Мета даної роботи сформувати швидку і масштабовану систему пакетної класифікації, яка б могла застосовуватись у різних випадках. Проблема класифікації пакетів має дві несхожі компоненти:

- співпадіння на основі пакетних заголовків;
- глибока перевірка пакетного навантаження;

Кожен з підходів має свої недоліки та переваги, які будуть обговорені в слідуючих розділах.

1. Методи аналізу заголовків

Декілька програмних розробок визначають правило яке може застосовуватись до пакету ґрунтуючись на значенні декількох полів IP-заголовку мережевого пакету. Швидкодія систем на основі перевірки списку правил для кожного пакету достатньо істотно деградує із збільшенням кількості правил чи створенні комплексних схем перевірки, які потребують співпадінь по багатьом критеріям. Попередні техніки що використовуються в класифікаторах також залежать від даного ефекту, але використовували різні оптимізації, зокрема представлення шаблону у вигляді бінарного дерева [10], або використання невизначеного автомату станів[2,3].

Але вони не спростовували залежності від кількості правил, що було неодноразово перевірено на практиці, коли кількість правил досягає декількох тисяч, як приклад у системі Snort NG[13]. Емпірична формула показує збільшення часу аналізу в 50 разів на кожні нові 2500 правил, тобто залежність порядку $\sqrt{N}$, де N – кількість нових правил. Більш того, різні додатки потребують різної поведінки при співпадіння пакета за рядом критеріїв. Частина застосунків можуть потребувати лише пакетної фільтрації, іншим може бути необхідна також класифікація. Попередні дослідження намагалися вирішити проблему шляхом перетворення(BPF, DPF, PathFinder) [4], та зосереджувалися на пакетній фільтрації. Опис алгоритму множинної класифікації за ключовими ознаками описано в [13], але слід зазначити що він не підтримує пріоритетів.

2. Постановка проблеми

За основні критерії в визначенні продуктивності скінченого автомата класифікатора в даній роботі є підтримка різних застосунків, час спрацювання та розмірність автомата. Представлений механізм дозволяє проводити класифікацію за допомогою спрощеного інтерфейсу взаємодії.

В алгоритмі використовується операція розкладу на стани, які засновані на понятті залишкового стану по відношенню до іншого. Якщо розглядати аналогії, то це операція подібно до взяття остачі від ділення цілих чисел, так само як поділ забезпечує основу. Такі конструкції важливі в плані оптимізації розташування правил при побудові автомата станів, оскільки в теорії це може мінімізувати розмірність кінцевого автомата, що в свою чергу впливає на швидкість роботи класифікатора[7].

Створення техніки вибору порядку розташування правил має велике значення в побудові, ефективного в плані часу обробки, автомата. Використання такого підходу може вносити похибку позитивного спрацювання може призвести, але в той же час це призводить до значного скорочення часу обробки великих списків правил, тобто зниження розміру автоматів для певних наборів правил. Якщо головною метою є встановлення однозначної відповідності того чи іншого пакета до окремого класу трафіку то необхідно застосовувати автомат побудований з меншим коофіцієнтом невизначених станів.

Використання часу як основної метрики є наслідком невпинного зростання швидкості каналів Інтернет та об'ємів даних що через них передаються. Створення ефективного підходу в організації вимірів мережевого трафіку дозволить встановлювати класифікатори не лише як апаратні системи аналізу трафіку, але як і програмні комплекси пост-обробки перехопленого трафіку.

Дану проблематику розглянуто такими аторами, як L. Bailey, B. Gopal, A. Pagels, L. Peterson, T. Lakshman, D. Stiliadis, Z. Chen, Y. Diao, T. Lakshman та ін.

3. Огляд рішень

Декілька систем мережевого моніторингу та систем мережевої безпеки обробляють набір пакетів як одне ціле в залежності від типу протоколу корисного навантаження верхнього рівня, базуючись на інформації яка біла отримана з пакетних заголовків. Прикладом таких програмних застосунків є система Snort[20], популярна відкрита IDS, яка порівнює пакети засновуючись на сигнатурах описаних через спеціальні правила. Розвиток вказаної системи йшов шляхом створення ефективних методів опису правил аналізу співпадіння за шаблоном па подальшого розпаралелення функції їх перевірки. Ці вирази можуть бути скомпільовані для застосування в ролі автомата станів DFA (Deterministic Finite Automata). Аналіз цих випадків детально розглянуто в [5]. Розроблений алгоритм дозволяє проводити класифікацію пакетів шляхом опису правил подібних до продукту Snort. В окремому випадку система дозволяє проводити класифікацію використовуючи схему співпадіння за регулярним виразом. В подальшому для спрощення опису рішення будемо вважати що, перевірка заголовків буде називатись перевіркою пакета, а перевірка навантаження того ж пакета – пакетною інспекцією.

3.1 Проблема класифікації

Опишемо два визначення для представлення умов створення правил класифікації, будемо їх називати фільтрами. Для спрощення представлення кожен фільтр буде ідентифікувати лише одне правило.

Визначення 1. (Синтаксис)

Форма запису виразу для фільтра має відповідати наступним вимогам:

- містить змінну що перевіряється (***x***) та одну чи декілька констант (***c***);
- дозволяє використання бітових масок(***x&c1=c***);
- підтримує можливість встановлення нерівності полів заголовку (***x≠c***);
- встановлення відношень між елементми заголовку пакета що перевіряється (***x<c***);

Таким чином, правило може бути записане як:

$$(\textbf{dport} = 80)\ \&\&\ (\textbf{sport} > 1024)\ ||\ (\textbf{flags}\&0xb = 0x3) \qquad (1)$$

Фільтр C записаний у вигляді вираза (1) може бути застосований до мережевого пакета P. Відповідна операція може біти записана у вигляді $C(P)$. За співпадіння будемо вважати позитивне логічне «істина», яке базується на застосуванні фільтра до пакетного навантаження.

Визначення 2. (Пріоритетна обробка)

Фільтр F це ланцюжок перевірок C. Набір фільтрів Ф формується з одиничних записів C, які відсортовані за пріоритетом. Пріоритет F оголошується як *pri*(F).

Для набору фільтрів Φ, ми вважаємо що $F \in \Phi$, відповідає співпадінню пакета P і означає: $M_{\Phi}(F,P)$, якщо $F(P)$ істина та $F'(P)$ хибне для $\forall F' \in \Phi$, що має вищий пріоритет ніж F. Таким чином пакет не може бути оброблений фільтром до того часу коли всі пріоритетні правила будуть оброблені. Наприклад, розглянемо набір правил:

- F_1 : (***icmp_type*** = *ECHO*)
- F_2 : (***icmp_type*** = *ECHO_REPLY*) || (***ip_ttl*** = *1*)
- F_3 : (***ip_ttl*** = *1*)

Будемо вважати F_1 відповідає пакету P_1, а F_2 в свою чергу P_2.

Якщо набір би мав не встановлені пріоритети то правило F_1 спрацювало б для P_1, F_2 для P_2, а F_3 перевірялося б для обох пакетів. Тобто,

$M_{\Phi}(P_1) = \{F_1, F_2\}$ та $M_{\Phi}(P_2) = \{F_2, F_3\}$,

якщо $pri(F_1) > pri(F_2) > pri(F_3)$ то $M_{\Phi}(P_1) = \{F_1\}$ та $M_{\Phi}(P_2) = \{F_2\}$,

якщо $pri(F_3) > pri(F_2) > pri(F_1)$ то $M_{\Phi}(P_1) = M_{\Phi}(P_2) = \{F_3\}$,

якщо $pri(F_1) = pri(F_2) > pri(F_3)$ то $M_{\Phi}(P_2) = \{F_3\}$ та $M_{\Phi}(P_1) = \{F_1\} \wedge M_{\Phi}(P_1) = \{F_2\}$.

3.2 Визначення пріоритетності

Пакетна фільтрація може відбуватися за умови встановлення рівного пріоритету для всіх фільтрів. Таким чином співпадіння буде можливим одразу після першого правила що співпадає. Співпадіння за списком також можливе, для цього необхідно встановити пріоритет для кожного правила. Як наслідок перших двох, можливе співпадіння за багатьма критеріями заголовку[11].

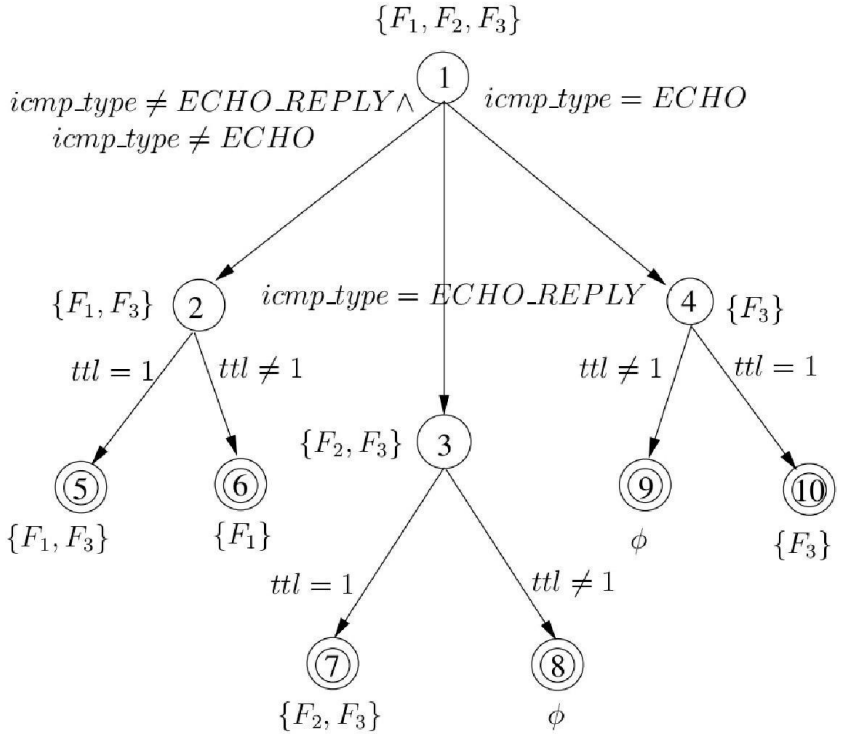


Рис. 1 Автомат скінченних станів

Схеми автоматів для наборів фільтрів з різним пріоритетом співпадіння по шаблону наведено на Рис.1 та Рис. 2. Такі автомати відомі як класифікаційні. Переходи між станами у невизначеному автоматі на Рис. 2 характеризуються наявністю початкових умов в вхідних правилах.

Під час виконання класифікації будемо вважати що деякий мережевий пакет задовольняє умові T_i, наприклад, ***icmp_type*** = *ECHO*. Перехід відбувається якщо T_i, не трапляється раніше у наборі правил або перед ним не стоїть правила з більшим пріоритетом. Перевірку полів заголовка ір-пакета можливо виконувати в будь-якій послідовності. Якщо продовжувати розглядати, наприклад з ***icmp_type,*** то поле буде розташоване за зміщенням 35 байтів, рахуючи від початку ір-заголовка, перед полем ***ttl***, яке розташоване за 13 байт раніше.

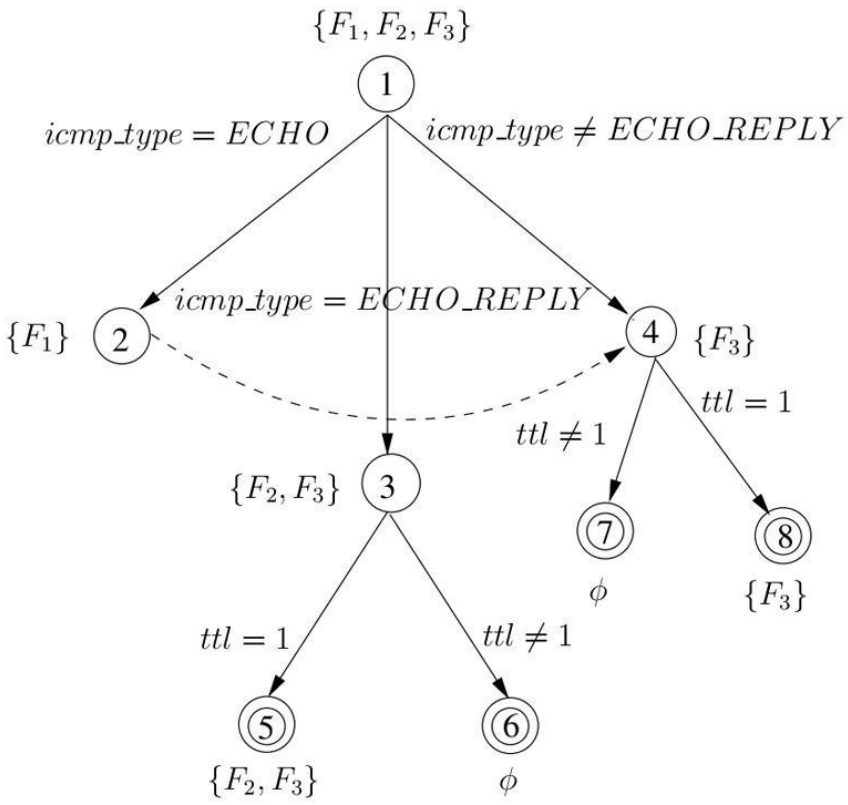


Рис.2 Невизначений автомат збігу станів

Використовуючи формат опису правил подібний до того що описано в [6], процес встановлення правил зводиться до прямого опису зміщень полів які необхідно досліджувати. Перед компіляцією кінцевих правил поводиться аналіз синтаксису на предмет взаємо-виключаючих станів.

Тобто збережена логічна послідовність проведення тестів:
*{**proto** = ICMP} : (**icmp_type** = ECHO),*
перевірка типу повідомлення можлива, якщо тип пакету – ICMP.

Існує дві кількісні характеристики роботи автомата – його розмір та час спрацювання. Загалом більшість фільтрів містить невелику кількість тестів в той час як кількість самих фільтрів досить велика. В результаті довжина шляху проходу в автоматі є короткою в порівнянні з його шириною.

Час спрацювання автомата прямо залежить від його довжини. Середнє значення буде залежати від розподілу пакетів в час проведення тесту.

4. Результати

Якщо застосовувати запис полів за відомим зміщенням, можна проводити аналіз вмісту пакета за допомогою перевірки відповідності текстовому регулярному виразу. Визначення збігу з описом в регулярних виразах є досить

популярним в системах глибокого аналізу пакетів (DPI). Для розширення можливостей обробника необхідно виконання двох умов:

- фіксований результат збігу;
- відома адреса зміщення в пакетному навантаженні;

Виконання першої умови важливо з огляду на попередню компіляцію правил, оскільки це задається жорсткою умовою при формуванні скінченого автомата подій. Виявлення збігу за регулярним виразом найбільш доцільно виконувати за алгоритмом Aho-Corasick [1], основою роботи алгоритму є представлення всіх значень у вигляді дерева. В результаті кінцева схема підтримки пошуку по текстовим полям в пакетному навантаження зводиться до оголошення додаткових перевірок, а саме - пошуку збігу. Процедура генерації фільтра не описується, оскільки залежить від операційної системи, ознайомитись з виразами можна на офіційному сайті документації [12]. Процедура перехоплення мережених пакетів неодноразово описувалась [5], [7]. Аналізатор може використовуватись в режимі IDS або як класифікатор трафіку для виявлення його приналежності до відповідного сервісу (поля DSCP або ToS).

Висновки

Підвищення ефективності роботи комплексів для класифікації мережевих пакетів має велике значення для систем аналізу з'єднань та запобігання вторгнень в мережу. Запропонований алгоритм показує покращення швидкості роботи класифікатора при попередній компіляції наборів правил.

Алгоритм підтримує встановлення пріоритетів для правил (для фільтрів) та можливість роботи у режимі пошуку максимальної кількості співпадінь. Запропонована техніка показує що зменшення розміру скінченого автомата досягається за рахунок мінімізації повторення проходів через класифікатор для правил що стосуються ключового поля IP-пакета. Як результат це також сприяє зменшенню часу спрацювання.

Розглянуті механізми можуть застосовуватись у побудові масштабованих систем мереженої класифікації пакетів. В подальшому планується провести інтеграцію з системою Snort, для можливості роботі системи в режимі IDS, та з набором утиліт класифікатора OpenDPI (набір бібліотек для класифікації мережевих пакетів в режимі реального часу). Застосування оптимізованих правил дозволяє проводити ефективну роботу в напрямку перехоплення пакетів на високошвидкісних каналах зв'язку Інтернет мережі та проводити статистичний аналіз цих даних з метою встановлення розподілів застосування протоколів та проведення планування ємності каналів з метою їх подальшого розширення. В рамках проведеного дослідження встановлено можливість створення програмної реалізації, яка б працювала незалежно від мереженого стеку операційної системи в режимі фільтру, за основу доцільно використовувати комплекс NetGraph (операційна система FreeBSD)[8] або DFA (операційна система LINUX)[9]. Вказані підходи дозволять абстрагуватися від апаратного забезпечення мережевого класифікатора, оскільки процедури обробки пакетів будуть виконуватись в ядрі операційної системи.

ЛІТЕРАТУРА

1. L. Bailey, B. Gopal, A. Pagels, L. Peterson. Pathfinder: A pattern-based packet classifier., Operating Systems Design and Implementation, p. 115-123, 1994.
2. A. Begel, S. McCanne, L. Graham. BPF+: Exploiting global data-flow optimization in generalized packet filter architecture. In SIGCOMM, p. 123-134, 1999.
3. S. Chandra, P. McCann. Packet types. Workshop on Compiler Support for Systems Software (WCSSS), May 1999.
4. P. Gustafsson, K. Sagonas. Efficient manipulation of binary data using pattern matching., J. Funct. Program., p. 16-35, 2006.
5. J. Hopcroft, R. Motwani. Introduction to Automata Theory, Languages, and Computation. Addison Wesley, 2001.
6. C. Kruegel, T. Toth. Using decision trees to improve signature-based intrusion detection., Symposium on Recent Advances in Intrusion Detection (RAID), 2003.
7. T. Lakshman, D. Stiliadis. High-speed policy-based packet forwarding using efficient multi-dimensional range matching. In SIGCOMM, p. 203-214, 1998.
8. S. McCanne, Van Jacobson. The BSD packet filter: A new architecture for user-level packet capture. In USENIX Winter, p. 259-270, 1993.
9. G. Varghes. Packet classification using multidimensional cutting. SIGCOMM, 2003.
10. U. Manber. A fast algorithm for multi-pattern searching. Technical Report TR94-17, 1994.
11. Z. Chen, Y. Diao, T. Lakshman. Fast and memory-efficient regular expression matching for deep packet inspection. In Architectures for Networking and Communications Systems, p. 93-100, 2006.
12. FreeBSD Kernel Interfaces Manual [Електронний ресурс] . – Режим доступу: http://www.freebsd.org/cgi/man.cgi?bpf%284%29
13. Snort - open source network IDS/IPS [Електронний ресурс] . – Режим доступу: http://www.snort.org/

Надійшла у першій редакції 02.06.2014, в останній - 30.10.2014.

УДК 51-72

Температурные напряжения, возникающие в бесконечном стержне в рамках пространственно нелокальной модели термоупругости

Л. И. Брацыхина, М. В. Сынах, Л. А. Фильштинский
Сумской государственный университет, Украина

В статье рассматривается одномерная пространственно-нелокальная задача термоупругости. Для решения задачи применялся метод интегральных преобразований и численное интегрирование. Получены графики распределения смещения и температуры для различных значений временного параметра.

Ключевые слова: *фрактальное уравнение теплопроводности, дробная производная Капуто, дробная производная Рисса, задача нелокальной термоупругости.*

В статі розглядається одновимірна просторово-нелокальна задача термопружності. Для розв'язання задачі було застосовано метод інтегральних перетворень та чисельне інтегрування. Отримано графіки розподілу переміщення і температури для різних значень часового параметру.

Ключові слова: *фрактальне рівняння теплопровідності, дробова похідна Капуто, дробова похідна Ріcса, задача нелокальної термопружності.*

In the article, 1D spatially nonlocal thermoelasticity problem is considered. The method of integral transforms and numerical integration were applied to solve this problem. Distributions of dispalcement and temperature for various values of time parameter are obtained.

Key words: *fractional heat conduction equation, Caputo fractional derivative, Riesz fractional derivative, nonlocal thermoelasticity problem.*

1. Общая постановка задачи

В классической теории термоупругости связь между напряжениями, перемещениями и температурой определяется соотношениями Дюгамеля-Неймана

$$\sigma_{ij} = \mu\left(\partial_j u_i + \partial_i u_j\right) + \left[\lambda e - \beta T\right]\delta_{ij}, \tag{1}$$

где λ, μ – постоянные Ламе, β – коэффициент температурных напряжений, σ_{ij} – компоненты тензора напряжений, u_i – компоненты вектора перемещений, $e = \partial_1 u_1 + \partial_2 u_2 + \partial_3 u_3$, $\partial_i u_j = \partial u_j / \partial x_i$, T – температура.

В одномерном случае (1) запишется в виде

$$\sigma\left(x,t\right) = \left(2\mu + \lambda\right)e - \beta T\left(x,t\right),\ e = \partial u/\partial x. \tag{2}$$

Следуя работе [1], обобщим (2) следующим образом

$$\sigma\left(x,t\right) = \left(2\mu + \lambda\right)e_\gamma - \beta T,$$

$$e_\gamma = \frac{1}{2}\left[{}^{C}D^{\gamma}_{a+}u\left(x,t\right) - {}^{C}D^{\gamma}_{b-}u\left(x,t\right)\right],\ 0 < \gamma < 1. \tag{3}$$

Подставляя (3) в уравнение движения $\rho\partial^2 u(x,t)/\partial t^2 = \partial\sigma/\partial x$, будем иметь

$$\frac{\partial^2 u}{\partial t^2} - E_0 \frac{\partial}{\partial x}\left[{}^C D^{\gamma}_{a+} u(x,t) - {}^C D^{\gamma}_{b-} u(x,t)\right] = -B\frac{\partial T}{\partial x}, \tag{4}$$

где $E_0 = (2\mu+\lambda)/(2\rho)$, $B = \beta/\rho$.

Температура T определяется из дробно-дифференциального уравнения теплопроводности [2] (при отсутствии тепловых источников)

$$\frac{\partial T}{\partial t} = \frac{\partial^{\beta} T}{\partial |x|^{\beta}},\ 1 < \beta < 2. \tag{5}$$

Соотношения (3)-(5) описывают одномерную модель динамической несвязанной термоупругости, учитывающую нелокальные упругие взаимодействия между частицами среды, а также аномальный характер теплопроводности.

В качестве примера реализации данной модели рассмотрим задачу Коши для фрактального уравнения теплопроводности и вычислим соответствующие температурные напряжения.

Математическую формулировку задачи запишем следующим образом

$$\frac{\partial^2 u}{\partial t^2} - E_0 \frac{\partial}{\partial x}\left[{}^C D^{\gamma}_{+} u(x,t) - {}^C D^{\gamma}_{-} u(x,t)\right] = -B\frac{\partial T}{\partial x},\ 0 < \gamma < 1, \tag{6}$$

$$\frac{\partial T}{\partial t} = \frac{\partial^{\beta} T}{\partial |x|^{\beta}},\ 1 < \beta < 2,\ -\infty < x < +\infty,\ t > 0, \tag{7}$$

начальные условия

$$T(x,0) = \delta(x),\ u(x,0) = 0, \tag{8}$$

где $\delta(x)$ – дельта-функция Дирака [3].

2. Истоки исследования авторов

Процессы переноса частиц и энергии (диффузия и теплопроводность соответственно), возникающие в пористых материалах, аморфных полупроводниках, перколяционных кластерах, полимерных структурах, называют аномальными, а иногда и фрактальными из-за их связи с дробно-дифференциальным исчислением [4-8].Наиболее естественным и удобным математическим аппаратом описания процессов аномальной диффузии (теплопроводности) на некотором множестве являются уравнения в частных дробных производных как по пространственным координатам, так и по времени [9, 10].

Экспериментально доказано, что в некоторых неоднородных средах процессы переноса не могут быть описаны законом Фика и уравнением теплопроводности, экспериментальные данные свидетельствуют о наличии больших «хвостов», связанных с полетами Леви [11, 12]. Последние относятся к случайным блужданиям со смещениями частиц, распределенных в соответствии с устойчивыми законами Леви. Пространственно-дробное уравнение аномальной теплопроводности (диффузии) можно получить, исходя из универсального уравнения (Паули) [13], либо на основании модели случайных блужданий, переходя к макромасштабному пределу.

Уравнения фрактальной термоупругости получены в работах [14, 15,16]. Ю. Повстенко записал уравнения квазистатической термоупругости, основанные на дробно-диффузионном уравнении теплопроводности, а также рассмотрел несколько задач об определении напряжений во фрактальных средах [17, 18, 19].

В настоящее время особый интерес представляют граничные задачи фрактальной теплопроводности, термоупругости, диффузии. В литературе представлены, в основном, только численные методы их решения, основанные на аппроксимациях дробной производной Рисса-Феллера. Поэтому важной задачей является разработка аналитических либо численно-аналитических методов решения дробно-дифференциальных уравнений и соответствующих граничных задач.

3. Основные понятия дробного интегро-дифференцирования

В данном разделе мы кратко перечислим основные определения и свойства дробного интегро-дифференцирования, необходимые для рассмотрения неклассических (фрактальных) задач теплопроводности и термоупругости.

3.1 Дробные производные и интегралы Римана-Лиувилля

Пусть $\Omega = [a,b]$ – конечный интервал на действительной оси R^1. Дробные интегралы $I_{a+}^{\alpha} f$ и $I_{b-}^{\alpha} f$ Римана-Лиувилля порядка α ($\mathrm{Re}\,\alpha > 0$) определяются следующим образом

$$\left(I_{a+}^{\alpha} f\right)(x) = \frac{1}{\Gamma(\alpha)} \int_a^x \frac{f(t)\,dt}{(x-t)^{1-\alpha}} \quad (x > a) \tag{9}$$

и

$$\left(I_{b-}^{\alpha} f\right)(x) = \frac{1}{\Gamma(\alpha)} \int_x^b \frac{f(t)\,dt}{(t-x)^{1-\alpha}} \quad (x < b), \tag{10}$$

соответственно. Здесь $\Gamma(\alpha)$ – Гамма-функция [20]. Эти интегралы называются левосторонним и правосторонним дробными интегралами.

С учетом (9), (10) дробные производные Римана-Лиувилля $D_{a+}^{\alpha} y$ и $D_{b-}^{\alpha} y$ порядка α $(\mathrm{Re}\,\alpha \geq 0)$ определяются так

$$\left(D_{a+}^{\alpha}y\right)(x)=\left(\frac{d}{dx}\right)^{n}\left(I_{a+}^{n-\alpha}y\right)(x)$$

$$=\frac{1}{\Gamma(n-\alpha)}\left(\frac{d}{dx}\right)^{n}\int_{a}^{x}\frac{y(t)\,dt}{(x-t)^{\alpha-n+1}}\quad\left(n=\left[\mathrm{Re}(\alpha)\right]+1;x>a\right),\quad(11)$$

$$\left(D_{b-}^{\alpha}y\right)(x)=\left(-\frac{d}{dx}\right)^{n}\left(I_{b-}^{n-\alpha}y\right)(x)$$

$$=\frac{1}{\Gamma(n-\alpha)}\left(-\frac{d}{dx}\right)^{n}\int_{x}^{b}\frac{y(t)\,dt}{(t-x)^{\alpha-n+1}}\quad\left(n=\left[\mathrm{Re}(\alpha)\right]+1;x<b\right),\qquad(12)$$

где $\left[\mathrm{Re}(\alpha)\right]$ – целая часть $\mathrm{Re}(\alpha)$.

При действительных $0<\alpha<1$ формулы (11) и (12) существенно упрощаются:

$$\left(D_{a+}^{\alpha}y\right)(x)=\frac{1}{\Gamma(1-\alpha)}\frac{d}{dx}\int_{a}^{x}\frac{y(t)\,dt}{(x-t)^{\alpha}},\ (0<\alpha<1;\ x>a)$$

$$\left(D_{b-}^{\alpha}y\right)(x)=-\frac{1}{\Gamma(1-\alpha)}\frac{d}{dx}\int_{x}^{b}\frac{y(t)\,dt}{(t-x)^{\alpha}}\ (0<\alpha<1;\ x<b).$$

На всей действительной оси (т. е. для $-\infty<x<+\infty$) имеют место следующие формулы при $0<\alpha<1$

$$\left(I_{+}^{\alpha}f\right)(x)=\frac{1}{\Gamma(\alpha)}\int_{-\infty}^{x}\frac{f(t)\,dt}{(x-t)^{1-\alpha}},\ \left(I_{-}^{\alpha}f\right)(x)=\frac{1}{\Gamma(\alpha)}\int_{x}^{\infty}\frac{f(t)\,dt}{(t-x)^{1-\alpha}},$$

$$\left(D_{+}^{\alpha}y\right)(x)=\frac{1}{\Gamma(1-\alpha)}\frac{d}{dx}\int_{-\infty}^{x}\frac{y(t)\,dt}{(x-t)^{\alpha}},\ \left(D_{-}^{\alpha}y\right)(x)=-\frac{1}{\Gamma(1-\alpha)}\frac{d}{dx}\int_{x}^{\infty}\frac{y(t)\,dt}{(t-x)^{\alpha}}.$$

В частном случае, при $\alpha=n\in N$ имеем

$$\left(D_{+}^{0}y\right)(x)=\left(D_{-}^{0}y\right)(x)=y(x),$$

$$\left(D_{+}^{n}y\right)(x)=y^{(n)}(x),\ \left(D_{-}^{n}y\right)(x)=(-1)^{n}y^{(n)}(x).$$

Интегральное преобразование Фурье дробных производных Римана-Лиувилля вычисляется с помощью следующих соотношений

$$F\left[D_{\pm}^{\alpha}f(x)\right](\xi)=(\mp i\xi)^{\alpha}F\left[f(x)\right](\xi),\ (\mp i\xi)^{\alpha}=|\xi|^{\alpha}e^{\mp\alpha\pi i\,\mathrm{sgn}(\xi)/2},\quad(13)$$

где $\mathrm{Re}(\alpha)>0$, $F\left[f(x)\right](\xi)=\int\limits_{-\infty}^{+\infty} f(x)e^{i\xi x}dx$.

3.2 Дробные производные в смысле Капуто и их связь с производными Римана-Лиувилля.

Дробные производные Капуто $\left({}^{C}D_{a+}^{\alpha}f\right)(x)$ и $\left({}^{C}D_{b-}^{\alpha}f\right)(x)$ выражаются через соответствующие производные Римана-Лиувилля следующим образом:

$$\left({}^{C}D_{a+}^{\alpha}f\right)(x)=\left(D_{a+}^{\alpha}\left[f(t)-\sum_{k=0}^{n-1}\frac{f^{(k)}(a)}{k!}(t-a)^{k}\right]\right)(x),$$

$$\left({}^{C}D_{b-}^{\alpha}f\right)(x)=\left(D_{b-}^{\alpha}\left[f(t)-\sum_{k=0}^{n-1}\frac{f^{(k)}(b)}{k!}(b-t)^{k}\right]\right)(x),$$

где $n=[\mathrm{Re}\,\alpha]+1$ для $\alpha\notin N$, $n=\alpha$ для $\alpha\in N$.

Важное прикладное значение в теории дифференциальных уравнений имеет дробная производная Капуто, определенная на положительной полуоси:

$$\left({}^{C}D_{0+}^{\alpha}f\right)(x)=D_{C}^{\alpha}f(x)=$$

$$=\frac{1}{\Gamma(n-\alpha)}\int_{0}^{x}(x-t)^{n-\alpha-1}\frac{d^{n}}{dt^{n}}f(t)dt,\ n-1<\alpha<n. \qquad (14)$$

Для вычисления преобразования Лапласа от производной (14) достаточно знать начальные значения функции $f(x)$ и её производных целого порядка k ($k=1,2,...,n-1$)

$$L\left[D_{C}^{\alpha}f(x)\right](p)=p^{\alpha}L\left[f(x)\right](p)-\sum_{k=0}^{n-1}f^{(k)}\left(0^{+}\right)p^{\alpha-1-k},\ n-1<\alpha<n. \qquad (15)$$

Функция Миттаг-Леффлера $E_{\alpha}\left[\lambda(x-a)^{\alpha}\right]$ [21] является инвариантом для оператора ${}^{C}D_{a+}^{\alpha}$, то есть при $\alpha>0$, $a\in R^{1}$ и $\lambda\in C$ имеет место равенство

$$\left({}^{C}D_{a+}^{\alpha}E_{\alpha}\left[\lambda(t-a)^{\alpha}\right]\right)(x)=\lambda E_{\alpha}\left[\lambda(x-a)^{\alpha}\right].$$

3.3 Операторы дробного интегрирования и дифференцирования Рисса

Для $\alpha\in C\setminus\{0\}$ и "достаточно хороших" функций $f(x)=f(x_1,...,x_n)$ дробная степень оператора Лапласа $\Delta=\sum_{j=1}^{n}\frac{\partial^{2}}{\partial x_{j}^{2}}$ определяется с помощью преобразования Фурье [22]

$$\left(-\Delta\right)^{-\alpha/2} f\left(x\right) = F^{-1}\left[\left|x\right|^{-\alpha}\right] F\left[f\right] = \begin{cases} I^{\alpha} f, \\ \mathbf{D}^{-\alpha} f, \ \mathrm{Re}\left(\alpha\right) > 0. \end{cases} \tag{16}$$

Операторы I^{α} и $\mathbf{D}^{\alpha}$ в правой части (16) называются операторами дробного интегрирования и дифференцирования Рисса.

Из (16) следует, что $F\left[\mathbf{D}^{\alpha} y\left(x\right)\right]\left(\xi\right) = \left|\xi\right|^{\alpha} F\left[y\left(x\right)\right]\left(\xi\right)$.

Если $x \in R^1$, то иногда производную Рисса представляют в виде [2]

$$\frac{d^{\alpha} f\left(x\right)}{d\left|x\right|^{\alpha}} = -\frac{\sin\left(\alpha\pi/2\right)}{\sin\left(\alpha\pi\right)}\left[D_{+}^{\alpha} f\left(x\right) + D_{-}^{\alpha} f\left(x\right)\right],$$

и

$$F\left[d^{\alpha} f\left(x\right) \big/ d\left|x\right|^{\alpha}\right]\left(\xi\right) = -\left|\xi\right|^{\alpha} F\left[f\left(x\right)\right]\left(\xi\right). \tag{17}$$

4. Используемый метод решения задачи термоупругости и его программная реализация

Применим к уравнениям (6), (7) интегральные преобразования Фурье и Лапласа с учетом начальных условий (8). Принимая во внимание формулы (13), (15) и (17), получим

$$\left\{p^2 - E_0\left(-i\omega\right)\left(\left(-i\omega\right)^{\gamma} - \left(i\omega\right)^{\gamma}\right)\right\} U\left(\omega, p\right) = \left(i\omega B\right) Q\left(\omega, p\right), \tag{18}$$

$$\left(\left|\omega\right|^{\beta} + p\right) Q = 1, \ U = LF[u], \ Q = LF[T]. \tag{19}$$

Выполним обращение интегральных преобразований в уравнениях (18), (19), после чего будем иметь

$$T\left(x,t\right) = \frac{H\left(t\right)}{\pi}\int_0^{\infty} e^{-\omega^{\beta} t} \cos\omega x d\omega, \tag{20}$$

$$u\left(x,t\right) = \frac{B}{\pi}\int_0^{\infty} \frac{\omega}{A^2\omega^{\gamma+1} + \omega^{2\beta}}\left\{e^{-\omega^{\beta} t} - \cos A\omega^{\frac{\gamma+1}{2}} t + \frac{\omega^{\beta}}{A\omega^{\gamma+1}} \sin A\omega^{\frac{\gamma+1}{2}} t\right\} \sin\omega x d\omega \tag{21}$$

где $H\left(t\right)$ – функция Хевисайда, $A = \sqrt{2E_0 \sin\dfrac{\gamma\pi}{2}}$.

На рис. 1-6 представлены результаты расчетов перемещения u и температуры T для различных значений показателей β и γ ($B = E_0 = 1$, $t_1 = 1c$, $t_2 = 3c$, $t_3 = 5c$, $t_4 = 7c$).

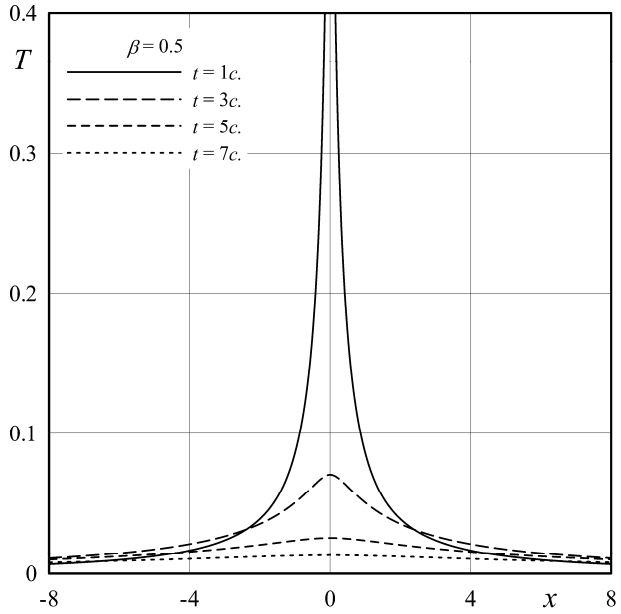


Рис. 1 Распределение температуры в стержне для различных моментов времени при $\beta = 0.5$.

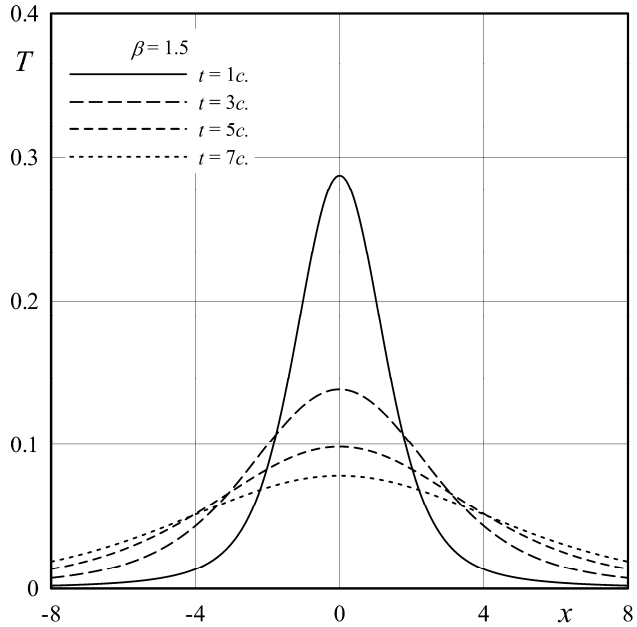


Рис. 2 Распределение температуры в стержне для различных моментов времени при $\beta = 1.5$.

Графики на рис. 1-2 характеризуют аномальный процесс теплопроводности, соответствующий явлению супердиффузии ($1/\beta > 1/2$), причем при $\beta = 0.5$ тепло переносится "быстрее", чем при $\beta = 1.5$.

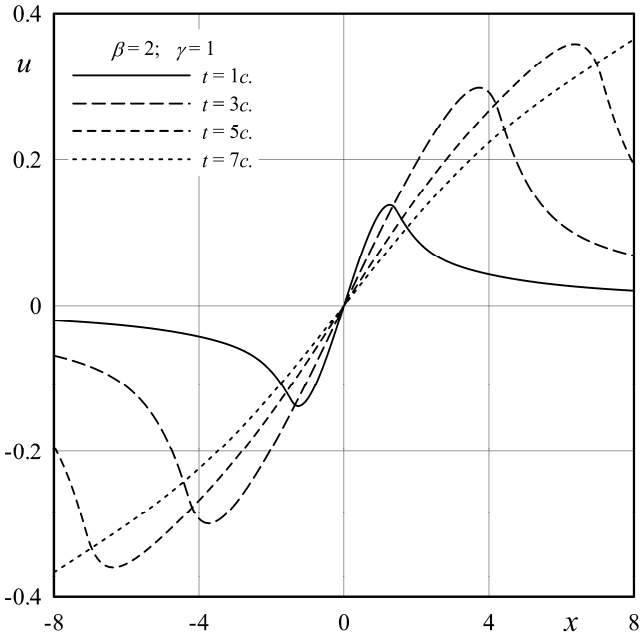


Рис. 3 Смещения в точках стержня для разных моментов времени при . $\beta = 2.,\ \gamma = 1$ *(классическая термоупругость).*

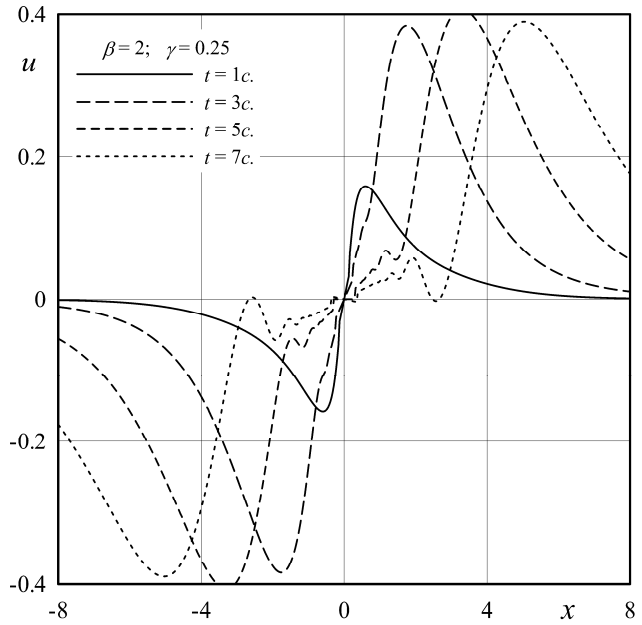


Рис. 4 Смещения в точках стержня для разных моментов времени при $\beta = 2$, $\gamma = 0.25$.

Сравнивая графики на рис. 3-4, видим, что нелокальность упругих взаимодействий приводит к колебаниям функции смещений вблизи нуля. Следует заметить, что в данном примере напряжения появляются только вследствие теплового расширения, однако дробный показатель γ оказывает

влияние на решение задачи даже при рассмотрении классической теплопроводности ($\beta = 2$).

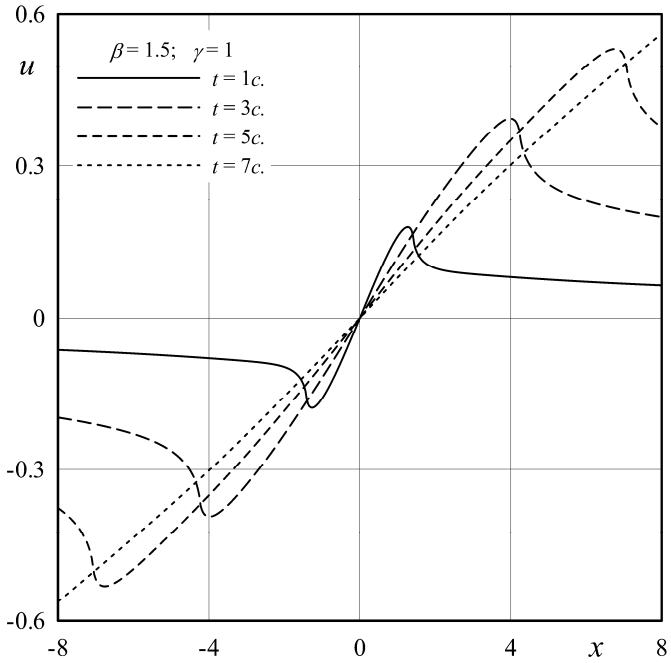


Рис. 5 Смещения в точках стержня для разных моментов времени при $\beta = 1.5$, $\gamma = 1$.

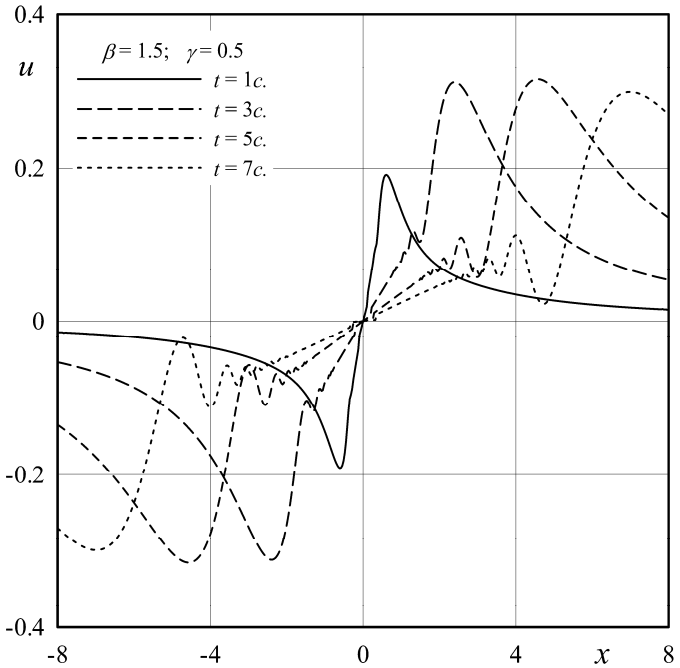


Рис.6 Смещения в точках стержня для разных моментов времени при $\beta = 1.5$, $\gamma = 0.5$.

Учет пространственной нелокальности тепловых и упругих взаимодействий одновременно (рис. 6) приводит к существенным изменениям результирующего поля перемещений.

5. Выводы по результатам и направления дальнейших исследований

В данной работе решена одномерная пространственно-нелокальная задача термоупругости, полученная на основании модели Атанковича [1]. При решении применялся метод интегральных преобразований. Результирующее перемещение и температура получены в виде интегралов (20), (21), представлены результаты численных расчетов. Учет пространственной нелокальности тепловых и упругих взаимодействий одновременно приводит к существенным изменениям результирующего поля перемещений. Дальнейшим развитием данной работы будет рассмотрение двумерной модели нелокальной термоупругости Атанковича.

ЛИТЕРАТУРА

1. Atanackovic T.M., Stankovic B. Generalized wave equation in nonlocal elasticity. // Acta Mechanica. — 2009. – 208. – P. 1-10.
2. Povstenko Y.Z. Thermoelasticity which uses fractional heat conduction equation. // Мат. методи та фіз.-мех. поля. — 2008. – 51, №2. – С. 239-246.
3. Владимиров В.С. Уравнения математической физики. — М.: Наука, 1971. – 512 с.
4. R. Metzler, J. Klafter. The random walk’s guide to anomalous diffusion: a fractional dynamics approach // Phys. Rep. – 2000. – 339, p. 1-77.
5. S. Lepri, R. Livi, A. Politi. Anomalous heat conduction, in book: Anomalous transport: foundations and applications edited by R. Klages, G. Radons, I.M. Sokolov, — Wiley, VCH(Berlin), 2008, – 584 p.
6. А.М. Нахушев. Дробное исчисление и его применение. — М.: Физматлит, 2003. — 272 с.
7. Учайкин В.В. Автомодельная аномальная диффузия и устойчивые законы. // УФН, – 2003. – т. 173, №8, – С. 847-876.
8. Смирнов Б.М. Энергетические процессы в макроскопических фрактальных структурах. // УФН. – 1993. – т. 161, №6. – С. 171-200.
9. Бейбалаев В.Д. Математическая модель теплопереноса в средах с фрактальной структурой. // Математическое моделирование. — 2009. – 21: 5. – С. 55-62.
10. Boyadjiev L., Scherer R. Fractional extensions of the temperature field problem in oil strata. // Kuwait. J. Sci. Eng. — 2004. – 31 (2). – p. 15-32.
11. Benson D.A., Wheatcraft S.W., Meerschaert M.M. The fractional-order equation of Levy motion, Water Resour. Res. — 2000. – 36. – p. 1413-1424.
12. Benson D.A., Schumer R., Meerschaert M.M., Wheatcraft S.W. Fractional dispersion, Levy motion, and the MADE tracer tests // Transp. Por. Med. — 2003. – 42. – p. 211-240.
13. Scalas E., Gorenflo R., Mainardi F. Uncoupled continuous-time random walks: solutions and limiting behaviour of the master equation, Phys. Rev. E. — 2004. – 692. – 011107.

14. Ostoja-Starzewski M. Towards thermoelasticity of fractal media. // Journal of thermal stresses. — 2007. – 30. – p. 889-896.
15. Sherief H.H., El-Sayed A.M.A., Abd El-Latief A.M. Fractional order theory of thermoelasticity // International Journal of Solids and Structures, 47 (2010), pp. 269-275.
16. Youssef H. M. Theory fractional order generalized thermoelasticity // J. Heat Transfer (ASME), 132(6), 2010, doi: 10.1115/3.4000705.
17. Povstenko Y.Z. Theory of thermoelasticity based on the space-time-fractional heat conduction equation. // Phys. Scr. — 2009. – T 136. – p. 014017-014023.
18. Povstenko Y.Z. Fundamental solutions to central symmetric problems for fractional heat conduction equation and associated thermal stresses. // Journal of thermal stresses. — 2008. – vol. 31, issue 2. – p. 127-148.
19. Povstenko Y. Time-fractional radial heat conduction in a cylinder and associated thermal stresses // Arch. Appl. Mech., doi: 10.1007/s00419-011-0560-x.
20. Абрамовиц М., Стиган И. Справочник по специальным функциям с формулами, графиками и математическими таблицами. – Москва: Наука, 1979. – 832 с.
21. Бейтмен Г., Эрдейи А. Высшие трансцендентные функции. Эллиптические и автоморфные функции. Функции Ламе и Матье. (Серия: «Справочная математическая библитотека»). — М.: Наука, 1967. – 300 с.
22. Kilbas A., Srivastava H.M., Trujillo J.J. Theory and applications of fractional differential equations. — North-Holland, Mathematics studies 204, 2006, – 524 p.

Надійшла – 10.06.2014.

УДК 004.655

Мультимножественная табличная алгебра: дополнительные операции

Д. Б. Буй[*], И. Н. Глушко[**]

[*] *Киевский национальный университет имени Тараса Шевченко, Киев, Украина*

[**] *Нежинский государственный университет имени Николая Гоголя, Нежин, Украина*

Сигнатура мультимножественной табличной алгебры пополнена новыми операциями: операциями внутренних и внешних соединений, операцией полусоединения, агрегатными операциями. Задана формальная математическая семантика указанных операций. Для задания внешних соединений введен особый элемент универсального домена NULL.

Ключевые слова: *реляционные базы данных, мультимножественная табличная алгебра, расширенная мультимножественная табличная алгебра.*

Сигнатура мультимножинної табличної алгебри поповнена новими операціями: операціями внутрішніх і зовнішніх з'єднань, операцією напівз'єднання, агрегатними операціями. Задано формальну математичну семантику вказаних операцій. Для задання зовнішніх з'єднань введено особливий елемент універсального домену NULL

Ключові слова: *реляційні бази даних, мультимножинна таблична алгебра, розширена мультимножинна таблична алгебра.*

The signature of multiset table algebra is filled up with new operations such as inner and outer joins, semijoin and aggregate operations. A formal mathematical semantics of these operations is defined. The special element NULL is introduced into the universal domain to define of outer join.

Key words: *relation databases, multiset table algebra, extending multiset table algebra.*

1. Общая постановка задачи и её актуальность

Реляционная модель данных в настоящее время широко используется как в научных исследованиях в базах данных, так и на практике. Данная модель основана на множествах кортежей, то есть не позволяет дубликаты кортежей в отношении [1]. Однако многие языки, ориентированные на работу с базами данных, требуют реляционную модель данных с мультимножественной семантикой (multiset semantics). Это предполагает понимание таблиц как мультимножеств, т.е. совокупностей с дубликатами. Вопросу использования мультимножеств в базах данных уделяли внимание G. Lamperti, M. Melchiori, M. Zanella [2], Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом [3], A. Silbeschatz, H. Korth, S. Sudarshan [4], а также отечественные ученые Д.Б. Буй, С.А. Поляков, Ю.Й. Брона, В.Н. Редько [5]. Обзор литературы об использовании мультимножеств в базах данных проведен в статье [6], которая насчитывает 9 источников по данной теме.

Вместе с тем, этот вопрос требует уточнения и расширения, поскольку ни в одной из указанных работ не уделяется достаточное внимание операциям внутренних и внешних соединений, операции полусоединения, внешним мультимножественным операциям, а также агрегатным операциям над таблицами мультимножественной табличной алгебры.

2. Основные понятия теории мультимножеств

Приведем основные понятия мультимножеств в терминах работ [5, 7]. Зафиксируем множество U. Под мультимножеством α с основой U понимаем отображение вида $\alpha : U \to \{1,2,...\}$. Пусть D – универсум элементов основ мультимножеств, тогда булеан $P(D)$ – универсум основ мультимножеств.

Под характеристической функцией мультимножества α понимаем функцию вида $\chi_\alpha : D \to \{0,1,2,...\}$, значение которой задается кусочной схемой :

$$\chi_\alpha(d) = \begin{cases} \alpha(d), \textit{если } d \in \mathrm{dom}\,\alpha, \\ 0, \textit{иначе}; \end{cases}$$

для всех $d \in D$, где $\mathrm{dom}\,\alpha$ – область определения мультимножества α, т.е. его основа.

Мультимножество называется пустым и обозначается как $\varnothing_m$, если его основа – пустое множество.

Мультимножества, областью значений которых является пустое множество или одноэлементное множество вида $\{1\}$ называются 1-мультимножествами. Эти мультимножества есть аналогами обычных множеств.

Договоримся мультимножество α с основой $\{d_1,...,d_k\}$ записывать как $\{d_1^{n_1},...,d_k^{n_k}\}$, где n_i – количество дубликатов (экземпляров) элемента d_i в мультимножестве α, т.е. $n_i = \alpha(d_i)$, $i = 1,...,k$.

Под рангом конечного мультимножества α понимаем количество дубликатов элементов его основы $\|\alpha\| = \sum_{d \in \mathrm{dom}\,\alpha} \alpha(d)$; при этом $\|\varnothing_m\| = 0$.

Скажем, что мультимножество β включается в мультимножество α ($\beta \preceq \alpha$), если: $\beta \preceq \alpha \Leftrightarrow U_\beta \subseteq U_\alpha \;\&\; \forall d\big(d \in U_\beta \Rightarrow \beta(d) \le \alpha(d)\big)$. Здесь U_α и U_β основы мультимножеств α и β соответственно.

Если $\beta \preceq \alpha$, то мультимножество β называется подмультимножеством мультимножества α, а мультимножество α – надмультимножеством мультимножества β.

В роботе [5] операции над мультимножествами определены в терминах характеристических функций. Авторы определяют операции объединения $\bigcup_1$, пересечения $\bigcap_1$, разности $\setminus_1$ мультимножеств, которые строят 1-мультимножества, основы которых получаются соответственноc теоретико-множественными объединением, пересечением и разницей основ мультимножеств-аргументов. Кроме того, вводятся операции объединения $\bigcup_{All}$, пересечения $\bigcap_{All}$, разности $\setminus_{All}$ мультимножеств, которые строят мультимножества общего вида. Также задано операцию декартового соединения мультимножеств $\otimes$ и операцию $Dist(\alpha)$, которая строит 1-мультимножество, основа которого совпадает с основой исходного мультимножества. Наконец, в этой работе вводится аналог полного образа (множества относительно функции) для мультимножеств.

3. Мультимножественная табличная алгебра

Рассмотрим два множества: $\boldsymbol{A}$ – множество атрибутов (имен) и $\boldsymbol{D}$ – универсальный домен (множество денотатов). Произвольное (конечное) множество атрибутов $R \subseteq \boldsymbol{A}$ назовём схемой. Под строкой схемы R понимаем именное множество на паре $\boldsymbol{A}$ и $\boldsymbol{D}$ [5], проекция которого по первой компоненте совпадает с R, т. е. строка схемы R – это функция вида $s : R \to \boldsymbol{D}$. Множество всех строк схемы R обозначим $S(R)$, а множество всех строк – S.

Под таблицей схемы R понимаем пару $\langle \psi, R \rangle$, где первая компонента ψ – это произвольное мультимножество, основой которого $\Theta(\psi)$ является произвольное множество, в частности, бесконечное, строк схемы R, а вторая компонента R – схема (таблицы).

Под мультимножественной табличной алгеброй понимаем алгебру $\langle \Psi, \Omega_{P,\Xi} \rangle$, где $\Psi = \bigcup\limits_{R \subseteq \boldsymbol{A}} \Psi(R)$ – множество всех таблиц, $\Psi(R)$ – множество всех таблиц схемы R,

$$\Omega_{P,\Xi} = \left\{ \bigcup\nolimits_{All}^{R}, \bigcap\nolimits_{All}^{R}, \setminus_{All}^{R}, \sigma_{p,R}, \pi_{X,R}, \underset{R_1,R_2}{\otimes}, Rt_{\xi,R,\sim_R} \right\}_{X,R,R_1,R_2 \subseteq \boldsymbol{A}}^{p \in P, \xi \in \Xi}$$ – сигнатура; P, Ξ – множества параметров. Операции мультимножественной табличной алгебры задано в [8].

Пополним сигнатуру мультимножественной табличной алгебры новыми операциями: операциям внутренних и внешних соединений, операцией полусоединения, агрегатными операциями.

4. Операции внутреннего соединения

Под декартовым соединением Cj (Cartesian Join) таблиц схем R_1 и R_2, причем $R_1 \cap R_2 = \varnothing$, понимаем бинарную параметрическую операцию вида $\underset{R_1,R_2}{Cj} : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2)$, $\langle \psi_1, R_1 \rangle \underset{R_1,R_2}{Cj} \langle \psi_2, R_2 \rangle = \langle \psi', R_1 \cup R_2 \rangle$, где $\langle \psi_1, R_1 \rangle \in \Psi(R_1)$, $\langle \psi_2, R_2 \rangle \in \Psi(R_2)$. Основой мультимножества ψ' является множество строк $\Theta(\psi') = \{ s \mid \exists s_1 \exists s_2 (s_1 \in \Theta(\psi_1) \wedge s_2 \in \Theta(\psi_2) \wedge s = s_1 \cup s_2) \}$. Количество дубликатов определяется так: $Occ(s, \psi') = Occ(s_1, \psi_1) \cdot Occ(s_2, \psi_2)$, где $s \in \Theta(\psi')$ и $s = s_1 \cup s_2$ (очевидно, что это разложение строки s единственно).

Ниже используется бинарное отношение совместимости строк, которое вводится так: $s_1 \approx s_2 \overset{def}{\Leftrightarrow} s_1 \mid R' = s_2 \mid R'$, де $R' = R_1 \cap R_2$, а R_1, R_2 – схемы строк s_1, s_2 соответственно [5]. Основное свойство этого отношения состоит в следующем: $s_1 \cup s_2 \in S(R_1 \cup R_2) \Leftrightarrow s_1 \approx s_2$ [5].

Под внутренним естественным соединением (Inner Natural Join) таблиц схем R_1 и R_2 понимаем бинарную параметрическую операцию $\underset{R_1,R_2}{\otimes}$, значениями

которой являются таблицы схемы $R_1 \cup R_2$, которые, говоря содержательно, содержат все объединения совместных строк исходных таблиц. Таким образом, $\underset{R_1,R_2}{\otimes} : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2)$, $\langle \psi_1, R_1 \rangle \underset{R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle = \langle \psi', R_1 \cup R_2 \rangle$, где $\langle \psi_1, R_1 \rangle \in \Psi(R_1)$, $\langle \psi_2, R_2 \rangle \in \Psi(R_2)$. Основой мультимножества ψ' является множество строк вида

$$\Theta(\psi') = \{ s \mid \exists s_1 \exists s_2 (s_1 \in \Theta(\psi_1) \wedge s_2 \in \Theta(\psi_2) \wedge s_1 \approx s_2 \wedge s = s_1 \cup s_2) \}.$$

Количество дубликатов определяется, как и ранее, так: $Occ(s, \psi') = Occ(s_1, \psi_1) \cdot Occ(s_2, \psi_2)$, где $s \in \Theta(\psi')$ и $s = s_1 \cup s_2$ (как и ранее, приведенное разложение единственно).

Под внутренним соединением по атрибутам $A_1, ..., A_n$ (Inner Join using $A_1, ..., A_n$), причем все $A_1, ..., A_n$ попарно различны, $n \geq 1$, таблиц схем R_1 и R_2, где $R_1 \cap R_2 \supseteq \{A_1, ..., A_n\}$ (подразумевается, что общие атрибуты, отличающиеся от атрибутов $A_1, ..., A_n$, перед соединением переименовываются) понимаем бинарную параметрическую операцию вида

$$\underset{A_1,...,A_n,R_1,R_2}{\otimes} : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2),$$

причем $\langle \psi_1, R_1 \rangle \underset{A_1,...,A_n,R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle = \langle \psi', R_1 \cup R_2 \rangle$, где $\langle \psi_1, R_1 \rangle \in \Psi(R_1)$, $\langle \psi_2, R_2 \rangle \in \Psi(R_2)$. Основой мультимножества ψ' является множество строк

$$\Theta(\psi') = \left\{ s \mid \exists s_1 \exists s_2 \left(s_1 \in \Theta(\psi_1) \wedge s_2 \in \Theta(\psi_2) \wedge \bigwedge_{i=1}^{n} s_1(A_i) = s_2(A_i) \wedge s = s_1 \cup s_2 \right) \right\}.$$

Количество дубликатов определяется, как и ранее, так: $Occ(s, \psi') = Occ(s_1, \psi_1) \cdot Occ(s_2, \psi_2)$, где $s \in \Theta(\psi')$ и $s = s_1 \cup s_2$ (как и ранее, приведенное разложение единственно).

Отметим, если таблицы-аргументы имеют еще и другие общие атрибуты, отличающиеся от атрибутов $A_1, ..., A_n$, то перед соединением их нужно переименовать.

Пусть $p : S \times S \xrightarrow{\sim} \{true, false\}$ – вообще говоря частичный бинарный предикат на множестве всех строк S, такой, что выполняется импликация $\forall s_1 \forall s_2 \left(\langle s_1, s_2 \rangle \in \operatorname{dom} p \wedge p(s_1, s_2) = true \Rightarrow s_1 \approx s_2 \right)$.

Под внутренним соединением по предикату p (Inner Join on p) таблиц схем R_1 и R_2 понимаем частичную бинарную параметрическую операцию вида $\underset{p,R_1,R_2}{\otimes} : \Psi(R_1) \times \Psi(R_2) \xrightarrow{\sim} \Psi(R_1 \cup R_2)$, $\operatorname{dom} \underset{p,R_1,R_2}{\otimes} = \{ \langle \langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle \rangle \mid \Theta(\psi_1) \times \Theta(\psi_2) \subseteq \operatorname{dom} p \}$, $\langle \psi_1, R_1 \rangle \underset{p,R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle = \langle \psi', R_1 \cup R_2 \rangle$. Основой

мультимножества ψ' является множество строк $\Theta(\psi') = \{s \mid \exists s_1 \exists s_2 (s_1 \in \Theta(\psi_1) \wedge s_2 \in \Theta(\psi_2) \wedge p(s_1, s_2) \simeq true \wedge s = s_1 \bigcup s_2)\}$. Количество дубликатов определяется, как и ранее, так: $Occ(s, \psi') = Occ(s_1, \psi_1) \cdot Occ(s_2, \psi_2)$, где $s \in \Theta(\psi')$ и $s = s_1 \bigcup s_2$. Выше предполагалось, что пара таблиц $\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle$ принадлежит указанной области определенности.

Отметим следующий очевидный факт. Операция естественного соединения $\underset{R_1,R_2}{\otimes}$ является расширением произвольной другой операции соединения в следующем смысле: $\langle \psi_1, R_1 \rangle \underset{R_1,R_2}{Cj} \langle \psi_2, R_2 \rangle = \langle \psi_1, R_1 \rangle \underset{R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle$,

$\langle \psi_1, R_1 \rangle \underset{A_1,\ldots,A_n,R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle = \langle \psi_1, R_1 \rangle \underset{R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle$,

$\left(\langle \psi_1, R_1 \rangle \underset{p,R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle \right)_1 \preceq \left(\langle \psi_1, R_1 \rangle \underset{R_1,R_2}{\otimes} \langle \psi_2, R_2 \rangle \right)_1$ [1],

при условии, что значения операций в левых частях этих двух равенств и включения определены.

Под операцией полусоединения (Semijoin) двух таблиц схем R_1 и R_2 понимаем бинарную параметрическую операцию $\square_{R_1,R_2}$, значением которой является таблица схемы R_1, которая содержит те строки первой таблицы, которые входят в (естественное) соединение таблиц-аргументов.

Следовательно, $\square_{R_1,R_2} : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1)$, $\langle \psi_1, R_1 \rangle \square_{R_1,R_2} \langle \psi_2, R_2 \rangle = \langle \psi', R_1 \rangle$, где $\langle \psi_1, R_1 \rangle \in \Psi(R_1)$, $\langle \psi_2, R_2 \rangle \in \Psi(R_2)$. Основой мультимножества ψ' является множество строк $\Theta(\psi') = \{s_1 \mid s_1 \in \Theta(\psi_1) \wedge \exists s_2 (s_2 \in \Theta(\psi_2) \wedge s_1 \approx s_2)\}$. Количество дубликатов определяется так: $Occ(s, \psi') = Occ(s, \psi_1)$, где $s \in \Theta(\psi')$.

5. Операции внешнего соединения

При применении операций внутреннего соединения возможна потеря информации, поскольку строки одной таблицы, которые не соединяются со строками другой таблицы, не будут включены в результирующую таблицу. В тех случаях, когда необходимо учесть строки таблиц-аргументов, которые не попали в результат исходного внутреннего соединения, используют операции внешнего соединения.

Для обозначения отсутствующих значений атрибутов строк результирующей таблицы используем особый элемент универсального домена $NULL$. Обозначим как $s_{R,NULL}$ константную строку схемы R вида $s_{R,NULL} : R \to \{NULL\}$, присваивающую всем атрибутам своей схемы значения $NULL$.

---

[1] Запись $(\langle \psi, R \rangle)_1$ обозначает первую компоненту пары $\langle \psi, R \rangle$, то есть мультимножество ψ.

Используем логическую схему определения операций внешнего соединения из [5], следуя которой четыре операции внешнего соединения вводятся как операции, подчиненные одной операции внутреннего соединения.

Пусть $\varphi : \Psi(R_1) \times \Psi(R_2) \xrightarrow{\sim} \Psi(R_1 \cup R_2)$ – некоторая частичная бинарная операция на множестве таблиц, причем выполняется включение $\left(\varphi\left(\langle \psi_1, R_1\rangle, \langle \psi_2, R_2\rangle\right)\right)_1 \preceq \left(\langle \psi_1, R_1\rangle \underset{R_1,R_2}{\otimes} \langle \psi_2, R_2\rangle\right)_1$ для всех $\left\langle \langle \psi_1, R_1\rangle, \langle \psi_2, R_2\rangle \right\rangle \in \mathrm{dom}\,\varphi$.

Отметим, что операции внутреннего соединения $\underset{R_1,R_2}{Cj}$, $\underset{R_1,R_2}{\otimes}$, $\underset{A_1,\ldots,A_n,R_1,R_2}{\otimes}$, $\underset{p,R_1,R_2}{\otimes}$ именно такие.

Зафиксируем таблицы $\langle \psi_1, R_1\rangle$, $\langle \psi_2, R_2\rangle$ из области определенности операции φ. Тогда таблица $\langle \psi_1, R_1\rangle$ предполагает следующее представление: $\langle \psi_1, R_1\rangle = \left\langle \psi_1 \underset{\varphi}{\cap} \psi_2, R_1 \right\rangle \cup_{All}^{R_1} \left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle$, где $\left\langle \psi_1 \underset{\varphi}{\cap} \psi_2, R_1 \right\rangle = \langle \psi', R_1\rangle$, основой мультимножества ψ' является множество строк

$$\Theta(\psi') = \{s_1 \mid s_1 \in \Theta(\psi_1) \wedge \exists s_2 \left(s_2 \in \Theta(\psi_2) \wedge s_1 \cup s_2 \in \Theta\left(\left(\varphi\left(\langle \psi_1, R_1\rangle, \langle \psi_2, R_2\rangle\right)\right)_1\right)\right)\},$$

а количество дубликатов $Occ(s_1, \psi') = Occ(s_1, \psi_1)$, $s_1 \in \Theta(\psi')$ и $\left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle = \langle \psi'', R_1\rangle$, основой мультимножества ψ'' является множество строк

$$\Theta(\psi'') = \{s_1 \mid s_1 \in \Theta(\psi_1) \wedge \forall s_2 \left(s_2 \in \Theta(\psi_2) \Rightarrow s_1 \cup s_2 \notin \Theta\left(\left(\varphi\left(\langle \psi_1, R_1\rangle, \langle \psi_2, R_2\rangle\right)\right)_1\right)\right)\},$$

а количество дубликатов определяется так: $Occ(s_1, \psi'') = Occ(s_1, \psi_1)$, где $s_1 \in \Theta(\psi'')$.

Говоря содержательно, строки таблицы $\left\langle \psi_1 \underset{\varphi}{\cap} \psi_2, R_1 \right\rangle$ используются при формировании результата (внутреннего) соединения, а строки таблицы $\left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle$ – не используются. Аналогичное представление таблицы $\langle \psi_2, R_2\rangle$ получим, поменяв роли таблиц $\langle \psi_1, R_1\rangle$ и $\langle \psi_2, R_2\rangle$ в представлении таблицы $\langle \psi_1, R_1\rangle$.

Отметим, что если операция φ совпадает с операцией $\underset{R_1,R_2}{\otimes}$, то $\left\langle \psi_1 \underset{\varphi}{\cap} \psi_2, R_1 \right\rangle = \langle \psi_1, R_1\rangle \,\square_{R_1,R_2}\, \langle \psi_2, R_2\rangle$, т.е. таблица в левой части последнего

равенства получается в результате применения операции полусоединения к исходным таблицам.

Ниже для упрощения записи будем считать, что операции соединения имеют больший приоритет, чем операции объединения.

Определим четыре операции внешнего соединения (Outer Join), индуцированные одной операцией внутреннего соединения φ. Для этого рассмотрим следующие естественные соединения:

$$\left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle \underset{R_1, R_2 \setminus R_1}{\otimes} \left\langle \{s^1_{R_2 \setminus R_1, NULL}\}, R_2 \setminus R_1 \right\rangle = \left\langle \psi', R_1 \cup R_2 \right\rangle,$$

где основой мультимножества ψ' является множество строк $\Theta(\psi') = \{s_1 \cup s_{R_2 \setminus R_1, NULL} \mid s_1 \in \Theta(\psi_1 \underset{\varphi}{-} \psi_2)\}$, а количество дубликатов $Occ(s', \psi') = Occ(s_1, \psi_1 \underset{\varphi}{-} \psi_2)$, $s' \in \Theta(\psi')$ и $s' = s_1 \cup s_{R_2 \setminus R_1, NULL}$ (очевидно, что здесь и далее аналогичные представления строк единственны), а также

$$\left\langle \psi_2 \underset{\varphi}{-} \psi_1, R_2 \right\rangle \underset{R_2, R_1 \setminus R_2}{\otimes} \left\langle \{s^1_{R_1 \setminus R_2, NULL}\}, R_1 \setminus R_2 \right\rangle = \left\langle \psi'', R_1 \cup R_2 \right\rangle,$$

где основой мультимножества ψ'' является множество строк $\Theta(\psi'') = \{s_{R_1 \setminus R_2, NULL} \cup s_2 \mid s_2 \in \Theta(\psi_2 \underset{\varphi}{-} \psi_1)\}$, а количество дубликатов $Occ(s'', \psi'') = Occ(s_2, \psi_2 \underset{\varphi}{-} \psi_1)$, $s'' \in \Theta(\psi'')$ и $s'' = s_{R_1 \setminus R_2, NULL} \cup s_2$.

Выше верхний индекс 1 в записи $\left\langle \{s^1_{R_2 \setminus R_1, NULL}\}, R_2 \setminus R_1 \right\rangle$ указывает на то, что строка $s_{R_2 \setminus R_1, NULL}$ входит в исходную таблицу только один раз, т.е. $\left\langle \{s^1_{R_2 \setminus R_1, NULL}\}, R_2 \setminus R_1 \right\rangle$ – константная таблица схемы $R_2 \setminus R_1$, первая компонента которой $\{1\}$-мультимножество с основой $\{s_{R_2 \setminus R_1, NULL}\}$. Для таблицы $\left\langle \{s^1_{R_1 \setminus R_2, NULL}\}, R_1 \setminus R_2 \right\rangle$ полностью аналогично.

Под внешним левым соединением (Outer Left Join), индуцированным операцией φ, понимаем частичную бинарную операцию вида $\varphi_l : \Psi(R_1) \times \Psi(R_2) \xrightarrow{\sim} \Psi(R_1 \cup R_2)$, где $\mathrm{dom}\,\varphi_l = \mathrm{dom}\,\varphi$, $\varphi_l\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) =$

$$= \varphi\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) \cup^{R_1 \cup R_2}_{All} \left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle \underset{R_1, R_2 \setminus R_1}{\otimes} \left\langle \{s^1_{R_2 \setminus R_1, NULL}\}, R_2 \setminus R_1 \right\rangle.$$

Под внешним правым соединением (Outer Right Join), индуцированным операцией φ, понимаем частичную бинарную операцию вида

$\varphi_r : \Psi(R_1) \times \Psi(R_2) \xrightarrow{\sim} \Psi(R_1 \cup R_2)$, где $\mathrm{dom}\,\varphi_r = \mathrm{dom}\,\varphi$, $\varphi_r\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) =$

$= \varphi\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) \bigcup_{All}^{R_1 \cup R_2} \left\langle \psi_2 \underset{\varphi}{-} \psi_1, R_2 \right\rangle \underset{R_2, R_1 \setminus R_2}{\otimes} \left\langle \{s^1_{R_1 \setminus R_2, NULL}\}, R_1 \setminus R_2 \right\rangle$.

Под внешним полным соединением (Outer Full Join), индуцированным операцией φ, понимаем частичную бинарную операцию вида $\varphi_f : \Psi(R_1) \times \Psi(R_2) \xrightarrow{\sim} \Psi(R_1 \cup R_2)$, где $\mathrm{dom}\,\varphi_f = \mathrm{dom}\,\varphi$,

$$\varphi_f\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) = \varphi\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) \bigcup_{All}^{R_1 \cup R_2} \left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle \underset{R_1, R_2 \setminus R_1}{\otimes}$$

$$\underset{R_1, R_2 \setminus R_1}{\otimes} \left\langle \{s^1_{R_2 \setminus R_1, NULL}\}, R_2 \setminus R_1 \right\rangle \bigcup_{All}^{R_1 \cup R_2} \left\langle \psi_2 \underset{\varphi}{-} \psi_1, R_2 \right\rangle \underset{R_2, R_1 \setminus R_2}{\otimes}$$

$$\underset{R_2, R_1 \setminus R_2}{\otimes} \left\langle \{s^1_{R_1 \setminus R_2, NULL}\}, R_1 \setminus R_2 \right\rangle.$$

Под внешним соединением объединением (Outer Union Join), индуцированным операцией φ, понимаем частичную бинарную операцию вида $\varphi_\cup : \Psi(R_1) \times \Psi(R_2) \xrightarrow{\sim} \Psi(R_1 \cup R_2)$, где $\mathrm{dom}\,\varphi_\cup = \mathrm{dom}\,\varphi$, $\varphi_\cup\left(\langle \psi_1, R_1 \rangle, \langle \psi_2, R_2 \rangle\right) =$

$$= \left\langle \psi_1 \underset{\varphi}{-} \psi_2, R_1 \right\rangle \underset{R_1, R_2 \setminus R_1}{\otimes} \left\langle \{s^1_{R_2 \setminus R_1, NULL}\}, R_2 \setminus R_1 \right\rangle \bigcup_{All}^{R_1 \cup R_2} \left\langle \psi_2 \underset{\varphi}{-} \psi_1, R_2 \right\rangle \underset{R_2, R_1 \setminus R_2}{\otimes}$$

$$\underset{R_2, R_1 \setminus R_2}{\otimes} \left\langle \{s^1_{R_1 \setminus R_2, NULL}\}, R_1 \setminus R_2 \right\rangle.$$

6. Агрегатные операции

Широко используемыми (параметрическими) агрегатными операциями являются *Sum*, *Avg*, *Min*, *Max*, *Count*. Их аргументы – это конечные таблицы, а значения – одноатрибутные таблицы с одной строкой. Так, операция *Sum* рассчитывает сумму значений в соответствующем столбце заданной таблицы, при этом значения *NULL* игнорируются. Операция *Avg* определяет среднее арифметическое значений в соответствующем столбце заданной таблицы, при этом значения *NULL* игнорируются. Операции *Min* и *Max* находят наименьшее и наибольшее значения в соответствующем столбце заданной таблицы, при этом значения *NULL* также игнорируются. Операция *Count* определяет количество значений, отличных от *NULL*, в соответствующем столбце заданной таблицы. Операция *Count*(∗) определяет количество строк в заданной таблице.

Пусть $\langle \psi, R \rangle \in \Psi(R)$, где ψ – конечное мультимножество и $A \in R$. Обозначим как α_A – мультимножество, которое содержит все элементы с учетом дубликатов столбца с атрибутом A таблицы $\langle \psi, R \rangle$. Тогда $\Theta(\alpha_A) = \{d \mid \exists s(s \in \Theta(\psi) \wedge \langle A, d \rangle \in s)\} = \{d \mid \{\langle A, d \rangle\} \in \Theta\left(\left(\pi_{\{A\}, R}\left(\langle \psi, R \rangle\right)\right)_1\right)\}$ – аналог

активного домена атрибута A относительно таблицы [5, 9]. Количество дубликатов элемента основы $d \in \Theta(\alpha_A)$ в мультимножестве α_A определяется как $\alpha_A(d) = Occ(\{\langle A, d\rangle\}, (\pi_{\{A\},R}(\langle \psi, R\rangle))_1) = \sum_{\substack{s\in\Theta(\psi),\\ s(A)=d}} Occ(s,\psi)$. Пусть $2_m^{\boldsymbol{D}'} = \{\alpha \mid \Theta(\alpha) \in 2^{\boldsymbol{D}'}\}$ – семейство всех мультимножеств, основы которых являются конечными подмножествами множества $\boldsymbol{D}'$; здесь $\boldsymbol{D}' \subseteq \boldsymbol{D}$ – подмножество универсального домена.

Пусть Num – числовое подмножество универсального домена $\boldsymbol{D}$, замкнутое относительно сложения. Множество Num расширим включением особого элемента $NULL$, но операцию сложения на случай, когда хотя бы один из аргументов является $NULL$, расширять не будем.

Зададим агрегатные операции Sum, Avg, Min, Max, $Count$. Общая схема: на конечном мультимножестве определяются функции суммирования, взятия наименьшего и наибольшего значений, определения среднего арифметического и количества элементов, а затем эти функции переносятся на таблицы. Заметим, что функции суммирования и нахождения среднего арифметического определены на конечном числовом мультимножестве.

Под операцией агрегирования $Sum_{A,R}$ по атрибуту A (конечных) таблиц схемы R, $A \in R$, понимаем унарную параметрическую операцию вида $Sum_{A,R} : \Psi(R) \to \Psi(\{A\})$, $Sum_{A,R}(\langle \psi, R\rangle) = \left\langle \left\{ \left\langle A, Sum(\alpha_A)\right\rangle \right\}^1, \{A\} \right\rangle$, где $\langle \psi, R\rangle \in \Psi(R)$, а Sum – функция, которая возвращает сумму значений столбца с атрибутом A таблицы $\langle \psi, R\rangle$ (эти значения могут повторяться), которые отличаются от значения $NULL$, кроме того предполагается, что этот столбец содержит только данные числового типа. Следовательно, $Sum : 2_m^{Num} \to Num$,

$$Sum(\alpha_A) = \begin{cases} NULL, \text{ если } \Theta(\alpha_A) = \varnothing; \\ NULL, \text{ если } \Theta(\alpha_A) = \{NULL\}; \\ \sum\limits_{d\in\Theta(\alpha_A)\setminus\{NULL\}} d\alpha_A(d), \text{ если } \Theta(\alpha_A) \setminus \{NULL\} \neq \varnothing. \end{cases}$$

Как и раньше верхний индекс 1 в записи $\left\langle \left\{ \left\langle A, Sum(\alpha_A)\right\rangle \right\}^1, \{A\} \right\rangle$ указывает на то, что строка $\{\langle A, Sum(\alpha_A)\rangle\}$ входит в исходную таблицу только один раз, т.е. $\{\{\langle A, Sum(\alpha_A)\rangle\}^1\}$ – $\{1\}$-мультимножество.

Таким образом, имеем $Sum(\varnothing_m) = NULL$, $Sum(\{NULL^n\}) = NULL$, $Sum\left(\left\{d_1^{n_1}, \ldots, d_k^{n_k}\right\}\right) = \sum_{i=1}^{k} d_i n_i$, в предположении, что все элементы d_i отличны от элемента $NULL$.

Для случая пустой таблицы $\langle \psi_{\varnothing}, R \rangle$ операция агрегирования $Sum_{A,R}$ применяется так: $Sum_{A,R}(\langle \psi_{\varnothing}, R \rangle) = \left\langle \left\{ \langle A, NULL \rangle \right\}^{1}, \{A\} \right\rangle$, здесь $\psi_{\varnothing} = \varnothing_m$.

Пусть $\leq$ – линейный порядок на универсальном домене $\boldsymbol{D}$. Под операцией агрегирования $Min_{A,R}$ по атрибуту A (конечных) таблиц схемы R, $A \in R$, понимаем унарную параметрическую операцию вида $Min_{A,R} : \Psi(R) \to \Psi(\{A\})$, $Min_{A,R}(\langle \psi, R \rangle) = \left\langle \left\{ \langle A, Min(\alpha_A) \rangle \right\}^{1}, \{A\} \right\rangle$, где $\langle \psi, R \rangle \in \Psi(R)$, а Min – функция, которая возвращает наименьшее значение среди значений столбца с атрибутом A таблицы $\langle \psi, R \rangle$, которые отличаются от значения $NULL$, т.е. $Min : 2_m^{\boldsymbol{D}} \to \boldsymbol{D}$,

$$Min(\alpha_A) = \begin{cases} NULL, \text{если } \Theta(\alpha_A) = \varnothing; \\ NULL, \text{ если } \Theta(\alpha_A) = \{NULL\}; \\ \min\{d \mid d \in \Theta(\alpha_A) \setminus \{NULL\}\}, \text{если } \Theta(\alpha_A) \setminus \{NULL\} \neq \varnothing. \end{cases}$$

Таким образом, имеем $Min(\varnothing_m) = NULL$, $Min(\{NULL^n\}) = NULL$, $Min\left(\left\{ d_1^{n_1}, \ldots, d_k^{n_k} \right\}\right) = \min\{d_1, \ldots, d_k\}$, в предположении, что все элементы d_i, $i = \overline{1,k}$, отличны от элемента $NULL$.

Для случая пустой таблицы $\langle \psi_{\varnothing}, R \rangle$ операция агрегирования $Min_{A,R}$ применяется так: $Min_{A,R}(\langle \psi_{\varnothing}, R \rangle) = \left\langle \left\{ \langle A, NULL \rangle \right\}^{1}, \{A\} \right\rangle$.

Под операцией агрегирования $Max_{A,R}$ по атрибуту A (конечных) таблиц схемы R, $A \in R$, понимаем унарную параметрическую операцию вида $Max_{A,R} : \Psi(R) \to \Psi(\{A\})$, $Max_{A,R}(\langle \psi, R \rangle) = \left\langle \left\{ \langle A, Max(\alpha_A) \rangle \right\}^{1}, \{A\} \right\rangle$, где $\langle \psi, R \rangle \in \Psi(R)$, а Max – функция, которая возвращает наибольшее значение среди значений столбца с атрибутом A таблицы $\langle \psi, R \rangle$, которые отличаются от $NULL$, т.е. $Max : 2_m^{\boldsymbol{D}} \to \boldsymbol{D}$,

$$Max(\alpha_A) = \begin{cases} NULL, \text{если } \Theta(\alpha_A) = \varnothing; \\ NULL, \text{ если } \Theta(\alpha_A) = \{NULL\}; \\ \max\{d \mid d \in \Theta(\alpha_A) \setminus \{NULL\}\}, \text{если } \Theta(\alpha_A) \setminus \{NULL\} \neq \varnothing. \end{cases}$$

Таким образом, имеем $Max(\varnothing_m) = NULL$, $Max(\{NULL^n\}) = NULL$, $Max\left(\left\{ d_1^{n_1}, \ldots, d_k^{n_k} \right\}\right) = \max\{d_1, \ldots, d_k\}$, в предположении, что все элементы d_i, $i = \overline{1,k}$, отличны от элемента $NULL$.

Для случая пустой таблицы $\langle \psi_{\varnothing}, R \rangle$ операция агрегирования $Max_{A,R}$ применяется так: $Max_{A,R}(\langle \psi_{\varnothing}, R \rangle) = \left\langle \left\{ \left\langle A, NULL \right\rangle \right\}^{1}, \{A\} \right\rangle$.

Отметим, что функции *Min* и *Max* определяют наименьший или наибольший элементы основы мультимножества, которые отличаются от значения *NULL*, поэтому сравнимость особого элемента с остальными элементами универсального домена в данном случае несущественна. Это свойство важно при задании семантики фразы ORDER BYоператора запросов SELECT, отвечающей за упорядочение строк.

Под операцией агрегирования $Count_{A,R}$ по атрибуту A (конечных) таблиц схемы R, $A \in R$, понимаем унарную параметрическую операцию вида $Count_{A,R} : \Psi(R) \to \Psi(\{A\})$, $Count_{A,R}(\langle \psi, R \rangle) = \left\langle \left\{ \left\langle A, Count(\alpha_A) \right\rangle \right\}^{1}, \{A\} \right\rangle$, где $\langle \psi, R \rangle \in \Psi(R)$, а *Count* – функция, которая возвращает количество значений, которые отличаются от значения *NULL*, с учетом дубликатов в столбце с атрибутом A таблицы $\langle \psi, R \rangle$, т.е. $Count : 2_m^{\boldsymbol{D}} \to N$, $Count(\alpha_A) = \sum_{d \in \Theta(\alpha_A) \setminus \{NULL\}} \alpha_A(d)$; полагается по определению (и это естественно), что сумма пустого множества слагаемых равна нулю.

Таким образом, имеем $Count(\varnothing_m) = 0$, $Count(\{NULL^n\}) = 0$, $Count\left(\left\{d_1^{n_1}, \ldots, d_k^{n_k}\right\}\right) = n_1 + \ldots + n_k$, в предположении, что все элементы d_i, $i = \overline{1,k}$, отличны от элемента *NULL*.

Для случая пустой таблицы $\langle \psi_{\varnothing}, R \rangle$ операция агрегирования $Count_{A,R}$ применяется так: $Count_{A,R}(\langle \psi_{\varnothing}, R \rangle) = \left\langle \left\{ \left\langle A, 0 \right\rangle \right\}^{1}, \{A\} \right\rangle$.

Допустим, что числовое подмножество *Num* универсального домена замкнуто относительно (частичной операции) деления $/ : Num \times Num \xrightarrow{\sim} Num$. Доопределим операцию деления так, что когда первый аргумент равен *NULL*, то функция принимает значение *NULL*. Это связано с тем, что мы будем осуществлять суперпозиции и вместо первого аргумента подставлять значение функции *Sum*, а вместо второго – значение функции *Count*, учитывая, что функция *Count* не может дать *NULL*.

Под операцией агрегирования $Avg_{A,R}$ по атрибуту A (конечных) таблиц схемы R, $A \in R$, понимаем унарную параметрическую операцию вида $Avg_{A,R} : \Psi(R) \to \Psi(\{A\})$, $Avg_{A,R}(\langle \psi, R \rangle) = \left\langle \left\{ \left\langle A, Avg(\alpha_A) \right\rangle \right\}^{1}, \{A\} \right\rangle$, где $\langle \psi, R \rangle \in \Psi(R)$, а *Avg* – функция, которая возвращает среднее арифметическое значение элементов столбца с атрибутом A таблицы $\langle \psi, R \rangle$, которые

отличаются от значения $NULL$, с учетом дубликатов, т.е. $Avg: 2_m^{Num} \to Num$, $Avg(\alpha_A) = Sum(\alpha_A) / Count(\alpha_A)$.

Таким образом, из определений следуют равенства $Avg(\varnothing_m) = Sum(\varnothing_m) / Count(\varnothing_m) = NULL / 0 = NULL$,

$Avg(\{NULL^n\}) = Sum(\{NULL^n\}) / Count(\{NULL^n\}) = NULL / 0 = NULL$,

$Avg\left(\left\{d_1^{n_1}, \ldots, d_k^{n_k}\right\}\right) = Sum\left(\left\{d_1^{n_1}, \ldots, d_k^{n_k}\right\}\right) / Count\left(\left\{d_1^{n_1}, \ldots, d_k^{n_k}\right\}\right) = \frac{1}{(n_1 + \ldots + n_k)} \sum_{i=1}^{k} d_i n_i$, в предположении, что все элементы d_i, $i = \overline{1,k}$, отличны от элемента $NULL$.

Для случая пустой таблицы $\langle \psi_\varnothing, R \rangle$ операция агрегирования $Avg_{A,R}$ применяется так: $Avg_{A,R}\left(\langle \psi_\varnothing, R \rangle\right) = \left\langle \left\{\left\{\langle A, NULL \rangle\right\}\right\}, \{A\} \right\rangle$.

Под операцией агрегирования $Count_{A,R}(*)$ (конечных) таблиц схемы R понимаем унарную параметрическую операцию вида $Count_{A,R}(*): \Psi(R) \to \Psi(\{A\})$, $Count_{A,R}(*)\left(\langle \psi, R \rangle\right) = \left\langle \left\{\left\langle A, \|\psi\| \right\rangle^1\right\}, \{A\} \right\rangle$, где $\langle \psi, R \rangle \in \Psi(R)$, а $\|\psi\|$ – это введенный ранее ранг мультимножества ψ.

Для случая пустой таблицы $\langle \psi_\varnothing, R \rangle$ операция агрегирования $Count_{A,R}(*)$ применяется так: $Count_{A,R}(*)\left(\langle \psi_\varnothing, R \rangle\right) = \left\langle \left\{\left\langle A, \|\varnothing_m\| \right\rangle^1\right\}, \{A\} \right\rangle = \left\langle \left\{\left\langle A, 0 \right\rangle^1\right\}, \{A\} \right\rangle$.

7. Выводы по результатам и направления дальнейших исследований

Существует ряд прикладных задач, особенностью которых является множественность и повторяемость данных. Например, социологические опросы различных групп населения, вычисления на ДНК, уточнения таблиц с дубликатами строк и другие. Математическим уточнением совокупностей с повторениями выступают мультимножества. Естественно возникает потребность в расширении возможностей реляционных баз данных за счет использования мультимножеств.

В данной работе рассматривается мультимножественная табличная алгебра, сигнатура которой пополнена новыми операциями: операциям внутренних и внешних соединений, операцией полусоединения, агрегатными операциями. Для определения внешних операций введен особый элемент универсального домена $NULL$.

Следует также отметить, что параметром агрегатной операции может выступать не только отдельный атрибут, но и некоторая функция над строкой. Соответствующее уточнение не трудно провести на основе представленных в работе построений.

В дальнейшем планируется исследовать свойства операций мультимножественной табличной алгебры.

ЛИТЕРАТУРА

1. Codd E.F. Relational Completeness of Data Base Sublanguages / Codd E.F. // Data Base Systems. – New York: Prentice-Hall. – 1972. – P. 65-93.
2. Lamperti G. On Multisets in Database Systems / G. Lamperti, M. Melchiori, M. Zanella // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 147-215.
3. Гарсиа-Молина Г. Системы баз данных: [полный курс: пер. с англ.] / Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом. – Москва: "Вильямс", 2004. – 1088 с.
4. Silbeschatz A., Korth H., Sudarshan S. Database System Concepts. – McGraw-Hill, 2011. –1376 p.
5. Реляційні бази даних: табличні алгебри та SQL-подібні мови / В.Н. Редько, Ю.Й. Брона, Д.Б. Буй, С.А. Поляков. – Київ: Видавничий дім "Академперіодика", 2001. – 198 с.
6. Буй Д.Б. Сучасний стан теорії мультимножин / Д.Б. Буй, Ю.О. Богатирьова // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2010. – Вип. 1. – С. 51-58.
7. Петровский А.Б. Пространства множеств и мультимножеств / А.Б. Петровский. – Москва: "Едиториал УРСС", 2003. – 248 с.
8. Глушко І.М. Мультимножинна таблична алгебра / І.М. Глушко // Proceedings of the International Scientific Conference of Student and Young Scientists "Theoretical and Applied Aspects of Cybernetics" (TAAC'2011, Kyiv, February 21–25, 2011). – P. 77-79.
9. Мейер Д. Теория реляционных баз данных: [пер. с англ.] / Д. Мейер. – Москва: Мир, 1987. – 608 с.

Надійшла у першій редакції 15.05.2014, в останній - 17.07.2014.

УДК 004.056.55

Моделирование алгебраической структуры шифра AES с использованием цепных дробей

Ю. И. Горбенко, А. А. Кузнецов, С. В. Костенко
Харьковский национальный университет имени В.Н. Каразина, Украина

В данной работе с использованием математического аппарата цепных дробей исследуется алгебраическая структура шифра AES. Приводится краткое описание шифра AES (FIPS-197), рассматриваются основные преобразования, используемые в этом крипто алгоритме, и его алгебраическая структура. С использованием полиномиального описания вводится алгебраическая форма нелинейного узла замен шифра, позволяющая существенно упростить систему уравнений, связывающих значения открытого текста, секретного ключа и полученного шифр-текста.

Ключевые слова: *цепные дроби, алгебраическая структура, шифр-текст.*

В даній роботі з використанням математичного апарату ланцюгових дробів досліджується алгебраїчна структура шифру AES. Приводиться короткий опис шифру AES (FIPS-197), розглядаються основні перетворення, які використовуються в цьому крипто алгоритмі, та його алгебраїчна структура. З використанням поліноміального опису вводиться алгебраїчна форма нелінійного вузла замін шифру, яка дозволяє суттєво спростити систему рівнянь, які зв'язують значення відкритого тексту, таємного ключа та отриманого шифр-тексту.

Ключові слова: *ланцюгові дроби, алгебраїчна структура, шифр-текст.*

In the paper, the algebraic structure of cipher AES is studied using the mathematical apparatus of continued fractions. Provided here brief description of the cipher AES (FIPS-197) covers the basic transformations used in this cryptographic algorithm and its algebraic structure. Polynomial description of the Algebraic form of nonlinear input node substitutions cipher greatly simplifies the system of equations connecting the values of the plaintext, the secret key and the resulting cipher text.

Key words:*continued fraction, algebraic structure, cipher text.*

1. Введение

В качестве стандарта шифрования по результатам открытого конкурса AES, проведенного Национальным институтом стандартов и технологий США в 1997-2001 г.г. был принят Advanced Encryption Standard (AES) - симметричный алгоритм блочного шифрования (размер блока 128 бит, размер ключа 128/192/256 бит) [1 - 3]. Целью данной работы является анализ алгебраической структуры шифра AES, т.е. вывод алгебраических уравнений, связывающих значения открытого текста, секретного ключа и полученного шифр-текста. При этом предлагается использовать математический аппарат цепных дробей.

2. Краткое описание алгоритма шифрования

Конкурс AES проводился Национальным институтом стандартов и технологий США в 1997-2001 г.г. и победителем объявлен алгоритм Rijndael (Рэндал) [1 - 3]. Фактически шифр AES, стандартизированный в FIPS-197, представляет собой один из вариантов алгоритма Rijndael.

Различные преобразования в алгоритме Rijndael (AES) оперируют с промежуточным результатом, называемым *Состояние* (State). Состояние может быть изображено как прямоугольный массив байтов. Этот массив имеет 4 строки, количество столбцов обозначается через *Nb* и равно длине блока, деленной на 32 (для AES используется *Nb* = 4). Ключ шифра также изображается как прямоугольный массив с 4 строками. Количество столбцов Ключа шифра обозначается через *Nk* и равно длине ключа, деленной на 32 (для AES используется *Nk* = 4, 6 или 8).

Вход и выход, используемые в Rijndael в его внешнем интерфейсе, являются одноразмерными массивами 8-битных байтов, пронумерованных в восходящем порядке от 0 до $4 \cdot Nb - 1$. Эти блоки поэтому имеют длины 16, 14 и 32 байт и индексы массивов в пределах 0..15, 0..23, 0..31.

Ключ шифра рассматривается как одноразмерные массивы 8-битных байтов, пронумерованные в восходящем порядке от 0 до $4 \cdot Nk - 1$. Эти блоки поэтому имеют длины 16, 24 и 32 байт и индексы массивов в пределах 0..15, 0..23, 0..31.

Входные байты шифра («открытый текст», если используется режим зашифровывания ECB) отображаются в байты состояния в порядке $a_{0,0}, a_{1,0}, a_{2,0}, a_{3,0}, a_{0,1}, a_{1,1}, a_{2,1}, a_{3,1}, a_{4,1} \ldots$, и байты ключа шифра отображаются в массив в порядке $k_{0,0}, k_{1,0}, k_{2,0}, k_{3,0}, k_{0,1}, k_{1,1}, k_{2,1}, k_{3,1}, k_{4,1} \ldots$ В конце оперирования шифра его выход извлекается из состояния выбором байтов состояния в таком же порядке.

Следовательно, если одноразмерным индексом байта в блоке является *n*, и двуразмерным индексом является (*i,j*), мы имеем:

$$i = n \bmod 4; \quad j = \lfloor n/4 \rfloor; \qquad n = i + 4 * j .$$

Далее, индекс *i* есть также номером байта в 4-байтном векторе или слове и *j* есть индекс вектора или слова в отдельно взятом блоке.

Количество раундов обозначено через *Nr* и зависит от значений *Nb* и *Nk*. Для шифра AES используются следующие параметры: *Nr* = 10 (при *Nk* = 4), *Nr* = 12 (при *Nk* = 6), *Nr* = 14 (при *Nk* = 8).

Раундовое преобразование состоит из четырех различных преобразований: *ByteSub(State); ShiftRow(State); MixColumn(State); AddRoundKey(State, RoundKey).* Последний раунд с удаленным шагом *MixColumn(State).* В этих обозначениях «функции» (Round, ByteSub, ShiftRow, …) оперируют с массивами, на которые указывают указатели (State, RoundKey).

Преобразование ByteSub есть нелинейной заменой байтов, которое оперирует с каждым байтом Состояния независимо. Таблица замена (или S-блок) инвертируема и создана композицией двух преобразований:

1. Вычисляется мультипликативно обратный элемент в поле $GF(2^8)$ (в полиномиальном представлении элементов поля операции производятся по модулю неприводимого бинарного многочлена $m(x) = x^8 + x^4 + x^3 + x + 1$), при этом элемент ‘00’ отображается в себя;

2. Применяется аффинное преобразование (над двоичным полем GF(2)), которое определено как:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \quad (1)$$

Применение описанного S-блока ко всем байтам состояния обозначается как ByteSub(State). Обратным к ByteSub является замена байтов с применением инвертированной таблицы. Это достигается обращением аффинного отображения, за которым следует взятие мультипликативно обратного в поле $GF(2^8)$.

В ***преобразовании ShiftRow*** строки Состояния сдвигаются на различное количество позиций. Строка 0 не смещается, строка 1 смещается на C1 байт, строка 2 – на C2 байт и строка 3 – на C3 байт. Для шифра AES: C1 = 1, C2 = 2, C3 = 3. Операция сдвига строк Состояния на определенную величину обозначается через ShiftRow(State). Обратным к ShiftRow есть циклический сдвиг 3 нижних строк на *Nb* – C1, *Nb* – C2, *Nb* – C3, байт соответственно так, что байт на позиции *j* в строке *i* двигается на позицию (*j* + *Nb* – Ci) mod *Nb*.

В MixColumn столбцы Состояния рассматриваются как многочлены над полем $GF(2^8)$ и умножаются по модулю x^4+1 с фиксированным многочленом $c(x)$, заданным как $c(x) = '03'x^3 + '01'x^2 + '01'x + '02'$.

Этот многочлен является взаимно-простым с x^4+1 и поэтому инвертируемым. Это может быть записано как матричное умножение. Пусть $b(x) = c(x) \otimes a(x)$,

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Применение этой операции ко всем столбцам Состояния обозначается через MixColumn(State). Инверсия преобразования MixColumn сходна к MixColumn. Каждый столбец преобразуется умножением его на особый многочлен $d(x)$, определенный через $\left('03'x^3 + '01'x^2 + '01'x + '02'\right) \otimes d(x) = '01'$.

Он задан таким образом:

$$d(x) = '0B'x^3 + '0D'x^2 + '09'x + '0E'.$$

Сложение с раундовым ключом. В этой операции к Состоянию применяется RoundKey простым побитным сложением по модулю 2. Раундовый ключ

производится из ключа шифра посредством процедуры формирования ключей. Длина раундового ключа равна длине блока *Nb*. Преобразование, которое состоит из побитового сложения Состояния и раундового ключа обозначается через AddRoundKey(State,RoundKey) оно является своей собственной инверсией.

Процедура формирования ключей. Раундовые ключи производятся из ключа шифра посредством процедуры формирования ключей. Она включает в себя два компонента: Расширение ключа и Выбор раундового ключа. Основной принцип заключается в следующем:

- общее количество бит раундового ключа равно длине блока, умноженной на количество раундов плюс 1 (т.е. для блока длины 128 бит и 10 раундов необходимо 1408 раундовых ключей);
- Ключ шифра расширяется в Расширенный ключ;
- раундовые ключи берутся из Расширенного ключа следующим образом: первый раундовый ключ состоит из первых *Nb* слов, второй – из следующих *Nb* слов, и так далее.

Расширенный ключ есть линейный массив 4-байтных слов и обозначается как *W[Nb*(Nr+1)]*. Первые *Nk* слов содержат Ключ шифра. Все другие слова определяются рекурсивно по величинам слов с меньшими индексами. Функция Расширения ключа зависит от величины *Nk*: есть версия для *Nk* меньше или равно 6, и есть версия для *Nk* больше 6.

Для $Nk \le 6$ расширенный ключ формируется по следующему правилу:

*W[i] = (Key[4*i],Key[4*i+1], Key[4*i+2],Key[4*i+3]), i = 0, …, Nk – 1;*

*W[i] = W[i - Nk] ^ temp, i = Nk, …, Nb * (Nr + 1) – 1,*

при этом если *(i)mod(Nk)=0*:

temp = SubByte(RotByte(temp)) ^ Rcon[i / Nk],

если *(i)mod(Nk)≠0*:

temp = W[i - 1].

В этом представлении *SubByte(W)* есть функция, которая возвращает 4-байтовое слово, в котором каждый байт есть результат применения S-блока к байту на соответствующей позиции во входном слове. Функция *RotByte(W)* возвращает слово, в котором байты являются циклической перестановкой байтов на входе таким образом, что входное слово (a, b, c, d) продуцируется в выходное слово (b, c, d, a).

Таким образом, первые *Nk* слов заполнены Ключом шифра. Каждое следующее слово *W[i]* равно сумме по модулю 2 с предыдущим словом *W[i–1]* и словом, расположенным на *Nk* позиций ранее *W[i–Nk]*. Для слов с позициями, кратными *Nk*, перед сложением по модулю 2 к слову *W[i–1]* применяется преобразование и прибавляется по модулю 2 раундовая константа. Это преобразование состоит из циклического сдвига байт с слове (*RotByte*), за которым следует применение табличного поиска ко всем 4 байтам слова (*SubByte*).

Отличием при формировании расширенного ключа $Nk > 6$ есть то, что для *i*–4, кратного *Nk*, перед сложением по модулю 2 к W[i–1] применяется SubByte.

Раундовые константы не зависят от *Nk* и определены как:

$$\mathrm{Rcon}[i] = \left(\mathrm{RC}[i], '00', '00', '00'\right),$$

где RC[i] представляют собой элемент поля GF(2^8) со значением $x^{(i-1)}$, такой, что

$$\mathrm{RC}[1] = 1 \ \ (\text{i.e.}\,'01')$$
$$\mathrm{RC}[i] = \mathrm{x} \ \ (\text{i.e.}\,'02') \bullet \left(\mathrm{RC}[i-1]\right) = x^{(i-1)}.$$

Выбор раундового ключа. Раундовый ключ i задан буфером слов раундового ключа от *W[Nb*i]* до *W[Nb*(i+1)]*.

Шифр Rijndael состоит из:

- Начального добавления раундового ключа
- $Nr - 1$ раундов
- Окончательного раунда.

Аспекты реализации. Шифр Rijndael приспособлен для эффективной реализации на широком спектре процессоров и специализированном аппаратном обеспечении. В частности, для 32-х битной платформы различные шаги раундового преобразования могут быть скомбинированы в единственный набор (множество) выборочных таблиц, приводящих к быстрой реализации на процессорах с длиной слова 32 или выше.

Выразим один столбец выхода раунда e в величинах байт входа раунда a (значением $a_{i,j}$ обозначим байт в строке i и столбце j, значением a_j обозначим столбец j Состояния a). Для прибавления ключей и для преобразования MixColumn получим:

$$\begin{bmatrix} e_{0,j} \\ e_{1,j} \\ e_{2,j} \\ e_{3,j} \end{bmatrix} = \begin{bmatrix} d_{0,j} \\ d_{1,j} \\ d_{2,j} \\ d_{3,j} \end{bmatrix} \oplus \begin{bmatrix} k_{0,j} \\ k_{1,j} \\ k_{2,j} \\ k_{3,j} \end{bmatrix} \ and \ \begin{bmatrix} d_{0,j} \\ d_{1,j} \\ d_{2,j} \\ d_{3,j} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} c_{0,j} \\ c_{1,j} \\ c_{2,j} \\ c_{3,j} \end{bmatrix}.$$

Для преобразований ShiftRow и ByteSub имеем:

$$\begin{bmatrix} c_{0,j} \\ c_{1,j} \\ c_{2,j} \\ c_{3,j} \end{bmatrix} = \begin{bmatrix} b_{0,j} \\ b_{1,j+C1} \\ b_{2,j+C2} \\ b_{3,j+C3} \end{bmatrix} \ and \ b_{i,j} = \mathrm{S}\left[a_{i,j}\right].$$

В последнем выражении индексы столбцов должны быть взяты по модулю Nb. Сгруппировав выражения, приведенные выше, получим:

$$\begin{bmatrix} e_{0,j} \\ e_{1,j} \\ e_{2,j} \\ e_{3,j} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} \mathrm{S}\left[a_{0,j}\right] \\ \mathrm{S}\left[a_{1,j+C1}\right] \\ \mathrm{S}\left[a_{2,j+C2}\right] \\ \mathrm{S}\left[a_{3,j+C3}\right] \end{bmatrix} \oplus \begin{bmatrix} k_{0,j} \\ k_{1,j} \\ k_{2,j} \\ k_{3,j} \end{bmatrix}.$$

Матричное умножение выразим как линейную комбинацию векторов:

$$\begin{bmatrix} e_{0,j} \\ e_{1,j} \\ e_{2,j} \\ e_{3,j} \end{bmatrix} = \mathrm{S}\left[a_{0,j}\right]\begin{bmatrix} 02 \\ 01 \\ 01 \\ 03 \end{bmatrix} \oplus \mathrm{S}\left[a_{1,j+C1}\right]\begin{bmatrix} 03 \\ 02 \\ 01 \\ 01 \end{bmatrix} \oplus \mathrm{S}\left[a_{2,j+C2}\right]\begin{bmatrix} 01 \\ 03 \\ 02 \\ 01 \end{bmatrix} \oplus \mathrm{S}\left[a_{3,j+C3}\right]\begin{bmatrix} 01 \\ 01 \\ 03 \\ 02 \end{bmatrix} \oplus \begin{bmatrix} k_{0,j} \\ k_{1,j} \\ k_{2,j} \\ k_{3,j} \end{bmatrix}$$

Множители $\mathrm{S}[a_{i,j}]$ четырех векторов получены составлением выборочной таблицы на входах байтах $a_{i,j}$ в таблице S-блока S[256].

Определим 4 таблицы с 256-ю вводами 4-байтовых слов от T_0 до T_3:

$$\mathrm{T}_0[a] = \begin{bmatrix} \mathrm{S}[a] \bullet 02 \\ \mathrm{S}[a] \\ \mathrm{S}[a] \\ \mathrm{S}[a] \bullet 03 \end{bmatrix} \quad \mathrm{T}_1[a] = \begin{bmatrix} \mathrm{S}[a] \bullet 03 \\ \mathrm{S}[a] \bullet 02 \\ \mathrm{S}[a] \\ \mathrm{S}[a] \end{bmatrix}$$

$$\mathrm{T}_2[a] = \begin{bmatrix} \mathrm{S}[a] \\ \mathrm{S}[a] \bullet 03 \\ \mathrm{S}[a] \bullet 02 \\ \mathrm{S}[a] \end{bmatrix} \quad \mathrm{T}_3[a] = \begin{bmatrix} \mathrm{S}[a] \\ \mathrm{S}[a] \\ \mathrm{S}[a] \bullet 03 \\ \mathrm{S}[a] \bullet 02 \end{bmatrix}.$$

Эти 4 таблицы занимают до 4 Кбайт памяти, с их использованием раундовое преобразование может быть выражено в виде:

$$e_j = \mathrm{T}_0\left[a_{0,j}\right] \oplus \mathrm{T}_1\left[a_{1,j+C1}\right] \oplus \mathrm{T}_2\left[a_{2,j+C2}\right] \oplus \mathrm{T}_3\left[a_{3,j+C3}\right] \oplus k_j . \qquad (2)$$

Следовательно, реализация выборочных таблиц с 4 Кбайтами памяти требует только 4 сложения по модулю 2 на столбец каждого раунда.

Можно заметить, что $\mathrm{T}_i[a]$ = RotByte($\mathrm{T}_{i-1}[a]$). Тогда заплатив тремя дополнительными ротациями на столбец в каждом раунде реализация шифра может быть выполнена с одной таблицей, общим размером в 1 Кбайт памяти:

$$e_j = k_j \oplus \mathrm{T}_0\left[b_{0,j}\right] \oplus$$
$$\oplus \texttt{Rotbyte}\left(\mathrm{T}_0\left[b_{1,j+C1}\right] \oplus \texttt{Rotbyte}\left(\mathrm{T}_0\left[b_{2,j+C2}\right] \oplus \texttt{Rotbyte}\left(\mathrm{T}_0\left[b_{3,j+C3}\right]\right)\right)\right)$$

Т.к. в последнем раунде нет операции MixColumn, взамен Т-таблиц должна быть использована S-таблица.

3. Алгебраическая структура шифра

Рассмотрим алгебраическую структуру шифра. При этом ограничимся исследованием шифра AES с $Nk = 4$, т.е. будем рассматривать Rijndael с длиной ключа и длиной блока в 128 бит, значения Состояния и Ключа шифра представляются в виде таблиц 4×4. Число раундов преобразования $Nr = 10$.

Целью проводимых исследований является анализ алгебраической структуры шифра, т.е. вывод алгебраических уравнений, связывающих значения открытого текста, секретного ключа и полученного шифр-текста. Для удобства дальнейших вычислений обозначим байты Состояния и Ключа шифра на выходе каждого *u* –го раунда (*u = 1, 2, ..., 10*) верхним индексом, т.е. в виде $a_{i,j}^{u}$ и $k_{i,j}^{u}$. Начальное значение Состояния (открытый текст до шифрования) обозначим в виде массива $a_{i,j}^{0}$. Значения $a_{i,j}^{10}$ Состояния на 10-м раунде будут соответствовать байтам полученного шифр-текста.

Байты секретного ключа, в соответствии с приведенным выше правилом формирования раундовых ключей, будут записаны в массив значений $k_{i,j}^{1}$ Ключа шифра (на первом раунде):

$k_{0,0}^{1}$	$k_{0,1}^{1}$	$k_{0,2}^{1}$	$k_{0,3}^{1}$
$k_{1,0}^{1}$	$k_{1,1}^{1}$	$k_{1,2}^{1}$	$k_{1,3}^{1}$
$k_{2,0}^{1}$	$k_{2,1}^{1}$	$k_{2,2}^{1}$	$k_{2,3}^{1}$
$k_{3,0}^{1}$	$k_{3,1}^{1}$	$k_{3,2}^{1}$	$k_{3,3}^{1}$

Выразим алгебраическую зависимость значений $a_{i,j}^{1}$ Состояния (после первого раунда шифрования) от значений $a_{i,j}^{0}$ Состояния (байт открытого текста) и значений $k_{i,j}^{1}$ Ключа шифра. Используя (2) для всех $j = 0, ..., 3$ имеем [4]:

$$a_{0,j}^{1} = '02' S[a_{0,j}^{0}] \oplus '03' S[a_{1,j+1}^{0}] \oplus S[a_{2,j+2}^{0}] \oplus S[a_{3,j+3}^{0}] \oplus k_{0,j}^{1};$$

$$a_{1,j}^{1} = S[a_{0,j}^{0}] \oplus '02' S[a_{1,j+1}^{0}] \oplus '03' S[a_{2,j+2}^{0}] \oplus S[a_{3,j+3}^{0}] \oplus k_{1,j}^{1};$$

$$a_{2,j}^{1} = S[a_{0,j}^{0}] \oplus S[a_{1,j+1}^{0}] \oplus '02' S[a_{2,j+2}^{0}] \oplus '03' S[a_{3,j+3}^{0}] \oplus k_{2,j}^{1};$$

$$a_{3,j}^{1} = '03' S[a_{0,j}^{0}] \oplus S[a_{1,j+1}^{0}] \oplus S[a_{2,j+2}^{0}] \oplus '02' S[a_{3,j+3}^{0}] \oplus k_{3,j}^{1},$$

где индексы приводятся по модулю 4.

Рассмотрим алгебраическую структуру S-блока шифра AES.

В работе [5] показано, что с помощью интерполяции можно получить следующую алгебраическую форму:

$$S[x] = '63' + '8F'x^{127} + 'B5'x^{191} + '01'x^{223} + 'F4'x^{239} + '25'x^{247} + \\ + 'F9'x^{251} + '09'x^{253} + '05'x^{254}, \quad (3)$$

которая, однако, сложна для вывода окончательных выражений, связывающих секретных ключ, открытый и закрытый текст.

В работе [6] используется следующая форма представления S-блока:

$$S[x] = w_8 + \sum_{d=0}^{7} w_d x^{255-2^d}, \quad (4)$$

для некоторых констант $w_0,...,w_8$, что также приводит к чрезвычайно громоздким окончательным выражениям.

Некоторое развитие представление (4) получило в статье [7].

В данной работе предлагается иная алгебраическая форма представления S-блока шифра AES, которая, как будет показано ниже, значительно упрощает вывод итоговых уравнений.

Рассмотрим выражение (1). Воспользуемся полиномиальным представлением элементов конечных полей для описания алгебраической структуры S-блока. Для этого вход представим в виде многочлена $a(x) \in GF(2^8)$, его мультипликативно обратный элемент (по модулю $m(x)$) в поэлементной записи формально запишем в виде:

$$\frac{1}{a(x)} = b(x) = b_0 + b_1 x + ... + b_7 x^7,$$

где коэффициенты $b_0,..., b_7$ определяют значения вектора-столбца в произведении (1), т.е.

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix}.$$

С помощью подстановки получим следующее выражение:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} + \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{bmatrix} = \begin{bmatrix} b_0 + b_4 + b_5 + b_6 + b_7 \\ b_0 + b_1 + b_5 + b_6 + b_7 \\ b_0 + b_1 + b_2 + b_6 + b_7 \\ b_0 + b_1 + b_2 + b_3 + b_7 \\ b_0 + b_1 + b_2 + b_3 + b_4 \\ b_1 + b_2 + b_3 + b_4 + b_5 \\ b_2 + b_3 + b_4 + b_5 + b_6 \\ b_3 + b_4 + b_5 + b_6 + b_7 \end{bmatrix} + \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{bmatrix} =$$

$$= \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} + \begin{bmatrix} b_7 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} + \begin{bmatrix} b_6 \\ b_7 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} + \begin{bmatrix} b_5 \\ b_6 \\ b_7 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + \begin{bmatrix} b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{bmatrix}.$$

Сумму пяти векторов-столбцов в правой части выражения запишем в полиномиальном виде (с операциями умножения по модулю x^8+1), получим:

$$\begin{aligned} S[a(x)] &= b(x) + xb(x) + x^2b(x) + x^3b(x) + x^4b(x) + c(x) = \\ &= (1 + x + x^2 + x^3 + x^4)b(x) + c(x) = \frac{'1F'}{a(x)} + c(x). \end{aligned} \tag{5}$$

Полученная алгебраическая форма (5) является наиболее простым из известных выражений, например, по сравнению с (3) и (4). Фактически, выражение (5) устанавливает наиболее простой (из известных автору) способ вычисления S-блока шифра AES – для каждого входного элемента достаточно вычислить мультипликативно обратный и умножить полученный результат на константу (по модулю x^8+1) и сложить с константой. Для удобства в дальнейших вычислениях будем использовать форму $S[a]=\frac{1}{a}$, подразумевая под этой записью вычисление мультиптикативно обратного в поле $GF(2^8)$ с умножением полученного результата на многочлен $1+x+x^2+x^3+x^4$ (по модулю x^8+1) и сложения с константой.

Воспользуемся выражением (5) (в форме $S[a]=\frac{1}{a}$) для построения системы алгебраических уравнений, описывающих **первый раунд шифрования**:

$$a^1_{0,j} = A^0_{0,j} \oplus k^1_{0,j};\ a^1_{1,j} = A^0_{1,j} \oplus k^1_{1,j};\ a^1_{2,j} = A^0_{2,j} \oplus k^1_{2,j};\ a^1_{3,j} = A^0_{3,j} \oplus k^1_{3,j}, \quad (6)$$

где:

$$A^0_{0,j} = \frac{'02'}{a^0_{0,j}} \oplus \frac{'03'}{a^0_{1,j+1}} \oplus \frac{1}{a^0_{2,j+2}} \oplus \frac{1}{a^0_{3,j+3}},$$

$$A^0_{1,j} = \frac{1}{a^0_{0,j}} \oplus \frac{'02'}{a^0_{1,j+1}} \oplus \frac{'03'}{a^0_{2,j+2}} \oplus \frac{1}{a^0_{3,j+3}},$$

$$A^0_{2,j} = \frac{1}{a^0_{0,j}} \oplus \frac{1}{a^0_{1,j+1}} \oplus \frac{'02'}{a^0_{2,j+2}} \oplus \frac{'03'}{a^0_{3,j+3}},$$

$$A^0_{3,j} = \frac{'03'}{a^0_{0,j}} \oplus \frac{1}{a^0_{1,j+1}} \oplus \frac{1}{a^0_{2,j+2}} \oplus \frac{'02'}{a^0_{3,j+3}}.$$

Из этой линейной системы неизвестные байты секретного ключа $k^1_{0,j}$, $k^1_{1,j}$, $k^1_{2,j}$ и $k^1_{3,j}$ выражаются линейной комбинацией байт шифр-текста и инвертированных (в конечном поле) байт открытого текста. Для вычисления одного байта секретного ключа необходимо выполнить 4 инверсии, 2 умножения, 5 сложения в конечном поле $GF(2^8)$ и 4 умножения на многочлен $1 + x + x^2 + x^3 + x^4$ по модулю $x^8 + 1$.

Для второго раунда система уравнений будет иметь вид:

$$a^2_{0,j} = A^1_{0,j} \oplus k^2_{0,j};\ a^2_{1,j} = A^1_{1,j} \oplus k^2_{1,j};\ a^2_{2,j} = A^1_{2,j} \oplus k^2_{2,j};\ a^2_{3,j} = A^1_{3,j} \oplus k^2_{3,j}, \quad (7)$$

где:

$$A^1_{0,j} = \frac{'02'}{A^0_{0,j} \oplus k^1_{0,j}} \oplus \frac{'03'}{A^0_{1,j+1} \oplus k^1_{1,j+1}} \oplus \frac{1}{A^0_{2,j+2} \oplus k^1_{2,j+2}} \oplus \frac{1}{A^0_{3,j+3} \oplus k^1_{3,j+3}},$$

$$A^1_{1,j} = \frac{1}{A^0_{0,j} \oplus k^1_{0,j}} \oplus \frac{'02'}{A^0_{1,j+1} \oplus k^1_{1,j+1}} \oplus \frac{'03'}{A^0_{2,j+2} \oplus k^1_{2,j+2}} \oplus \frac{1}{A^0_{3,j+3} \oplus k^1_{3,j+3}},$$

$$A^1_{2,j} = \frac{1}{A^0_{0,j} \oplus k^1_{0,j}} \oplus \frac{1}{A^0_{1,j+1} \oplus k^1_{1,j+1}} \oplus \frac{'02'}{A^0_{2,j+2} \oplus k^1_{2,j+2}} \oplus \frac{'03'}{A^0_{3,j+3} \oplus k^1_{3,j+3}},$$

$$A^1_{3,j} = \frac{'03'}{A^0_{0,j} \oplus k^1_{0,j}} \oplus \frac{1}{A^0_{1,j+1} \oplus k^1_{1,j+1}} \oplus \frac{1}{A^0_{2,j+2} \oplus k^1_{2,j+2}} \oplus \frac{'02'}{A^0_{3,j+3} \oplus k^1_{3,j+3}}.$$

В системе (7) в качестве неизвестных переменных выступают байты секретного ключа $k^1_{0,j}$, $k^1_{1,j+1}$, $k^1_{2,j+2}$ и $k^1_{3,j+3}$, а также байты раундовых ключей $k^2_{0,j}$, $k^2_{1,j}$, $k^2_{2,j}$ и $k^2_{3,j}$. Выразим раундовые ключи (для второго раунда) в виде алгебраических уравнений от секретных ключей шифра (от ключей первого раунда):

$$k^2_{0,0} = k^1_{0,0} \oplus \mathrm{S}[k^1_{1,3}] \oplus '01' = k^1_{0,0} \oplus '01' \oplus \frac{1}{k^1_{1,3}};\ k^2_{1,0} = k^1_{1,0} \oplus \mathrm{S}[k^1_{2,3}] = k^1_{1,0} \oplus \frac{1}{k^1_{2,3}};$$

$$k^2_{2,0} = k^1_{2,0} \oplus \mathrm{S}[k^1_{3,3}] = k^1_{2,0} \oplus \frac{1}{k^1_{3,3}};\ k^2_{3,0} = k^1_{3,0} \oplus \mathrm{S}[k^1_{0,3}] = k^1_{3,0} \oplus \frac{1}{k^1_{0,3}};$$

$$k^2_{0,1} = k^1_{0,1} \oplus k^2_{0,0} = k^1_{0,1} \oplus k^1_{0,0} \oplus '01' \oplus \frac{1}{k^1_{1,3}};\ k^2_{1,1} = k^1_{1,1} \oplus k^2_{1,0} = k^1_{1,1} \oplus k^1_{1,0} \oplus \frac{1}{k^1_{2,3}};$$

$$k^2_{2,1} = k^1_{2,1} \oplus k^2_{2,0} = k^1_{2,1} \oplus k^1_{2,0} \oplus \frac{1}{k^1_{3,3}};\ k^2_{3,1} = k^1_{3,1} \oplus k^2_{3,0} = k^1_{3,1} \oplus k^1_{3,0} \frac{1}{k^1_{0,3}};$$

$$k^2_{0,2} = k^1_{0,2} \oplus k^2_{0,1} = k^1_{0,2} \oplus k^1_{0,1} \oplus k^1_{0,0} \oplus '01' \oplus \frac{1}{k^1_{1,3}};$$

$$k^2_{1,2} = k^1_{1,2} \oplus k^2_{1,1} = k^1_{1,2} \oplus k^1_{1,1} \oplus k^1_{1,0} \oplus \frac{1}{k^1_{3,2}};\ k^2_{2,2} = k^1_{2,2} \oplus k^2_{2,1} = k^1_{2,2} \oplus k^1_{2,1} \oplus k^1_{2,0} \oplus \frac{1}{k^1_{3,3}};$$

$$k^2_{3,2} = k^1_{3,2} \oplus k^2_{3,1} = k^1_{3,2} \oplus k^1_{3,1} \oplus k^1_{3,0} \oplus \frac{1}{k^1_{3,0}};$$

$$k^2_{0,3} = k^1_{0,3} \oplus k^2_{0,2} = k^1_{0,3} \oplus k^1_{0,2} \oplus k^1_{0,1} \oplus k^1_{0,0} \oplus '01' \oplus \frac{1}{k^1_{3,1}};$$

$$k^2_{1,3} = k^1_{1,3} \oplus k^2_{1,2} = k^1_{1,3} \oplus k^1_{1,2} \oplus k^1_{1,1} \oplus k^1_{1,0} \oplus \frac{1}{k^1_{2,3}};$$

$$k^2_{2,3} = k^1_{2,3} \oplus k^2_{2,2} = k^1_{2,3} \oplus k^1_{2,2} \oplus k^1_{2,1} \oplus k^1_{2,0} \oplus \frac{1}{k^1_{3,3}};$$

$$k^2_{3,3} = k^1_{3,3} \oplus k^2_{3,2} = k^1_{3,3} \oplus k^1_{3,2} \oplus k^1_{3,1} \oplus k^1_{3,0} \oplus \frac{1}{k^1_{0,3}};.$$

С учетом последних обозначений формулы (7) примут вид:

$$a^2_{0,j} = A^1_{0,j} \oplus \frac{1}{k^1_{1,3}} \oplus '01' \bigoplus_{w=0}^{j} k^1_{0,w};\ a^2_{1,j} = A^1_{1,j} \oplus \frac{1}{k^1_{2,3}} \bigoplus_{w=0}^{j} k^1_{1,w};$$

$$a^2_{2,j} = A^1_{2,j} \oplus \frac{1}{k^1_{3,3}} \bigoplus_{w=0}^{j} k^1_{2,w};\ a^2_{3,j} = A^1_{3,j} \oplus \frac{1}{k^1_{0,3}} \bigoplus_{w=0}^{j} k^1_{3,w}.$$

Таким образом, значения секретного ключа, открытого и закрытого текста после двух раундов шифрования связаны системой из 16 уравнений от 16 неизвестных. Каждое уравнение содержит от 4 до 8 неизвестных (в зависимости от значения индекса j).

На третьем раунде система уравнений будет иметь вид:

$$a^3_{0,j} = A^2_{0,j} \oplus k^3_{0,j};\ a^3_{1,j} = A^2_{1,j} \oplus k^3_{1,j};\ a^3_{2,j} = A^2_{2,j} \oplus k^3_{2,j};\ a^3_{3,j} = A^2_{3,j} \oplus k^3_{3,j},\ (8)$$

где:

$$A^2_{0,j} = \frac{'02'}{A^1_{0,j} \oplus k^2_{0,j}} \oplus \frac{'03'}{A^1_{1,j+1} \oplus k^2_{1,j+1}} \oplus \frac{1}{A^1_{2,j+2} \oplus k^2_{2,j+2}} \oplus \frac{1}{A^1_{3,j+3} \oplus k^2_{3,j+3}},$$

$$A^2_{1,j} = \frac{1}{A^1_{0,j} \oplus k^2_{0,j}} \oplus \frac{'02'}{A^1_{1,j+1} \oplus k^2_{1,j+1}} \oplus \frac{'03'}{A^1_{2,j+2} \oplus k^2_{2,j+2}} \oplus \frac{1}{A^1_{3,j+3} \oplus k^2_{3,j+3}},$$

$$A^2_{2,j} = \frac{1}{A^1_{0,j} \oplus k^2_{0,j}} \oplus \frac{1}{A^1_{1,j+1} \oplus k^2_{1,j+1}} \oplus \frac{'02'}{A^1_{2,j+2} \oplus k^2_{2,j+2}} \oplus \frac{'03'}{A^1_{3,j+3} \oplus k^2_{3,j+3}},$$

$$A^3_{3,j} = \frac{'03'}{A^1_{0,j} \oplus k^2_{0,j}} \oplus \frac{1}{A^1_{1,j+1} \oplus k^2_{1,j+1}} \oplus \frac{1}{A^1_{2,j+2} \oplus k^2_{2,j+2}} \oplus \frac{'02'}{A^1_{3,j+3} \oplus k^2_{3,j+3}}.$$

В системе (8) неизвестными являются байты раундовых ключей $k^2_{0,j}$, $k^2_{1,j+1}$, $k^2_{2,j+2}$, $k^2_{3,j+3}$, $k^3_{0,j}$, $k^3_{1,j}$, $k^3_{2,j}$ и $k^3_{3,j}$. Выше было показано, что $k^2_{0,j}$, $k^2_{1,j+1}$, $k^2_{2,j+2}$, $k^2_{3,j+3}$ выражаются в виде алгебраических уравнений от неизвестных байтов секретного ключа $k^1_{0,j}$, $k^1_{1,j+1}$, $k^1_{2,j+2}$ и $k^1_{3,j+3}$. Выразим по аналогии раундовые ключи третьего раунда $k^3_{0,j}$, $k^3_{1,j}$, $k^3_{2,j}$ и $k^3_{3,j}$:

$$k^3_{0,0} = k^2_{0,0} \oplus \mathrm{S}[k^2_{1,3}] \oplus '02' = k^1_{0,0} \oplus \frac{1}{k^1_{1,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{1,w} \oplus \frac{1}{k^1_{2,3}}} \oplus '03';$$

$$k^3_{1,0} = k^2_{1,0} \oplus \mathrm{S}[k^2_{2,3}] = k^1_{1,0} \oplus \frac{1}{k^1_{2,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{2,w} \oplus \frac{1}{k^1_{3,3}}};$$

$$k^3_{2,0} = k^2_{2,0} \oplus \mathrm{S}[k^2_{3,3}] = k^1_{2,0} \oplus \frac{1}{k^1_{3,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{3,w} \oplus \frac{1}{k^1_{0,3}}};$$

$$k^3_{3,0} = k^2_{3,0} \oplus \mathrm{S}[k^2_{0,3}] = k^1_{3,0} \oplus \frac{1}{k^1_{0,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{0,w} \oplus \frac{1}{k^1_{3,1}} \oplus '01'};$$

$$k^3_{0,1} = k^2_{0,1} \oplus k^3_{0,0} = k^1_{0,1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{1,w} \oplus \frac{1}{k^1_{2,3}}} \oplus '02';$$

$$k^3_{1,1} = k^2_{1,1} \oplus k^3_{1,0} = k^1_{1,1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{2,w} \oplus \frac{1}{k^1_{3,3}}};$$

$$k^3_{2,1} = k^2_{2,1} \oplus k^3_{2,0} = k^1_{2,1} \oplus \mathrm{S}[\bigoplus_{w=0}^{3} k^1_{3,w}] = k^1_{2,1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{3,w} \oplus \frac{1}{k^1_{0,3}}};$$

$$k_{3,1}^3 = k_{3,1}^2 \oplus k_{3,0}^3 = k_{3,1}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'};$$

$$k_{0,2}^3 = k_{0,2}^2 \oplus k_{0,1}^3 = k_{0,2}^1 \oplus k_{0,0}^1 \oplus \frac{1}{k_{1,3}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{1,w}^1 \oplus \frac{1}{k_{2,3}^1}} \oplus '03';$$

$$k_{1,2}^3 = k_{1,2}^2 \oplus k_{1,1}^3 = k_{1,2}^1 \oplus k_{1,0}^1 \oplus \frac{1}{k_{3,2}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{2,w}^1 \oplus \frac{1}{k_{3,3}^1}};$$

$$k_{2,2}^3 = k_{2,2}^2 \oplus k_{2,1}^3 = k_{2,2}^1 \oplus k_{2,0}^1 \oplus \frac{1}{k_{3,3}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{3,w}^1 \oplus \frac{1}{k_{0,3}^1}};$$

$$k_{3,2}^3 = k_{3,2}^2 \oplus k_{3,1}^3 = k_{3,1}^1 \oplus k_{3,0}^1 \oplus \frac{1}{k_{3,0}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'};$$

$$k_{0,3}^3 = k_{0,3}^2 \oplus k_{0,2}^3 = k_{0,3}^1 \oplus k_{0,1}^1 \oplus \frac{1}{k_{3,1}^1} \oplus \frac{1}{k_{1,3}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{1,w}^1 \oplus \frac{1}{k_{2,3}^1}} \oplus '02';$$

$$k_{1,3}^3 = k_{1,3}^2 \oplus k_{1,2}^3 = k_{1,3}^1 \oplus k_{1,1}^1 \oplus \frac{1}{k_{2,3}^1} \oplus \frac{1}{k_{3,2}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{2,w}^1 \oplus \frac{1}{k_{3,3}^1}};$$

$$k_{2,3}^3 = k_{2,3}^2 \oplus k_{2,2}^3 = k_{2,3}^1 \oplus k_{2,1}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{3,w}^1 \oplus \frac{1}{k_{0,3}^1}};$$

$$k_{3,3}^3 = k_{3,3}^2 \oplus k_{3,2}^3 = k_{3,3}^1 \oplus k_{3,1}^1 \oplus \frac{1}{k_{0,3}^1} \oplus \frac{1}{k_{3,0}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'}.$$

Выражение (8) связывает значения секретного ключа, открытого и закрытого текста после трех раундов шифрования системой из 16 алгебраических уравнений от 16 неизвестных.

Обобщим полученные выражения на большее число раундов.

На *i*-ом раунде система уравнений, связывающих значения секретного ключа, открытого и закрытого текста, будет иметь вид:

$$a_{0,j}^i = A_{0,j}^{i-1} \oplus k_{0,j}^i;\ a_{1,j}^i = A_{1,j}^{i-1} \oplus k_{1,j}^i;\ a_{2,j}^i = A_{2,j}^{i-1} \oplus k_{2,j}^i;\ a_{3,j}^i = A_{3,j}^{i-1} \oplus k_{3,j}^i, \quad (9)$$

где:

$$A_{0,j}^i = \frac{'02'}{A_{0,j}^{i-1} \oplus k_{0,j}^i} \oplus \frac{'03'}{A_{1,j+1}^{i-1} \oplus k_{1,j+1}^i} \oplus \frac{1}{A_{2,j+2}^{i-1} \oplus k_{2,j+2}^i} \oplus \frac{1}{A_{3,j+3}^{i-1} \oplus k_{3,j+3}^i},$$

$$A^i_{1,j} = \frac{1}{A^{i-1}_{0,j} \oplus k^i_{0,j}} \oplus \frac{'02'}{A^{i-1}_{1,j+1} \oplus k^i_{1,j+1}} \oplus \frac{'03'}{A^{i-1}_{2,j+2} \oplus k^i_{2,j+2}} \oplus \frac{1}{A^{i-1}_{3,j+3} \oplus k^i_{3,j+3}},$$

$$A^i_{2,j} = \frac{1}{A^{i-1}_{0,j} \oplus k^i_{0,j}} \oplus \frac{1}{A^{i-1}_{1,j+1} \oplus k^i_{1,j+1}} \oplus \frac{'02'}{A^{i-1}_{2,j+2} \oplus k^i_{2,j+2}} \oplus \frac{'03'}{A^{i-1}_{3,j+3} \oplus k^i_{3,j+3}},$$

$$A^i_{3,j} = \frac{'03'}{A^{i-1}_{0,j} \oplus k^i_{0,j}} \oplus \frac{1}{A^{i-1}_{1,j+1} \oplus k^i_{1,j+1}} \oplus \frac{1}{A^{i-1}_{2,j+2} \oplus k^i_{2,j+2}} \oplus \frac{'02'}{A^{i-1}_{3,j+3} \oplus k^i_{3,j+3}}.$$

Значения раундовых ключей (на четвертом раунде) выражаются следующими уравнениями:

$$k^4_{0,0} = k^3_{0,0} \oplus \mathrm{S}[k^3_{1,3}] \oplus '04' =$$

$$= k^1_{0,0} \oplus \frac{1}{k^1_{1,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{1,w} \oplus \frac{1}{k^1_{2,3}}} \oplus \frac{1}{k^1_{1,3} \oplus k^1_{1,1} \oplus \frac{1}{k^1_{2,3}} \oplus \frac{1}{k^1_{3,2}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{2,w} \oplus \frac{1}{k^1_{3,3}}}} \oplus '07';$$

$$k^4_{1,0} = k^3_{1,0} \oplus \mathrm{S}[k^3_{2,3}] = k^1_{1,0} \oplus \frac{1}{k^1_{2,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{2,w} \oplus \frac{1}{k^1_{3,3}}} \oplus \frac{1}{k^1_{2,3} \oplus k^1_{2,1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{3,w} \oplus \frac{1}{k^1_{0,3}}}};$$

$$k^4_{2,0} = k^3_{2,0} \oplus \mathrm{S}[k^3_{3,3}] =$$

$$= k^1_{2,0} \oplus \frac{1}{k^1_{3,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{3,w} \oplus \frac{1}{k^1_{0,3}}} \oplus \frac{1}{k^1_{3,3} \oplus k^1_{3,1} \oplus \frac{1}{k^1_{0,3}} \oplus \frac{1}{k^1_{3,0}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{0,w} \oplus \frac{1}{k^1_{3,1}}} \oplus '01'};$$

$$k^4_{3,0} = k^3_{3,0} \oplus \mathrm{S}[k^3_{0,3}] =$$

$$= k^1_{3,0} \oplus \frac{1}{k^1_{0,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{0,w} \oplus \mathrm{S}[k^1_{3,1}] \oplus '01'} \oplus \frac{1}{k^1_{0,3} \oplus k^1_{0,1} \oplus \frac{1}{k^1_{3,1}} \oplus \frac{1}{k^1_{1,3}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{1,w} \oplus \frac{1}{k^1_{2,3}}} \oplus '02'};$$

$$k^4_{0,1} = k^3_{0,1} \oplus k^4_{0,0} = k^1_{0,1} \oplus k^1_{0,0} \oplus \frac{1}{k^1_{1,3}} \oplus \frac{1}{k^1_{1,3} \oplus k^1_{1,1} \oplus \frac{1}{k^1_{2,3}} \oplus \frac{1}{k^1_{3,2}} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{2,w} \oplus \frac{1}{k^1_{3,3}}}} \oplus '05';$$

$$k^4_{1,1} = k^3_{1,1} \oplus k^4_{1,0} = k^1_{1,1} \oplus k^1_{1,0} \oplus \frac{1}{k^1_{2,3}} \oplus \frac{1}{k^1_{2,3} \oplus k^1_{2,1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k^1_{3,w} \oplus \frac{1}{k^1_{0,3}}}};$$

$$k_{2,1}^4 = k_{2,1}^3 \oplus k_{2,0}^4 = k_{2,1}^1 \oplus k_{2,0}^1 \oplus \frac{1}{k_{3,3}^1} \oplus \frac{1}{k_{3,3}^1 \oplus k_{3,1}^1 \oplus \frac{1}{k_{0,3}^1} \oplus \frac{1}{k_{3,0}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'}};$$

$$k_{3,1}^4 = k_{3,1}^3 \oplus k_{3,0}^4 = k_{3,1}^1 \oplus k_{3,0}^1 \oplus \frac{1}{k_{0,3}^1} \oplus \frac{1}{k_{0,3}^1 \oplus k_{0,1}^1 \oplus \frac{1}{k_{3,1}^1} \oplus \frac{1}{k_{1,3}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{1,w}^1 \oplus \frac{1}{k_{2,3}^1}} \oplus '02'};$$

$$k_{0,2}^4 = k_{0,2}^3 \oplus k_{0,1}^4 =$$

$$= k_{0,2}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{1,w}^1 \oplus \frac{1}{k_{2,3}^1}} \oplus k_{0,1}^1 \oplus \frac{1}{k_{1,3}^1 \oplus k_{1,1}^1 \oplus \frac{1}{k_{2,3}^1} \oplus \frac{1}{k_{3,2}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{2,w}^1 \oplus \frac{1}{k_{3,3}^1}}} \oplus '06';$$

$$k_{1,2}^4 = k_{1,2}^3 \oplus k_{1,1}^4 = k_{1,2}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{2,w}^1 \oplus \frac{1}{k_{3,3}^1}} \oplus k_{1,1}^1 \oplus \frac{1}{k_{2,3}^1 \oplus k_{2,1}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{3,w}^1 \oplus \frac{1}{k_{0,3}^1}}};$$

$$k_{2,2}^4 = k_{2,2}^3 \oplus k_{2,1}^4 =$$

$$= k_{2,2}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{3,w}^1 \oplus \frac{1}{k_{0,3}^1}} \oplus k_{2,1}^1 \oplus \frac{1}{k_{3,3}^1 \oplus k_{3,1}^1 \oplus \frac{1}{k_{0,3}^1} \oplus \frac{1}{k_{3,0}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'}};$$

$$k_{3,2}^4 = k_{3,2}^3 \oplus k_{3,1}^4 =$$

$$= \frac{1}{\bigoplus\limits_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'} \oplus \frac{1}{k_{0,3}^1 \oplus k_{0,1}^1 \oplus \frac{1}{k_{3,1}^1} \oplus \frac{1}{k_{1,3}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{1,w}^1 \oplus \frac{1}{k_{2,3}^1}} \oplus '02'};$$

$$k_{0,3}^4 = k_{0,3}^3 \oplus k_{0,2}^4 =$$

$$= k_{0,3}^1 \oplus k_{0,1}^1 \oplus \frac{1}{k_{3,1}^1} \oplus \frac{1}{k_{1,3}^1} \oplus k_{0,2}^1 \oplus k_{0,1}^1 \oplus \frac{1}{k_{1,3}^1 \oplus k_{1,1}^1 \oplus \frac{1}{k_{2,3}^1} \oplus \frac{1}{k_{3,2}^1} \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{2,w}^1 \oplus \frac{1}{k_{3,3}^1}}} \oplus '04';$$

$$k_{1,3}^4 = k_{1,3}^3 \oplus k_{1,2}^4 = k_{1,3}^1 \oplus \frac{1}{k_{2,3}^1} \oplus \frac{1}{k_{3,2}^1} \oplus k_{1,2}^1 \oplus \frac{1}{k_{2,3}^1 \oplus k_{2,1}^1 \oplus \frac{1}{\bigoplus\limits_{w=0}^{3} k_{3,w}^1 \oplus \frac{1}{k_{0,3}^1}}};$$

$$k_{2,3}^4 = k_{2,3}^3 \oplus k_{2,2}^4 = k_{2,3}^1 \oplus k_{2,2}^1 \oplus \frac{1}{k_{3,3}^1 \oplus k_{3,1}^1 \oplus \frac{1}{k_{0,3}^1} \oplus \frac{1}{k_{3,0}^1} \oplus \frac{1}{\bigoplus_{w=0}^{3} k_{0,w}^1 \oplus \frac{1}{k_{3,1}^1} \oplus '01'}};$$

$$k_{3,3}^4 = k_{3,3}^3 \oplus k_{3,2}^4 =$$

$$= k_{3,3}^1 \oplus k_{3,1}^1 \oplus \frac{1}{k_{0,3}^1} \oplus \frac{1}{k_{3,0}^1} \oplus \frac{1}{k_{0,3}^1 \oplus k_{0,1}^1 \oplus \frac{1}{k_{3,1}^1} \oplus \frac{1}{k_{1,3}^1} \oplus \frac{1}{\bigoplus_{w=0}^{3} k_{1,w}^1 \oplus \frac{1}{k_{2,3}^1}} \oplus '02'}.$$

Таким образом, в результате проведенных исследований получена алгебраических уравнений, связывающих значения открытого текста, секретного ключа и полученного шифр-текста. При этом был использован математический аппарат цепных дробей.

Для подтверждения корректности и достоверности полученных аналитических выражений в работе были проведены экспериментальные исследования, которые состояли в проверке полученных формул при перехвате разного количества крипто пар после осуществления следующего раунда. Экспериментально была проверена система уравнений (7), полученные результаты сведены в таблицу 1.

Табл. 1. – Число решений алгебраической системы уравнений (7)

Количество крипто пар	Число решений
1	4284867295
2	16771824
3	65291
4	267
5	1

Таким образом, при перехвате 5 крипто пар, существует только одно решение, которое удовлетворяет одновременно всем уравнениям системы (7). Эта оценка раскрывает сложность организации алгебраической атаки в виде числа пар «криптограмма-открытый текст», необходимых для однозначного восстановления секретного ключа шифрования.

4. Выводы

Проведенные исследования показали, что принятый в США национальный стандарт шифрования AES (FIPS-197) обладает специфической алгебраической структурой, которая может быть эффективно описана в терминах математического аппарата цепных дробей. В частности, удалось получить системы алгебраических уравнений (5-9), связывающих значения открытого текста, секретного ключа и полученного шифр-текста. Поиск неизвестных байтов ключа $k_{i,j}^u$ по известным байтам Состояния $a_{i,j}^u$ составляет задачу

криптографического анализа. В данном случае задача криптоанализа сведена к поиску решений системы нелинейных алгебраических уравнений.

Проведенные экспериментальные исследований подтвердили корректность аналитических выражений, кроме того, получена оценка сложности организации алгебраической атаки в виде числа пар «криптограмма-открытый текст», необходимых для однозначного восстановления секретного ключа шифрования. Перспективным направлением дальнейших исследований является разработка эффективных методов решения полученных систем нелинейных уравнений и их апробация на реальных тестовых примерах шифрования.

ЛИТЕРАТУРА

1. National Institute of Standards and Technology, “FIPS-197: Advanced Encryption Standard”, November 2001 [Электронный ресурс]. Режим доступа: http://csrc.nist.gov/publications/ fips/fips197/fips-197.pdf
2. Горбенко І.Д., Горбенко Ю.І. Прикладна криптологія. Монографія. Харків, ХНУРЕ, Форт, 2012 р., 1 та 2 видання, 868 с.
3. Есин В.И., Кузнецов А.А., Сорока Л.С. Безопасность информационных систем и технологий. Х.:ООО «ЭДЭНА», 2010. – 656с.
4. Кузнецов А.А., Иваненко Д.В., Костенко С.В. Алгебраическая структура шифра AES. Застосування інформаційних технологій у підготовці та діяльності сил охорони правопорядку: збірник тез доповідей. – Х.: Академія ВВ МВС України. – 2014. – С. 22 - 24.
5. T. Jakobsen and L.R. Knudsen, "The interpolation attack on block ciphers," Fast Software Encryption, LNCS 1267, E. Biham, Ed., Springer-Verlag, 1997, pp. 28-40.
6. Joan Daemen and Vincent Rijmen. AES proposal: Rijndael. In AES Round 1 Technical Evaluation, CD-1: Documentation. NIST, August 1998. See http://www.esat.kuleuven.ac.be/~rijmen/rijndael/ or http://www.nist.gov/aes.
7. Niels Ferguson, Richard Schroeppel, and Doug Whiting A simple algebraic representation of Rijndael // Selected Areas in Cryptography, Proc. SAC 2001, Lecture Notes in Computer Science #2259. — Springer Verlag, 2001. — P. 103–111.

Надійшла – 14.05.2014.

УДК 519.6

Решение обратной задачи интервального анализа поисковым методом

В. Ю. Дубницкий, А. М. Кобылин
Харьковский институт банковского дела УБС НБУ, Украина

Предложено для уменьшения неопределенности в процессе выполнения финансово-экономических расчетов использовать систему нестандартных интервальных арифметических операций при решении обратной задачи интервального анализа поисковым методом. Показана ее эффективность в сравнении с результатами аналогичных вычислений, выполненных на основе классической интервальной математики.

Ключевые слова: *Банковские операции, финансовая математика, интервальные вычисления, интервальный анализ.*

Запропоновано для зменшення невизначеності в процесі виконання фінансово-економічних розрахунків використовувати систему нестандартних інтервальних арифметичних операцій при вирішені зворотної задачі інтервального аналізу пошуковим методом. Показана її ефективність у порівнянні з результатами аналогічних обчислень, виконаних на основі класичної інтервальної математики.

Ключові слова: *Банківські операції, фінансова математика, інтервальні обчислення, інтервальний аналіз.*

In order to reduce uncertainty in financial and economic calculations, we propose to apply a system of non-standard interval arithmetic operations if search method is used for solving interval analysis inverse problem. The efficiency of such approach is shown as compared to results of similar calculations performed on the basis of classic interval mathematics.

Key words: *Banking operations, financial mathematics, interval calculations, interval analysis.*

1. Общая постановка задачи и её актуальность

Интервальный анализ как научное направление сформировался относительно недавно, в основном, как метод автоматического контроля ошибок округления на ЭВМ, обусловленный тем, что во многих вычислительных задачах возникла потребность не только вычисления приближенных решений, но и гарантированных оценок их близости к точным решениям. Ценность интервальных решений заключается в том, что они в целом позволяют получать наиболее достоверные решения исходных задач, учитывающие возможные диапазоны изменения исходных и вычисляемых значений [1, 2].

Из классической математики известно, что замкнутый числовой промежуток можно представить в виде интервала. Например, интервал между переменными $x_1 \in R$ и $x_2 \in R$ содержит все вещественные числа из множества R между x_1 и x_2, включая их самих, и обозначается как $[x_1, x_2]$. Соответственно, интервальную неопределенность можно понимать как состояние неполного (частичного) знания о какой-либо величине, когда возможно лишь указание ее принадлежности к данному интервалу. Иными словами, можно обозначить лишь границы возможных значений рассматриваемой величины (либо пределы ее изменения), и ширина получающегося интервала является естественной мерой интервальной неопределенности (неоднозначности). Выполнение

арифметических операций над величинами, имеющими интервальную неопределенность, приводит к интервальной неопределенности в ответе, и интервал результата должен содержать все возможные результаты выполнения операции над элементами исходных интервалов. Это значит, что в результате интервальных вычислений получающийся интервал гарантированно содержит множество всевозможных ответов «точечных» задач, данные к которым содержались в исходных интервалах.

Вычислительный эксперимент [1] на базе известных методов интервальной математики показал, что для сложных задач применение интервального анализа часто дает неудовлетворительные результаты из-за чрезмерной длины получаемых интервалов. Чаще всего это происходит из-за того, что «пессимистические» оценки точности оказываются на порядок хуже, чем реально достигаемая точность результатов [2, 3]. Кроме этого, возникает естественное противоречие между относительно большим диапазоном интервальных значений, отражающим низкую точность соответствующих значений, и предельно точным заданием границ интервалов □2-5□.

2. Истоки исследования авторов

Первая монография, полностью посвященная интервальному анализу, была опубликована Р. Е. Муром в 1966 г. [5]. В этой монографии были последовательно изложены основы нового направления в вычислительной математике, а также высказана точка зрения, что первым «интервальщиком» следует считать Архимеда, широко использовавшего в своих расчетах двусторонние приближения, в частности, для определения границ числа π(отношения длины окружности к ее диаметру). Предложенные Муром новые интересные постановки задач и поучительные применения интервальной техники оказали решающее влияние на становление и развитие нового научного направления во всем мире. Началом широкого распространения интервальных методов в компьютерных технологиях можно считать первый международный симпозиум по интервальному анализу, прошедший в Великобритании в январе 1968 года [6]. На русском языке первая достаточно известная работа по интервальным вычислениям была опубликована Ю.И. Шокиным в 1981 году [7]. «Интервальная идея» начала развиваться в XX веке в тесной связи с развитием и распространением практических инженерных вычислений. Но оформление интервального анализа в самостоятельную научную дисциплину стало возможным лишь с широким распространением ЭВМ. Последующие исследования показали, что методы интервального анализа могут служить не только для учета ошибок округления на ЭВМ, но и являются достаточно эффективными аналитическими методами для теоретических исследований [8]. Еще в 1931 году Р. Янг (Великобритания) [9] предложил арифметику для вычислений с множествами чисел. Американский ученый П. Двайер в 1951 г. рассматривал специальный случай замкнутых интервалов в связи с необходимостью учета погрешностей в численном анализе [10]. В 1956-58 гг. в Польше были опубликованы работы М. Вармуса [11] и Т. Сунаги[12], предложивших классическую интервальную арифметику и намечавшие ее приложения. При этом в работе Т. Сунаги [12] впервые были использованы и

современные термины «интервал», «интервальный». Кроме того, им были заложены основы интервального алгебраического формализма и даны весьма нетривиальные примеры применений новых методов, к примеру, в численном решении алгебраических уравнений и задачи Коши для обыкновенных дифференциальных уравнений [13]. В 1959- м году начал публиковать свои работы в области интервальных вычислений также и Р. Э. Мур, на базе чего к 1966 году была издана упомянутая выше монография [6].

В Советском Союзе «интервальную» историю можно отсчитывать с 20-х годов прошлого века, и связана она с именем видного русского математика Владимира Модестовича Брадиса, который широко известен своими математическими таблицами. В. М. Брадис предложил так называемый метод границ – способ организации вычислений, приводящий к достоверным двусторонним границам точного значения вычисляемого результата, фактически аналогичный интервальной арифметике. Работая в Тверском Педагогическом институте, он опубликовал целый ряд работ на эту тему [14-16]. В 1962-м году в одном из первых выпусков «Сибирского математического журнала» была опубликована статья Леонида Витальевича Канторовича [17], обозначившего эту тематику как одну из приоритетных для активно набирающей обороты вычислительной науки. В 1982 г. было издано учебное пособие Т. Н. Назаренко, Л.В. Марченко по интервальным методам [18], а в 1986 г. – монография С. Л. Калмыкова, Ю. И. Шокина, З. Х. Юлдашева [19]. Обширная и подробная библиография по интервальным анализу и вычислениям имеется, в частности, в работах [20-22]. К настоящему времени разработаны различные приемы интервальных вычислений [20, 22-25] и множество пакетов прикладных программ и алгоритмических макроязыков, реализующих элементы интервального анализа на машинном уровне для нескольких типов ЭВМ [24,25].

3. Нерешенные проблемы и цели работы

Одной из причин использования интервальных методов является то, что современные классические ЭВМ не учитывают, степень неточности большинства исходных данных. Даже невинно выглядящее дробное число 1/10 может порождать в определенных случаях вычислительную проблему, т. к. компьютер не может выполнять точные вычисления с этим числом [28]. В той мере, в какой точные вычислительные результаты используются для принятия критических решений, неучтенные ошибки вычислений означают повышенный риск. Очевидно, что чем сильнее зависимость точности входных данных от точности вычисляемых значений, чем более важной для последних является их корректность, и тем больше допускаемый риск. Например, широко известен такой классический пример, как катастрофа с американской зенитной ракетой Patriot 25 февраля 1991 года в Дхаране (Саудовская Аравия) [29]. Он показывает, что может произойти, если существующие вычисления с плавающей точкой будут и далее некритично применяться к новым задачам. В тот день батарея ракет Patriot не смогла перехватить иракскую ракету Scud, по официально названной причине: неадекватное вычисление в формате с плавающей точкой. Дело в том, что система управления ракеты Patriot имела внутренние системные часы, отсчитывающие время в десятых долях секунды, т. е. для перевода

времени в формат с целыми секундами компьютер просто умножает данные на 1/10. Однако, как уже упоминалось выше, на современных классических ЭВМ дробь 1/10 не имеет точного внутреннего представления, и должна быть приближена подходящей двоичной дробью. В качестве такого приближения американские разработчики взяли 24-битное двоичное число 0.00011001100110011001100, которое меньше, чем 1/10, примерно на одну миллионную. Эта на первый взгляд ничтожно малая погрешность постепенно накапливалась и, после четырех дней непрерывной работы расхождение системного времени с точным временем достигло 1/3 секунды, что, в конечном счете, привело к ошибке наведения в 700 метров. В результате ракета, выпущенная на перехват Scud, попала в помещение с американскими военнослужащими, убив 28 человек [29]. Вышеприведенный и подобные ему (не столь катастрофичные) случаи, наглядно демонстрируют, что являющиеся основой современных цифровых ЭВМ числа в формате с плавающей точкой, оказываются не вполне адекватными как реальному физическому миру, так и его математическим моделям, в частности, математическому понятию вещественного (действительного) числа. Основные недостатки современного представления чисел с плавающей точкой заключаются в следующем:

– большинство чисел вещественной оси не могут быть представлены точно числами с плавающей точкой, имеющими конечную длину мантиссы и, соответственно, свойства арифметических операций над числами с плавающей точкой отличаются (из-за неизбежных округлений) от свойств идеальных математических операций над вещественными числами;

– число в формате с плавающей точкой не несет никакой информации о точности той величины, которую оно представляет.

Получается, что существующая модель вычислений с плавающей точкой не предназначена ни для адекватного представления исходных значений, ни для отслеживания вычислительных ошибок. В связи с этим постепенно усиливается тенденция к переходу от точечных значений к интервальным, что влечет за собой стремительное развитие интервальной арифметики.

4. Используемые методы решения задач интервального анализа

Интервальный тип данных и интервальная арифметика реализуются на современных ЭВМ, как правило, с помощью представления интервала в виде пары чисел – одного для левого конца интервала, а другого для правого. При этом существующее аппаратное обеспечение, в частности, арифметика чисел с плавающей точкой, используются без каких-либо изменений, так как корректность получающейся интервальной арифметики может быть обеспечена так называемыми направленными округлениями. Например, там, где в задачах внешнего интервального оценивания в процессе вычислений требуется округление результата, нижняя граница интервала должна округляться вниз, а верхняя граница интервала – вверх. Таким образом, даже неизбежные ошибки округления при вычислениях с плавающей точкой будут строго и систематически учитываются в процессе выполнения интервальной программы. В качестве примера, на рис. 1 показано, как иррациональные числа в различных числовых шкалах представлены в виде различных интервалов (числовых

промежутков), причем наглядно проиллюстрировано изменение «ширины» этих интервалов. Интервалы чисел, представленных в вещественном формате (шкала floats) являются достаточными, так как в границы интервалов включены и ошибки округления исходных иррациональных чисел:

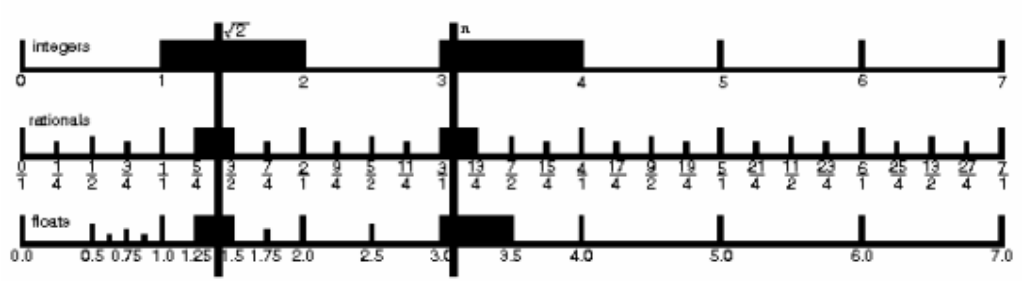


Рис. 1 – Представление иррациональных чисел в виде интервалов (integers – целые числа, rationales – рациональные числа, floats –вещественные числа) [28, с. 485].

Данный пример отражает основные положения интервального подхода к вычислениям на ЭВМ: исходные данные и промежуточные результаты представляются граничными значениями, над которыми и производятся все операции. При этом сами операции (прежде всего арифметические) определяются таким образом, что результат соответствующей точной операции обязательно лежит внутри вычисляемых границ.

$$\frac{1}{3} \in [0.33333, 0.33334]$$

$$\sqrt{2} \in [1.4142, 1.4143]$$

$$\pi \in [3.1415936, 3.1415937]$$

Традиционно аппаратное обеспечение компьютеров поддерживает две числовые системы: целые числа и числа с плавающей точкой. Целочисленная арифметика оперирует конечным подмножеством множества целых чисел и позволяет безошибочно осуществлять адресные вычисления, компиляцию и другие формы трансляции, а также реализовать различные алгоритмы типа поиска и сортировки [31].

Произвольное вещественное число представляется бесконечной систематической (например, десятичной или двоичной) дробью. На практике в научных и инженерных вычислениях вещественные числа приходится представлять в компьютере конечными дробями, чаще всего числами с плавающей точкой. Арифметика чисел с плавающей точкой поддерживается аппаратным обеспечением компьютеров и поэтому выполняется очень быстро, однако каждая операция с вещественным числом может вносить погрешности, накопление которых может существенно исказить результат [32].

Современные ЭВМ практически полностью базируются на двоичной логике и арифметике, обеспечивающих до недавнего времени практически все потребности компьютерных вычислений. Однако в 90-е годы прошлого века произошли качественные изменения, как в развитии логических основ, так и в

области компьютерных технологий, которые обусловили актуальность соответствующих изменений как в кодо-логическом [33], так и в алгоритмическом [34] базисе современных компьютерных технологий. Суть данных изменений может быть сведена к переходу от преобладания фиксированной точечной определенности к эволюционирующей множественности и неопределенности.

К недостаткам привычного, подхода можно отнеси, например, отсутствие ассоциативности в цепочке операций сложения и умножения. Выходит, что результат операций типа скалярного умножения будет разным в зависимости от особенностей компьютерного окружения — транслятора, процессора, выбранной разрядности, способов округления и т. д. Как следствие, выполнение одного и того же алгоритма в разных компьютерных окружениях приводит к различным, порой совершенно непохожим друг на друга, результатам.

Постепенно у критически настроенных исследователей все чаще стал возникать вопрос, вынесенный в заголовок обобщающей статьи немецкого математика проф. К. Никеля: «Can we trust the results of our computing?» («Можем ли мы доверять результатам наших вычислений?»). Действительно, беспристрастный анализ традиционного подхода к численным вычислениям и соответствующего инструментария (алгоритмов, языков программирования и аппаратного обеспечения), проведенный специалистами в области вычислительной математики, привел к неутешительному выводу о том, что алгоритм, сформулированный в привычных нам терминах, попросту не доопределён и потому обладает, вообще говоря, непредсказуемыми свойствами. (На одной из крупных научных конференций ректор Технического университета Вены проф. П. Скалички наполовину с юмором, а наполовину всерьез заявил, что с тех пор, как подробнее узнал о принятых способах выполнения машинных вычислений, очень опасается ходить по мостам и оказываться внутри других сложных инженерных сооружений…).

В рамках интервального подхода исходные данные и промежуточные результаты представляются граничными значениями, над которыми и производятся все операции. При этом сами операции (прежде всего арифметические) определяются таким образом, что результат соответствующей точной операции обязательно лежит внутри вычисляемых границ.

Приведем правила выполнения операций вещественной интервальной арифметики:

Арифметические операции с интервальными числами выполняют согласно правилам классической интервальной арифметики [19, 20]:

$$\boldsymbol{A}+\boldsymbol{B}=\left[\underline{\boldsymbol{a}};\overline{\boldsymbol{a}}\right]+\left[\underline{\boldsymbol{b}};\overline{\boldsymbol{b}}\right]=\left[\underline{\boldsymbol{a}}+\underline{\boldsymbol{b}};\overline{\boldsymbol{a}}+\overline{\boldsymbol{b}}\right]; \quad (1)$$

$$\boldsymbol{A}-\boldsymbol{B}=\left[\underline{\boldsymbol{a}};\overline{\boldsymbol{a}}\right]-\left[\underline{\boldsymbol{b}};\overline{\boldsymbol{b}}\right]=\left[\underline{\boldsymbol{a}}-\overline{\boldsymbol{b}};\overline{\boldsymbol{a}}-\underline{\boldsymbol{b}}\right]; \quad (2)$$

$$\boldsymbol{A}*\boldsymbol{B}=\left[\underline{\boldsymbol{a}};\overline{\boldsymbol{a}}\right]*\left[\underline{\boldsymbol{b}};\overline{\boldsymbol{b}}\right]=\left[\min\left\{\underline{\boldsymbol{a}}\cdot\underline{\boldsymbol{b}},\underline{\boldsymbol{a}}\cdot\overline{\boldsymbol{b}},\overline{\boldsymbol{a}}\cdot\underline{\boldsymbol{b}},\overline{\boldsymbol{a}}\cdot\overline{\boldsymbol{b}}\right\},\max\left\{\underline{\mathrm{a}}\cdot\underline{\boldsymbol{b}},\underline{\boldsymbol{a}}\cdot\overline{\boldsymbol{b}},\overline{\boldsymbol{a}}\cdot\underline{\boldsymbol{b}},\overline{\boldsymbol{a}}\cdot\overline{\boldsymbol{b}}\right\}\right]; \quad (3)$$

$$\boldsymbol{A}/\boldsymbol{B}=\left[\underline{\boldsymbol{a}};\overline{\boldsymbol{a}}\right]/\left[\underline{\boldsymbol{b}};\overline{\boldsymbol{b}}\right]=\left[\underline{\boldsymbol{a}},\overline{\boldsymbol{a}}\right]*\left[1/\overline{\mathrm{b}},1/\underline{\boldsymbol{b}}\right];\ 0\notin\boldsymbol{b} \quad (4)$$

Обоснование этих правил приведено в работе [19, 20].

Возникающая при вычислении границ погрешность учитывается с помощью направленных округлений: меньшая из вычисленных границ получается округлением до ближайшего машинного числа с недостатком, а большая – с избытком. Таким образом, интервальный подход позволяет единообразным способом учесть все виды погрешностей вычислительного процесса: приближенно известные исходные данные заключаются в гарантированно содержащие точное значение границы. Погрешности округлений лишь несколько расширяют границы промежуточных результатов, а сам вычислительный метод строится так, чтобы его погрешность также включалась в вычисленные границы конечного результата [20].

Полезно выписать определение интервального умножения в виде так называемой таблицы Кэли [21], дающей представление результата операции в зависимости от различных комбинаций значений операндов. Для этого выделим в ***IR*** следующие подмножества:

$$P := \left\{a \in IR \mid \underline{a} \geq 0 \text{ и } \overline{a} \geq 0\right\} \text{- неотрицательные интервалы} \quad (5)$$

$$Z := \left\{a \in IR \mid \underline{a} \leq 0 \leq \overline{a}\right\} \text{- нульсодержащие интервалы} \quad (6)$$

$$-P := \left\{a \in IR \mid -a \in P\right\} \text{- неположительные интервалы} \quad (7)$$

В целом ***IR*** = $P \cup Z \cup (-P)$. Тогда интервальное умножение (6) может быть описано с помощью Табл.1, особенно удобной при реализации этой операции на ЭВМ. В частности, при умножении интервала на число полезно помнить следующее простое правило:

$$\mu \cdot \boldsymbol{a} = \begin{cases} \left[\mu\underline{\boldsymbol{a}}, \mu\overline{\boldsymbol{a}}\right], & \text{если } \mu \geq 0 \\ \left[\mu\overline{\boldsymbol{a}}, \mu\underline{\boldsymbol{a}}\right], & \text{если } \mu < 0 \end{cases} \quad (8)$$

Таблица 1. Интервальное умножение

·	$b \in P$	$b \in Z$	$b \in -P$
$a \in P$	$[\underline{ab}, \overline{ab}]$	$[\overline{a}\underline{b}, \overline{ab}]$	$[\overline{a}\underline{b}, \underline{a}\overline{b}]$
$a \in Z$	$[\underline{a}\overline{b}, \overline{ab}]$	$[\min\{\underline{a}\overline{b}, \overline{a}\underline{b}\}, \max\{\underline{ab}, \overline{ab}\}]$	$[\overline{a}\underline{b}, \underline{ab}]$
$a \in -P$	$[\underline{a}\overline{b}, \overline{a}\underline{b}]$	$[\underline{a}\overline{b}, \underline{ab}]$	$[\overline{ab}, \underline{ab}]$

Как можно осуществить требуемое расширение классической интервальной арифметики? Здесь нам на выручку приходит абстрактная алгебра. С более общей точки зрения арифметика ***IR*** является коммутативной полугруппой как относительно сложения, так и относительно умножения. Известно (см., например, [38]), что всякая коммутативная полугруппа, в которой справедлив так называемый закон сокращения, может быть вложена в группу (или, что эквивалентно, расширена до группы), т. е. в действительно более богатую алгебраическую структуру, в которой каждый элемент имеет обратный. Интервальная арифметика ***IR*** как раз таки является коммутативной полугруппой, удовлетворяющей закону сокращения относительно сложения, а

относительно умножения полугруппу с законом сокращения образуют все интервалы, не содержащие нуля.

Все технические конструкции, необходимые для такого согласованного расширения интервальных полугрупп по сложению и умножению, были реализованы немецким исследователем Э. Каухером ещё в 70-е годы XX века. В работах [39, 40, 41] он построил алгебраическую систему, которую мы будем обозначать ***KR***, включающую в себя классическую интервальную арифметику ***IR*** как собственное подмножество. Она вполне удовлетворяет нашим требованиям, так как является группой по сложению и почти группой по умножению. Кроме того, в ***IR*** без каких-либо ограничений выполнимы операции взятия нижней и верхней граней относительно упорядочения интервалов по включению, т. е. ***KR*** обладает лучшими в сравнении с классической арифметикой ***IR*** порядковыми свойствами.

Э. Каухер при расширении ***IR*** опирался на свойство монотонности интервальных арифметических операций по включению и сохранил его в новой интервальной арифметике. Подчёркивая хорошие свойства новой алгебраической системы ***KR***, мы будем называть её полной интервальной арифметикой или, по имени её создателя, интервальной арифметикой Каухера.

Ещё одним замечательным свойством полной интервальной арифметики Каухера является то, что именно она является минимаксной интервальной арифметикой, в которой вычисление минимаксов может быть осуществлено на уровне сложения, вычитания, умножения и деления.

Элементами полной интервальной арифметики ***KR*** являются пары вещественных чисел $[\eta, \vartheta]$, не обязательно связанных соотношением $\eta \le \vartheta$. Таким образом, ***KR*** получается присоединением неправильных интервалов $[\eta, \vartheta]$, $\eta > \vartheta$, к множеству ***IR*** = $\{[\eta, \vartheta] \mid \eta, \vartheta \in \mathrm{R},\ \eta \le \vartheta\}$ правильных интервалов и вещественных чисел (отождествляемых с вырожденными интервалами нулевой ширины). Элементы арифметики Каухера и образуемые из них более сложные объекты (векторы, матрицы) мы будем выделять жирным шрифтом, как и обычные интервалы. При этом, если $\mathbf{a} = [\eta, \vartheta]$, то η называется левым (или нижним) концом интервала $\mathbf{a}$ и обозначается $\underline{\mathbf{a}}$ или **inf** $\mathbf{a}$, а ϑ называется правым (или верхним) концом интервала $\mathbf{a}$ и обозначается $\overline{\mathbf{a}}$ или **sup** $\mathbf{a}$. Как и прежде, интервал $\mathbf{a}$ назовём уравновешенным, если $\underline{\mathbf{a}} = -\overline{\mathbf{a}}$.

Определение 1. Абсолютной величиной (модулем) интервала a называется величина

$$|\boldsymbol{a}| = \max\left\{|\underline{\boldsymbol{a}}|, |\overline{\boldsymbol{a}}|\right\} \tag{9}$$

Правильные и неправильные интервалы, две половинки ***KR***, меняются местами в результате отображения дуализации dual: $\boldsymbol{KR} \to \boldsymbol{KR}$, меняющего местами (переворачивающего) концы интервала, т. е. такого что

$$\mathrm{dual}\,\boldsymbol{a} := \left[\overline{\boldsymbol{a}}, \underline{\boldsymbol{a}}\right] \tag{10}$$

Правильной проекцией интервала $\mathbf{a}$ называется величина

$$\text{pro}\, \boldsymbol{a} := \begin{cases} \boldsymbol{a}, & \text{если } \boldsymbol{a} \text{ - правильный,} \\ \text{dual}\, \boldsymbol{a}, & \text{иначе.} \end{cases} \tag{11}$$

Аналогично классической интервальной арифметике ***IR*** отношение включения одного интервала в другой определяется в ***KR*** следующим образом:

$$\boldsymbol{a} \subseteq \boldsymbol{b} \Leftrightarrow \underline{\boldsymbol{a}} \geq \underline{\boldsymbol{b}} \quad \text{и} \quad \overline{\boldsymbol{a}} \leq \overline{\boldsymbol{b}} \tag{12}$$

Например, $[3, 1] \subseteq [2, 2] = 2 \in \mathrm{R}$.

Помимо теоретико-множественного включения на множестве интервалов ***KR*** существует ещё одно частичное упорядочение, которое естественно обобщает линейный порядок на вещественной оси.

Определение 2. Для интервалов $\mathbf{a}$, $\mathbf{b} \in \boldsymbol{KR}$ условимся считать, что **a** не превосходит **b** и писать $\mathbf{a} \leq \mathbf{b}$ тогда и только тогда, когда $\underline{\mathbf{a}} \leq \underline{\mathbf{b}}$ и $\overline{\mathbf{a}} \leq \overline{\mathbf{b}}$. Интервал называется неотрицательным, т.е. «≥ 0», если неотрицательны оба его конца. Интервал называется неположительным, т.е. «≤ 0», если неположительны оба его конца.`

Пример. [1, 2] $\leq$ [3, 2], причём оба сравниваемых интервала [1, 2] и [3, 2] неотрицательны.

Сложение определяется в ***KR*** совершенно так же, как в классической интервальной арифметике:

$$\boldsymbol{a} + \boldsymbol{b} := \left[\underline{\boldsymbol{a}} + \underline{\boldsymbol{b}}, \overline{\boldsymbol{a}} + \overline{\boldsymbol{b}}\right] \tag{13}$$

Но теперь из факта существования неправильных интервалов следует то, что каждый элемент a из ***KR*** имеет единственный обратный по сложению (противоположный), обозначаемый через **opp a,** и из равенства **a** + **opp a** = **0** следует

$$\text{opp}\, \boldsymbol{a} := \left[-\underline{\boldsymbol{a}}, -\overline{\boldsymbol{a}}\right] \tag{14}$$

Таким образом, относительно сложения арифметика ***KR*** является коммутативной группой, изоморфной аддитивной группе стандартного линейного пространства R^2. Для краткости мы будем обозначать операцию, обратную сложению, так называемое внутреннее (алгебраическое) вычитание в ***KR***, через, $\oplus$ так что

$$\boldsymbol{a} \oplus \boldsymbol{b} := \boldsymbol{a} + \text{opp}\, \boldsymbol{b} = \left[\underline{\boldsymbol{a}} - \underline{\boldsymbol{b}}, \overline{\boldsymbol{a}} - \overline{\boldsymbol{b}}\right] \tag{15}$$

Для того чтобы выписать явные формулы для умножения в полной интервальной арифметике, выделим в ***KR*** следующие подмножества:

$$\boldsymbol{P} := \left\{\boldsymbol{a} \in \boldsymbol{KR} \mid \left(\underline{\boldsymbol{a}} \geq 0\right) \& \left(\overline{\boldsymbol{a}} \geq 0\right)\right\} - \text{ неотрицательные интервалы} \tag{16}$$

$$\boldsymbol{Z} := \left\{\boldsymbol{a} \in \boldsymbol{KR} \mid \underline{\boldsymbol{a}} \leq 0 \leq \overline{\boldsymbol{a}}\right\} - \text{нульсодержащие интервалы} \tag{17}$$

$$-\boldsymbol{P} := \left\{\boldsymbol{a} \in \boldsymbol{KR} \mid -\boldsymbol{a} \in \boldsymbol{P}\right\} - \text{неположитедьные интервалы} \tag{18}$$

$$\boldsymbol{dual\ Z} := \left\{\boldsymbol{a} \in \boldsymbol{KR} \mid \boldsymbol{dual} \in \boldsymbol{Z}\right\} - \text{ интервалы, содержащиеся в нуле} \tag{19}$$

В целом ***KR*** = $\mathrm{P} \cup \mathrm{Z} \cup (-\mathrm{P}) \cup (\text{dual Z})$. Тогда умножение в интервальной арифметике Каухера может быть описано таблицей [5], которая получается

дополнением таблицы 1 ещё одной строкой и ещё одним столбцом, соответствующем случаю операндов из множества dual Z.

Замечание. Умножение в арифметике Каухера допускает нетривиальные делители нуля. Например, [–1,2]•[5,–3]=0. Интервальное умножение в арифметике Каухера оказывается коммутативным и ассоциативным [39, 40, 41], но группу по умножению в ***KR*** образуют лишь интервалы $\mathbf{a}$, для которых $\underline{\mathbf{a}}\,\overline{\mathbf{a}} > 0$, поскольку закон сокращения не выполняется ни на каком более широком подмножестве ***KR***.

Рассмотренные методы интервального анализа имеют достаточно развитые методы для решения многих задач, но общий недостаток этих методов – широкие интервальные оценки результата, что иногда неприменимо не только для проведения практических расчетов, но и для дальнейшего анализа.

Таблица 2. Таблица Кэли для операции умножения в интервальной арифметике Каухера

	$b \in P$	$b \in Z$	$b \in -P$	$b \in dual\ Z$
$a \in P$	$\left[\underline{a}\underline{b}, \overline{a}\overline{b}\right]$	$\left[\overline{a}\underline{b}, \overline{a}\overline{b}\right]$	$\left[\overline{a}\underline{b}, \underline{a}\overline{b}\right]$	$\left[\underline{a}\underline{b}, \underline{a}\overline{b}\right]$
$a \in Z$	$\left[\underline{a}\overline{b}, \overline{a}\overline{b}\right]$	$\left[\min\left\{\underline{a}\overline{b}, \overline{a}\underline{b}\right\}, \max\left\{\underline{a}\underline{b}, \overline{a}\overline{b}\right\}\right]$	$\left[\overline{a}\underline{b}, \underline{a}\underline{b}\right]$	0
$a \in -P$	$\left[\underline{a}\overline{b}, \overline{a}\underline{b}\right]$	$\left[\underline{a}\overline{b}, \underline{a}\underline{b}\right]$	$\left[\overline{a}\overline{b}, \underline{a}\underline{b}\right]$	$\left[\overline{a}\overline{b}, \overline{a}\underline{b}\right]$
$a \in dual\ Z$	$\left[\underline{a}\underline{b}, \overline{a}\underline{b}\right]$	0	$\left[\overline{a}\overline{b}, \underline{a}\overline{b}\right]$	$\left[\max\left\{\underline{a}\underline{b}, \overline{a}\overline{b}\right\}, \min\left\{\underline{a}\overline{b}, \overline{a}\underline{b}\right\}\right]$

В работе [42] приведена структура, которая получила название системы правил нестандартной интервальной математики. Обозначим $\boldsymbol{M} = \left(\boldsymbol{I}(\boldsymbol{R}), +, -, \times, /, +^{-}, -^{-}, \times^{-}, /^{-}\right)$, где $\boldsymbol{I}(\boldsymbol{R}) = \left\{\left[\boldsymbol{a}^{-}, \boldsymbol{a}^{+}\right] \mid \boldsymbol{a}^{-} \le \boldsymbol{a}^{+}, \boldsymbol{a}^{-}, \boldsymbol{a}^{+} \in \boldsymbol{R}\right\}$- множество действительных интервалов; $(+, -, \times, /)$ и $\left(+^{-}, -^{-}, \times^{-}, /^{-}\right)$- стандартные и нестандартные интервальные операции сложения, вычитания, произведения и деления соответственно действительным интервалам.

Для программной реализации представим значения интервальных чисел $\boldsymbol{A}$ и $\boldsymbol{B}$ в форме центр-радиуса $\boldsymbol{A} = \langle \boldsymbol{a}, \boldsymbol{r}_a \rangle$, $\boldsymbol{B} = \langle \boldsymbol{b}, \boldsymbol{r}_b \rangle$, где

$$\boldsymbol{a} = \frac{\underline{\boldsymbol{a}} + \overline{\boldsymbol{a}}}{2}, \quad \boldsymbol{r}_a = \frac{\overline{\boldsymbol{a}} - \underline{\boldsymbol{a}}}{2} \quad \boldsymbol{b} = \frac{\underline{\boldsymbol{b}} + \overline{\boldsymbol{b}}}{2}, \quad \boldsymbol{r}_b = \frac{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}{2} \tag{20}$$

- центры и радиусы соответственно интервалов $\boldsymbol{A}$ и $\boldsymbol{B}$.

Нестандартная интервально-арифметическая операция сложения определяется так:

$$\boldsymbol{A} +^{-} \boldsymbol{B} = \langle \boldsymbol{a} + \boldsymbol{b}, \mid \boldsymbol{r}_a - \boldsymbol{r}_b \mid \rangle \tag{21}$$

Нестандартная интервально-арифметическая операция вычитания определяется так:

$$A -^{-} B = \langle a - b, | r_a - r_b | \rangle \quad (22)$$

Нестандартная интервально-арифметическая операция произведения определяется так:

$$A \times^{-} B = \langle ab - \mathrm{sgn}(ab) r_a r_b, | ar_b - \mathrm{sgn}(ab) b r_a | \rangle \text{ , если } \frac{|a|}{r_a} \geq 1, \frac{|b|}{r_b} \geq 1 \quad (23)$$

$$A \times^{-} B = \langle ab - \mathrm{sgn}(b) a r_b, | br_a - \mathrm{sgn}(b) r_a r_b | \rangle \text{ , если } \frac{|a|}{r_a} < 1, \frac{|a|}{r_a} < \frac{|b|}{r_b} \quad (24)$$

$$A \times^{-} B = \langle ab - \mathrm{sgn}(a) b r_b, | ar_a - \mathrm{sgn}(a) r_b r_b | \rangle \text{ , если } \frac{|b|}{r_b} < 1, \frac{|a|}{r_a} \geq \frac{|b|}{r_b} \quad (25)$$

При умножении интервала на число применяется такое правило:

$$\mu \cdot a = \begin{cases} \left[\mu \cdot \underline{a}, \mu \cdot \overline{a}\right], \textit{если} & \mu \geq 0, \\ \left[\mu \cdot \overline{a}, \mu \cdot \underline{a}\right], \textit{если} & \mu < 0. \end{cases} \quad (26)$$

Нестандартная интервально-арифметическая операция деления определяется так:

$$A /^{-} B = \frac{1}{b^2 - r_b^2} \langle ab - \mathrm{sgn}(ab) r_a r_b, | ar_b - \mathrm{sgn}(ab) b r_a | \rangle, \text{ если } \frac{|b|}{r_b} > 1, \frac{|a|}{r_a} \geq 1 \quad (27)$$

$$A /^{-} B = \frac{1}{b^2 - r_b^2} \langle ab - \mathrm{sgn}(b) a r_b, | br_a - \mathrm{sgn}(b) r_a r_b | \rangle, \text{ если } \frac{|b|}{r_b} > 1, \frac{|a|}{r_a} < 1 \quad (28)$$

$$A /^{-} B = \frac{1}{b^2 - r_b^2} \langle ab - \mathrm{sgn}(a) b r_a, | ar_b - \mathrm{sgn}(a) r_a r_b | \rangle, \text{ если } \frac{|b|}{r_b} < 1, \frac{|a|}{r_a} < 1 \quad (29)$$

При делении интервала на число применяется такое правило:

$$\mu / a = \begin{cases} \left[\mu \cdot \frac{1}{\underline{a}}, \mu \cdot \frac{1}{\overline{a}}\right], \textit{если} & \mu \geq 0, \\ \left[\mu \cdot \frac{1}{\overline{a}}, \mu \cdot \frac{1}{\underline{a}}\right], \textit{если} & \mu < 0. \end{cases} \quad (30)$$

5. План экспериментов и оценка точности численного решения рассмотренных аксиом интервальной математики

Авторами разработаны программные системы для решения практических задач, которые использовали следующие типы интервальной математики:

- классическая интервальная математика;
- интервальная математика на основе таблиц Кели;
- полная интервальная математика Каухера;
- нестандартная интервальна математика.

Результаты проведенных сравнительных вычислений сведены в таблицу 3.

Табл.3. Фрагмент таблицы сравнительных результатов вычислительного эксперимента для различных выражений перечисленными аксиомами интервальной математики при x=[1,2], y=[-1,0]

Выражение	Выражение в интервальном виде	Классика	Кели	Каухера	Нестандартная
$\frac{2x+3y}{4x-y}$	$\frac{[2,4]+[-3,0]}{[4,8]-[-1,0]}$	[-0,2, 0,8]	[-0,2, 0,8]	[0,25, 1,0]	[0,20, 0,25]
$\frac{x+3y}{2x-y}$	$\frac{[1,2]+[-3,0]}{[2,4]-[-1,0]}$	[-1,0, 1,0]	[-1,0, 1,0]	[-1,0 , 1,0]	[-0,25, 0,25]

Результаты вычислительных экспериментов представлены на рис.2 и 3.

Рис. 2. Сравнение ширина интервала для рассмотренных выражений в классической интервальной математике

*Рис.3. Сравнение ширина интервала для рассмотренных выражений в нестандартной интервальной математик*е

6. Вычислительный эксперимент: обоснование эффективности рассмотренных аксиом интервальной математики ее и реализация

Согласно работы [21] ширину интервала, определяющего число $A=\left[\underline{a},\overline{a}\right]$ вычисляют по формуле:

$$\Delta A = \overline{a} - \underline{a}\,, \tag{31}$$

середину интервала определяем по формуле:

$$m(A) = \frac{1}{2}\ (\overline{a} + \underline{a})\,, \tag{32}$$

тогда точность интервала ε определим следующим образом:

$$\varepsilon_{кл} = \Delta A / m(A)\,. \tag{33}$$

Для нестандартной интервальной математики это будет соответствовать условию:

$$\varepsilon_{нст} = r_a / a\,. \tag{34}$$

Эффективность ***ef*** предложенного процесса будем определять мерой уменьшения интервала окончательного результата, определенного с использованием нестандартной интервальной математики в сравнении с аналогичным, но определенным с использованием классической интервальной математики :

$$\boldsymbol{ef} = (1 - \frac{\boldsymbol{\varepsilon_{нст}}}{\boldsymbol{\varepsilon_{кл}}})100\%\,. \tag{35}$$

В таблице 4 представлены результаты численного эксперимента для финансовых функций по определению накопленных сумм на основе простой и учетной ставок разными методами интервальной математики.

Таблиця 4.Сравнение эффективности вычисления финансовых расчетов методами классической и нестандартной интервальной математики

Показатель	Методика расчета				Параметры эффективности		
	классическая интервальная математика		нестандартная интервальная математика				
	$\underline{a}$	$\overline{a}$	$\underline{a}$	$\overline{a}$	$\varepsilon_{кл}$	$\varepsilon_{нст}$	Ef
Накопленная сумма на основе простой процентной ставки	2325	2520	2362,5	2480	0,08	0,05	37
Накопленная сумма на основе сложной процентной ставки	2527,5	2757,3	2585,0	2696,0	0,08	0,04	50
Накопленная сумма на основе простой учетной ставки	5555,5	6000,0	5666,6	5882,3	0,07	0,04	43
Накопленная сумма на основе сложной учетной ставки	4477,0	4697	4566,5	4604,9	0,05	0,01	80

Приведенные расчеты показывают, что применение нестандартной интервальной математики существенно сужает интервал существования результата вычислений, т.е. эта система более эффективна, чем система, построенная на основе классической интервальной математики.

7. Решение обратной задачи интервального анализа поисковым методом

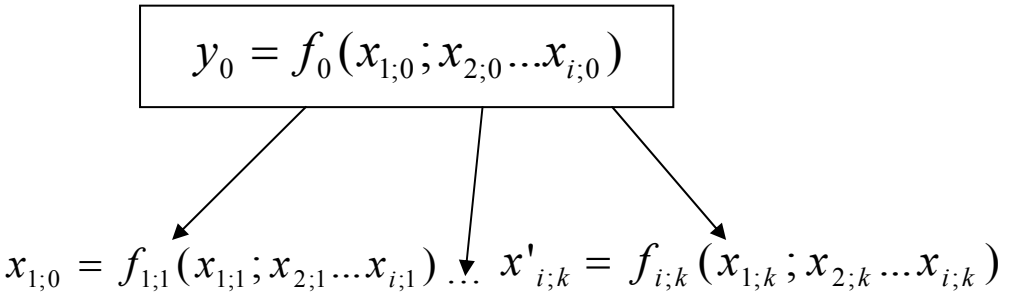
Назовем прямой задачею финансового анализа вычисления значения некоторого показателя, который является функцией от нескольких аргументов. Особенностью методов финансового анализа есть многоразовая суперпозиция функций, обусловленная особенностями расчетов показателей и учетных методов.

Обратная задача возникает при планировании деятельности фирмы, когда по заданному желательному значения окончательного показателя необходимо подобрать соответствующие значения аргументов.

Для решения задачи введем следующие обозначения. Итоговый показатель будем считать корнем дерева. Представим его в следующем виде:

$$y_0 = f_0(x_{1;0}; x_{2;0} \ldots x_{n;0}), \quad (36)$$

где y_0 - финансовый показатель первого уровня (итоговый), $x_{1;0}; x_{2;0} \ldots x_{n;0}$ - аргументы первого уровня. Построение дерева покажем на рисунке:4



(далее продолжать по аналогичной схеме).

Рис. 4. Построение дерева формул

Будем считать, что на каждом уровне в зависимости от вида $f_{i;k}(x_{1;0}; x_{2;0} \ldots x_{i;k})$, т.е. зависимость і-ого уравнения переменной входят все переменные данного уровня. Часть из них может быть нулевым.

Предлагается последовательность действий, состоящую со следующих шагов:

Шаг 1: Задаем желаемое значение итогового показателя y_0.

Шаг 2: Задаем допустимый интервал неопределенности пошагового показателя, т. е. интервалы изменения переменных нулевого уровня.

Шаг 3: Используя интервальный анализ вычисляем допустимый интервал для y_0.

Если $y_0 \in [y_н; y_в]$, где $y_н$, $y_в$ – левая и правая граница допустимых значений, y_0 то задача имеет решения.

Если $y_0 \notin [y_н; y_в]$, то задача не имеет решения.

Шаг 4: (этот шаг используется как общий шаг алгоритма) Если задача не имеет решения, то необходимо изменить интервалы неопределенности аргументов нулевого уровня. Если задача имеет решения, то подбор значений переменных начинают с уровня дерева с наибольшим номером.

Для подбора значений переменных используют особенность, что $ЛП_\tau$ – последовательность может быть заменена случайными равномерно распределенными числами.

Для каждой переменной каждого уровня возьмем ее допустимый интервал расчета, вычислим равномерно распределенные случайные числа.

Шаг 5: Для выбранных случайных чисел вычисляется соответствующие показатели и выбираются те значения, для которых модуль отклонения расчетного значения от желаемого не превышает заданной точности.

Шаг 6: Имея вычисленные интервалы неопределенности вычисляют интервал соответствующего показателя для уровня дерева с наименьшим номером.

Эти процедуры продолжаются до достижения нулевого уровня.

Если задача не решена, то процесс начинается с шага 1.

Пример. имеем число 100, которое складывается с двоих чисел 80 та 20, которые в свою очередь складываются с 20*4 и 10*2 соответственно (рис.5). Как необходимо изменить исходные числа, чтобы получить 105, если диапазон интервала каждого элемента нижнего уровня 5%?

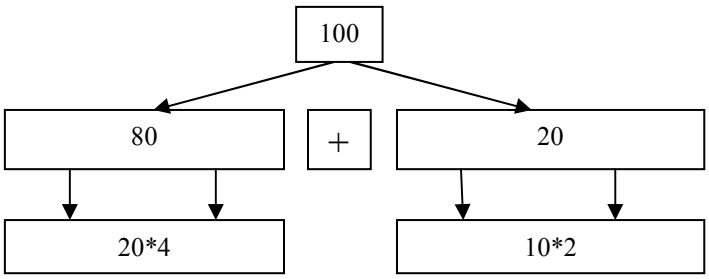


Рис. 5. Условия примера метода обратной задачи

Пояснения к рисунку: От значения каждого элемента вычитанием и сложением 5% получаем его интервальное значение и занесем их в таблицу. Следующим шагом найдем столбец E и F. Для этого проведем необходимые арифметические операции, в нашем случае это умножение. Далее находим столбец G. Его значение определяется сложением столбца E и F. В столбце H находим именно ту комбинацию столбцов A, B, C, и D, которая ближе всего находится до значения 105 (рис.6).

Обратная задача возникает при планировании деятельности фирмы, когда задан желаемый уровень значения итогового показателя, по которому необходимо подобрать соответствующие значения аргументов.

Рассмотрим задачу прогнозирования уровня рентабельности производства и решим ее методом интервальной обратной задачи. Условие задачи представлено деревом формул (рис. 7)

	A	B	C	D	E	F	G	H
1	**20**	**4**	**10**	**2**	**A*B**	**C*D**	**E+F**	**G-105**
2	19	3,8	9,5	1,9	72,2	18,05	90,25	-14,75
3	19,1	3,82	9,55	1,91	72,962	18,2405	91,2025	-13,7975
4	19,2	3,84	9,6	1,92	73,728	18,432	92,16	-12,84
5	19,3	3,86	9,65	1,93	74,498	18,6245	93,1225	-11,8775
6	19,4	3,88	9,7	1,94	75,272	18,818	94,09	-10,91
7	19,5	3,9	9,75	1,95	76,05	19,0125	95,0625	-9,9375
8	19,6	3,92	9,8	1,96	76,832	19,208	96,04	-8,96
9	19,7	3,94	9,85	1,97	77,618	19,4045	97,0225	-7,9775
10	19,8	3,96	9,9	1,98	78,408	19,602	98,01	-6,99
11	19,9	3,98	9,95	1,99	79,202	19,8005	99,0025	-5,9975
12	20	4	10	2	80	20	100	-5
13	20,1	4,02	10,05	2,01	80,802	20,2005	101,0025	-3,9975
14	20,2	4,04	10,1	2,02	81,608	20,402	102,01	-2,99
15	20,3	4,06	10,15	2,03	82,418	20,6045	103,0225	-1,9775
16	20,4	4,08	10,2	2,04	83,232	20,808	104,04	-0,96
17	20,5	4,1	10,25	2,05	84,05	21,0125	105,0625	0,0625
18	20,6	4,12	10,3	2,06	84,872	21,218	106,09	1,09
19	20,7	4,14	10,35	2,07	85,698	21,4245	107,1225	2,1225
20	20,8	4,16	10,4	2,08	86,528	21,632	108,16	3,16
21	20,9	4,18	10,45	2,09	87,362	21,8405	109,2025	4,2025
22	21	4,2	10,5	2,1	88,2	22,05	110,25	5,25

Рис. 6. Пример расчета методом обратной задачи. Ответ находится в строке 17 в рис. 6.

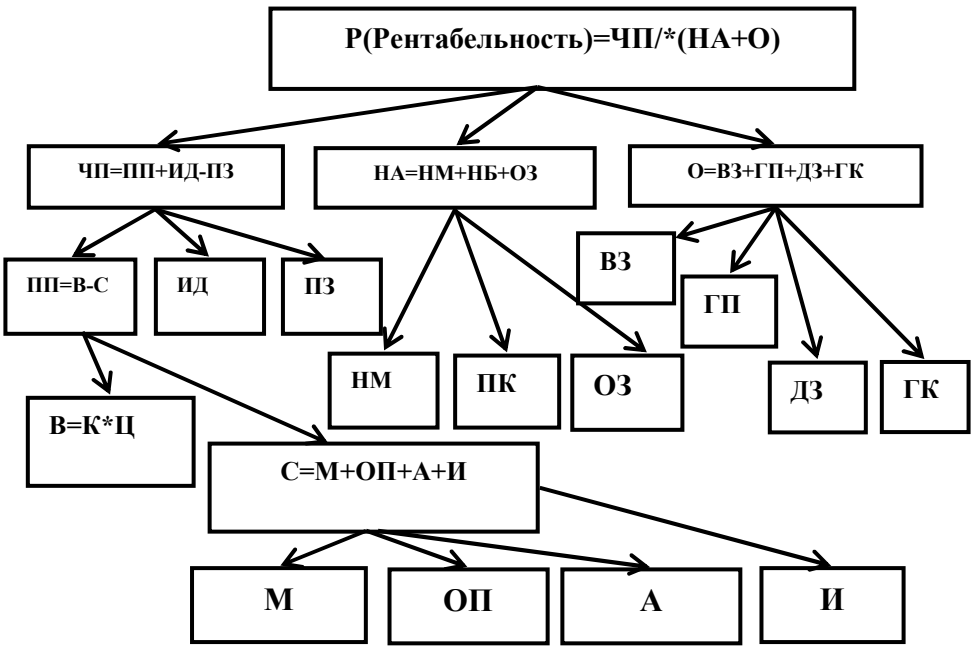


Рис. 7. Дерево показателей расчета рентабельности

Здесь: В=50327,3 грн.; М=32156 грн.; ОП=3278,7 грн.; А=1063,2 грн.; И= 589 грн.; ИД=5184,1 грн.; ПЗ=8884,6 грн.; ДФИ=1327,7 грн.; НМ=8,9 грн.; НБ=468,9 грн.; ОЗ=16456,7 грн.; ВЗ=992 грн.; ГП=1185,2 грн.; Т=7,4 грн.; ИАО=738,6 грн.; ГК=10,8 грн.; ДЗ=11102,8 грн..

Задача исследования: повысить уровень рентабельности на 7,5%, рассчитав рентабельность за формулою (37) і приведением к окончательному виду (37).

$$P = ЧП/(НА + О) \qquad (37)$$

Программная реализация системы расчета численных значений показателей финансовой деятельности фирмы, которые обеспечивают необходимый уровень прибыли предприятия, разработана на языке программирования VBA выполнена в среде MS Excel 2007 с разработкой макросов для решения как прямой так и обратной задачи.

8 Выводы по результатам и направления дальнейших исследований

1. Для уменьшения интервала заключительного результату выполнения финансовых расчетов предложено использовать правила нестандартной интервальной математики .
2. Разработана программную систему, которая реализует правила нестандартной интервальной математики.
3. Показано, что правила нестандартной интервальной математики дают возможность получить результирующий интервал на 37- 80 процентов меньше, чем аналогичный, но определенный согласно с правилами классической интервальной математики.
4. Обоснованы теоретические основы решение обратной задачи интервального анализа поисковым методом и целесообразность их использования в системе задач экономического факторного анализа.

ЛИТЕРАТУРА

1. Дубницький В.Ю., Кобилін А.М. Порівняльний аналіз результатів планування нормативів банківської безпеки засобами класичної та нестандартної інтервальної математики. / Радіоелектронні і комп'ютерні системи. №5(69) Науково-технічний журнал. Харків: «ХАІ» 2014. стр.29-33.
2. Добронец Б. С. Интервальная математика: Учеб. Пособие. Краснояр. гос. ун-т. – Красноярск. 2004. – 216 с.
3. Марчук Г.И. Методы вычислительной математики. – М.: Наука. 1977. – 456с.
4. Самарский А.А. Введение в теорию разностных схем. – М.: Наука, 1971. – 552с.
5. Moore R.E. Intervalanalysis. Eiiglewood Cliffs / R.E. Moore – N.J.:Prentic-e-llall, 1966.
6. Hansen E. Topics in Interval Analysis. – London: Oxford UniversityPress, 1969.
7. Шокин Ю. И. Интервальный анализ. / Ю. И. Шокин – Новосибирск: Наука, 1981. – 111 с.
8. Интервальный анализ и его приложения. Исторические заметки, http://www.sbras.ru/interval/index.php?j=Introduction/history.
9. Young R.C. Algebra of many-valued quantities // Mathematische Annalen / R. C. Young, 1931. S. 260-290.

10. Dwyer P.S. Linear Computations / P. S. Dwyer – New York: John Wiley & Sons, 1951. – 36 p.
11. Warmus M. Calculus of approximations // Bull. Acad. Polon. Sci./ M. Warmus – 1956, Cl. III, vol. IV, No. 5.
12. Sunaga T. Theoryof an interval algebra and its application to numerical analysis // RAAG Memoirs. – Vol. 2, Misc. II, 1958.
13. Markov S., Okumura K. The contribution of T.Sunaga to interval analysis and reliable computing // Developments in Reliable Computing / Cendes T., ed. – Dordrecht: Kluwer Academic Publishers, 1998.
14. Брадис В.М. Опыт обоснования некоторых практических правил действий над приближенными числами // Известия Тверского педагогического института. 1927. – Вып. 3.
15. Брадис В.М. Теория и практика вычислений. Пособие для высших педагогических учебных заведений. – Москва:Учпедгиз, 1937.
16. Брадис В.М. Средства и способы элементарных вычислений. – Москва: Издательство Академии педагогических наук РСФСР, 1948.
17. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сибирский Математический Журнал. – 1962. – Т. 3, №. 5.
18. Назаренко Т.И., Марченко Л.В. Введение в интервальные методы вычислительной математики. / Т.И. Назаренко, и др. – Иркутск: Изд-во Иркут. ун-та. 1982.
19. Калмыков С.Л., Шокин Ю.И., Юлдашев З.Х. Методы интервального анализа. /О.Л. Калмыкин и др. – Новосибирск: Наука. 1986. – 224 с.
20. Алефельд Г., Херцбергер Ю. Введение и интервальные вычисления. / Г. Алефельд и др. – М.: Мир, 1987. – 360 с.
21. Шарый С.П. Конечномерный интервальный анализ. – Новосибирск, Институт вычислительных технологий СО РАН, 2009. – 569 с.
22. Beierbaum, F., Schwierz. K.P. A bibliography on interval mathematics // J. Comput. Appl. Math. V. 4, N 1. P.59-86.
23. Moore R.E. Methods and Applications of Interval Analysis. SIAM. Philadelphia 1979.
24. Клатте Р., Кулиш У., Неага М., Рац Д., Ульрих Х. PASCAL-XSC Язык численного программирования. / Р. Клатте и др. – М.: ДМК Пресс, 2000.
25. Агеев М. П., Алик И. П., Марков Ю. И. Алгоритм 616. Процедуры интервальной математики // Библиотека алгоритмов 516-11/06. / М. П. Агеев и др. – М.: Сов. Радио, 1976.
26. Interval arithmetic - From Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Interval_arithmetic
27. IEEE Interval Standard Working Group - P1788. http://grouper.ieee.org/groups/1788/
28. Hayes B. A Lucid Interval. A reprint from American Scientist the magazine of Sigma Xi, the Scientific Research Society, Volume 91, Number 6, November–December, 2003, p.484-488.
29. Строгий учет ошибок округлений на цифровых ЭВМ, http://www.sbras.ru/interval/index.php?j=Introduction/RusIntro.

30. Кулиш У., Рац Д., Хаммер Р., Хокс М. Достоверные вычисления. Базовые численные методы М.: «Регулярная и хаотическая динамика», 2005.
31. Кнут Д. Е. Искусство программирования, том 3. Сортировка и поиск, 3-е издание / Д. Е. Кнут – Спб.: Диалектика, 2005.
32. Аноприенко А.Я. Тетралогика и тетракоды. // Сборник трудов факультета вычислительной техники и информатики. Вып.1. – Донецк: ДонГТУ. – 1996. С. 32-43.
33. Аноприенко А.Я. Расширенный кодо-логический базис компьютерного моделирования / В кн. «Информатика, кибернетика и вычислительная техника» (ИКВТ- 97). Сборник научных трудов ДонГТУ. Выпуск 1. Донецк, ДонГТУ, 1997, с. 59-64.
34. Аноприенко А.Я. Эволюция алгоритмического базиса вычислительного моделирования и сложность реального мира // Научные труды Донецкого национального технического университета. Выпуск 52. Серия «Проблемы моделирования и автоматизации проектирования динамических систем» (МАП-2002): Донецк: ДонНТУ, 2002. – С. 6-9.
35. Соболева А.Г. Когнитивная визуализация данных с помощью лиц Чернова // Збірка тез доповідей II Міжнародної наукової конференції студентів, аспірантів та молодих вчених «Комп'ютерний моніторинг та інформаційні технології 2006», 15-17 травня 2006 р. – Донецьк: ДонНТУ, 2006.
36. Аноприенко А.Я. Обобщенный кодо-логический базис в вычислительном моделировании и представлении знаний: эволюция идеи и перспективы развития // Научные труды Донецкого национального технического университета. Серия «Информатика, кибернетика и вычислительная техника» (ИКВТ-2005) выпуск 93: – Донецк: ДонНТУ, 2005. С. 289-316.
37. Юровицкий В.М. О компьютерной «вычислительной катастрофе», http://www.yur.ru/science/computer/Comcat.htm3
38. Курош А.Г. Лекции по общей алгебре. М.: Наука, 1973.
39. Kaucher E. Uber metriche und algebraische Eigenschaften einiger beim numersschen Technen auftretender Raume. Dr. Naturwissen/ Dissertation. – Karlsruhe: Universitat/ Karlsruhe, 1973.
40. Kaucher E. Algebraische Erweiterungen der Intervallrechnung unter Erhaltung Ordnungsund Verbandsstrukten // Grandlagen der Computer-Arithmetic / Albrecht R., Kulisch U., eds. – Wien: Springer, 1977. – P. 65-79. – (Computing Supplementum; 1)
41. Kaucher E. Interval analysis in externded interval space **IR** // Fundamentals of numerical computation (Computer-oriented numerical analysis) / Alefeld G., Grigorieff R.D/, eds. – Wien: Springer, 1980. – P. 33-49. –(Computing Supplement; 2)
42. Жуковская О.А. Исследование нестандартных интервальных арифметических операций / О.А. Жуковская // Системні дослідження та інформаційні технології. // Київ. Інститут прикладного системного аналізу НАН України та МООН України - 2005. –№2. - с.106-116.

Надійшла у першій редакції 18.07.2014, в останній - 25.10.2014.

УДК 539.3

Гибридный метод оптимизации в задаче отстройки цилиндрического резервуара от резонансных частот

О. Д. Егорова, Г. А. Шелудько
Харьковский национальный университет имени В.Н. Каразина, Украина

Решена задача отстройки от резонансных частот для цилиндрического резервуара. Данная задача рассмотрена с позиций методов оптимального проектирования. Для ее решения применен гибридный адаптивный метод оптимизации. Последовательность поисковых точек, величина шага и направление поиска генерируются с помощью специально разработанных процедур.

Ключевые слова: *цилиндрическая оболочка, резонансные частоты, оптимизация.*

Розв'язано задачу відбудови від резонансних частот для циліндричного резервуара. Ця задача розглянута з позицій методів оптимального проектування. Для її вирішення застосовано гібридний адаптивний метод оптимізації. Послідовність пошукових точок, величина кроку і напрямок пошуку генеруються за допомогою спеціально розроблених процедур.

Ключові слова: *циліндрична оболонка, резонансні частоти, оптимізація.*

The problem of detuning from resonance frequencies was solved for cylindrical tank. This problem was considered from the standpoint of optimal design methods. To solve it, the adaptive hybrid optimization method was applied. The sequence of search points as well as search step size and direction are generated by specially designed procedures.

Key words: *cylindrical shell, the resonance frequencies, optimization.*

1. Введение

Тонкостенные конструкции типа оболочек встречаются в различных областях техники, таких как машиностроение, приборостроение, авиастроение, ракетостроение, в различных видах наземного транспорта, в энергетическом строительстве. Формы объектов, которые могут быть причислены к этому классу, довольно разнообразны: корпуса различных машин, улитки турбин, корпуса ракет, судов, трубопроводы, цистерны, элементы нефте- и газопроводов, резервуары и другие емкости в промышленной аппаратуре. Большое распространение оболочек объясняется их экономичностью по сравнению с равнопрочными конструкциями, состоящими из плоских пластин. Цилиндрические оболочки – наиболее употребляемые в практике объекты, относящиеся к классу оболочек вращения. Цистерны, воздушные и газовые баллоны обычно представляют собой оболочки вращения цилиндрической формы[1]. Отметим также, что не только оценка состояния существующего оборудования, но и проектирование новых высокоэффективных машин и сооружений, обладающих необходимым уровнем надежности, требует определения прочностных характеристик их элементов. Эти данные позволяют оценить предел прочности конструкции при ударном либо сейсмическом воздействии, произвести отстройку от нежелательных резонансных частот, выявить еще на стадии проектирования наиболее опасные с точки зрения концентрации напряжений зоны.

2. Существующие методы прочностного расчета оболочек

Анализ работ, посвященных расчету оболочек, в том числе тех, которые моделируют резервуары с опасными наполнителями, позволяет выделить на данный момент как основные и наиболее эффективные методы численного интегрирования метод конечных разностей и методы конечных и граничных элементов. Методы численного интегрирования применяются, в основном, для расчета оболочек вращения. Эти методы основаны на сведении краевой задачи к ряду задач Коши, которые решаются хорошо разработанными методами численного интегрирования обычных дифференциальных уравнений.

В 1961 году С.К. Годунов предложил метод ортогонализации [2], который позволяет получить численное решение краевых задач для линейных дифференциальных уравнений, когда одновременно существуют быстро возрастающие и быстро убывающие решения. На основе этого метода разработаны эффективные алгоритмы расчета оболочек. К задачам статики несимметрично нагруженных изотропных оболочек вращения метод впервые был применен Я.М.Григоренко, А.Т. Василенко [3]. Универсальный алгоритм расчета на прочность, устойчивость и колебания осесимметрично нагруженных конструкций, составленных из набора изотропных и ортотропных оболочек вращения, соединенных между собой непосредственно или при помощи упругих шпангоутов, предложен А.В. Кармишиным, В.А. Лясковцом, В.И. Мяченковым, А.И. Фроловым [4].

3. Постановка задачи

В данной работе для решения задачи об отстройке цилиндрического резервуара от резонансных частот будем использовать гибридный метод оптимизации[5]. В работе определяются параметры цилиндрической оболочки, при которых наименьшая собственная частота будет больше заданной, чтобы избежать резонанса.

Приведем уравнения движения цилиндрической оболочки.

Определим положение произвольной точки M на срединной поверхности оболочки координатами $\alpha = x/R$ и $\beta = S/R$, где x - координата вдоль образующей, S-длина дуги в окружном направлении, R- радиус кривизны срединной поверхности (рис.1) [6].

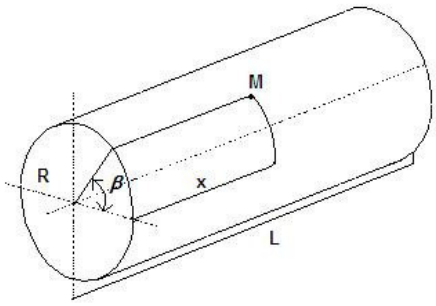


Рис. 1 Цилиндрическая оболочка

Уравнения движения цилиндрической оболочки могут быть записаны в таком виде:

$$\begin{cases} N_{11}U + N_{12}V + N_{13}W = \rho \cdot h \dfrac{\partial^2 U}{\partial t^2}; \\ N_{21}U + N_{22}V + N_{23}W = \rho \cdot h \dfrac{\partial^2 V}{\partial t^2}; \\ N_{31}U + N_{32}V + N_{33}W = \rho \cdot h \dfrac{\partial^2 W}{\partial t^2}; \end{cases} , \qquad (1)$$

где U, V, W - компоненты перемещения точки срединной поверхности в направлениях α- и β- координатных линий и по нормали; ρh - масса оболочки на единицу срединной поверхности; N_{ij} - дифференциальные операторы. Структура операторов N_{ij} для оболочек произвольной формы весьма сложна. Поэтому уравнения движения в виде (1), т.е. в перемещениях, имеет смысл привести только для простейшего случая цилиндрической оболочки постоянной толщины, для которой коэффициенты уравнений постоянны. В этом случае

$$\begin{cases} N_{11} = \dfrac{E \cdot h}{(1-\mu^2)R^2}\left(\dfrac{\partial^2}{\partial\alpha^2} + \dfrac{1-\mu}{2}\dfrac{\partial^2}{\partial\beta^2}\right); \\ N_{12} = N_{21} = \dfrac{E \cdot h}{(1-\mu^2)R^2}\dfrac{1+\mu}{2}\dfrac{\partial^2}{\partial\alpha\,\partial\beta}; \\ N_{13} = -N_{31} = \dfrac{E \cdot h}{(1-\mu^2)R^2}\ \mu\ \dfrac{\partial}{\partial\alpha}; \\ N_{22} = \dfrac{E \cdot h}{(1-\mu^2)R^2}\left\{\dfrac{1-\mu}{2}\dfrac{\partial^2}{\partial\alpha^2} + \dfrac{\partial^2}{\partial\beta^2} + a^2\left[2(1-\mu)\dfrac{\partial^2}{\partial\alpha^2} + \dfrac{\partial^2}{\partial\beta^2}\right]\right\}; \\ N_{23} = -N_{32} = \dfrac{E \cdot h}{(1-\mu^2)R^2}\left\{\dfrac{\partial}{\partial\beta} - a^2\left[(2-\mu)\dfrac{\partial^3}{\partial\alpha^2\,\partial\beta} + \dfrac{\partial^3}{\partial\beta^3}\right]\right\}; \\ N_{33} = -\dfrac{E \cdot h}{(1-\mu^2)R^2}\left[1 + a^2\left(\dfrac{\partial^2}{\partial\alpha^2} + \dfrac{\partial^2}{\partial\beta^2}\right)^2\right]; \end{cases} \qquad (2)$$

где $\alpha = x/R$; $\beta = S/R$ - безразмерные координаты точки на срединной поверхности;

$a^2 = h^2/(12R^2)$, где h – толщина оболочки;

E - модуль Юнга материала оболочки;

μ - коэффициент Пуассона материала;

ρ - плотность материала.

Особенностью уравнений движения оболочек является то, что, как это видно из формул (2), в эти уравнения входит малый, пропорциональный квадрату толщины оболочки, параметр a^2, пропорциональный квадрату толщины оболочки, на который умножаются старшие производные перемещений по координатам. Поэтому, если рассматриваются такие формы колебаний, при которых перемещения медленно меняются по координатам α и β, соответствующими моментными членами в уравнениях (1) можно пренебречь.

Аналитическое решение задачи о собственных колебаниях для замкнутой цилиндрической оболочки может быть получено при граничных условиях Навье. Согласно этим условиям, на торцах оболочки отсутствуют нормальные W и окружные V перемещения, а также продольная сила T_x в срединной поверхности и изгибающий момент M_x. Условиям Навье удовлетворяют следующие выражения компонентов перемещения:

$$\begin{aligned} U &= A\cdot\cos\left(\frac{m\cdot\pi\cdot R\cdot\alpha}{l}\right)\cdot\sin(n\cdot\beta)\cdot\cos(\omega\cdot t);\\ V &= B\cdot\sin\left(\frac{m\cdot\pi\cdot R\cdot\alpha}{l}\right)\cdot\cos(n\cdot\beta)\cdot\cos(\omega\cdot t); \quad , \\ W &= C\cdot\sin\left(\frac{m\cdot\pi\cdot R\cdot\alpha}{l}\right)\cdot\sin(n\cdot\beta)\cdot\cos(\omega\cdot t); \end{aligned} \tag{3}$$

где *n*- номер гармоники, *m*- количество узлов по окружной координате.

Подставив эти выражения в уравнения движения (1) и учитывая (2), придем к системе трех линейных алгебраических уравнений (4), (5), (6) относительно A, B, C [6]:

$$\begin{aligned} &A\cdot\left(\left(\frac{m\cdot\pi\cdot R}{l}\right)^2+\frac{1-\mu}{2}n^2\right)+B\cdot n\cdot\frac{1+\mu}{2}\cdot\frac{m\pi R}{l}-C\cdot\mu\cdot\frac{m\cdot\pi\cdot R}{l}=\\ &=\frac{A\cdot\rho\cdot\omega^2\cdot(1-\mu^2)R^2}{E}; \end{aligned} \tag{4}$$

$$\begin{aligned} &A\cdot n\cdot\frac{1+\mu}{2}\cdot\frac{m\cdot\pi\cdot R}{l}+B\left\{\frac{1-\mu}{2}\left(\frac{m\cdot\pi\cdot R}{l}\right)^2+n^2+\right.\\ &\left.+a^2\left[2(1-\mu)\cdot\left(\frac{m\cdot\pi\cdot R}{l}\right)^2+n^2\right]\right\}-Cn\left\{1+a^2\left[(2-\mu)\cdot\left(\frac{m\cdot\pi\cdot R}{l}\right)^2+\right.\right. \end{aligned} \tag{5}$$

$$\left.\left.+n^2\right]\right\}=\frac{B\cdot\rho\cdot\omega^2(1-\mu^2)R^2}{E};$$

$$-\mu\ A\frac{m\pi R}{l}-B\cdot n\left(1+a^2\left[(2-\mu)\left(\frac{m\cdot\pi\cdot R}{l}\right)^2+n^2\right]\right)+$$
$$+C\left\{1+a^2\left(\left(\frac{m\cdot\pi\cdot R}{l}\right)^2+n^2\right)^2\right\}=\frac{C\cdot\rho\cdot\omega^2(1-\mu^2)R^2}{E}; \quad (6)$$

Равенство нулю определителя этой системы приводит к кубическому уравнению относительно ω^2. Три корня этого уравнения соответствуют трем различным формам колебаний с одинаковыми числами узловых окружностей и образующих, но с различными соотношениями между A,B,C. Корни этого определителя связаны с собственными частотами колебаний резервуара таким соотношением:

$$\overline{\omega}=\frac{\rho\cdot\omega^2\cdot(1-\mu^2)R^2}{E}\ . \quad (7)$$

В дальнейшем будем искать частоты, обезразмеренные по формуле (7).

Пусть задано критическое значение обезразмеренной частоты колебаний ω^*. Требуется подобрать такие параметры цилиндрической оболочки, чтобы наименьшая ее частота колебаний была бы больше критического значения. При этом объем оболочки должен быть больше заданного значения V_0. С точки зрения методов оптимального проектирования нами сформулированы два ограничения на проектные параметры. Отметим, что проектными параметрами в данном случае будут R, l, h. Объем цилиндрической оболочки выражается через эти параметры следующим образом:

$$V=\pi R^2 l\ . \quad (8)$$

Таким образом, второе из сформулированных ограничений имеет следующий вид;

$$G_2=\pi R^2 l-V_0\ . \quad (9)$$

Первое ограничение в общем виде имеет следующую форму:

$$G_1=\min\{\overline{\omega}\}-\omega^*\ . \quad (10)$$

Отметим, что первое из ограничений G_1 задается алгоритмически, потому что оно не может быть представлено в виде аналитического выражения, так как аналитически не удается найти собственные значения матрицы, полученной из системы трех линейных алгебраических уравнений относительно A,B,C

4. Метод решения задачи

При оптимальном проектировании сложных многопараметрических объектов рассматриваемого типа удобно воспользоваться автоматическим гибридным поисковым методом оптимизации [5], предназначенным для отыскания локального оптимального вектора X^* задачи на условный экстремум

$$X^* = \arg \underset{X \in G}{extr}\, F(X) \tag{11}$$

в допустимой области

$$G = \left\{X : G_i(X) \geq 0, i = \overline{1,m}\right\} \neq \varnothing \tag{12}$$

Суть предлагаемого метода заключается, в общих чертах, в следующем. Имеется ряд методов-гибридиентов, которые составляют гибридную коалицию $\{M_i\}$. Задается критерий $Q(\boldsymbol{\sigma})$, выясняющий в процессе решения, какой из гибридиентов в данной ситуации $\boldsymbol{\sigma}$ наиболее эффективно может использоваться для достижения поставленной цели. Вводится функция управления $u = u(Q(\boldsymbol{\sigma}))$, устанавливающая адаптивную стратегию ввода в действие конкретного гибридиента $M_k \in \{M_i\}$, $i = 1,..., k,..., s$ (или группы гибридиентов).

Совместные действия гибридиентов обеспечивают более эффективное достижение цели, нежели каждый из гибридиентов коалиции в отдельности. Это достигается путем введения специального адаптивного управления, которое осуществляет получение векторов минимизирующей последовательности $\left\{X_k^r\right\}$, направлений поиска $\mathrm{Dir}\left\{X_k^r\right\}$ и поисковых адаптирующихся шагов h_k^r, в соответствии с изменяющейся ситуацией $\boldsymbol{\sigma}$. В общем случае адаптивное управление u можно представить в виде

$$\begin{Bmatrix} X_k^r \\ Dir X_k^r \\ h_k^r \end{Bmatrix} = \sum_{i=1}^{s} u_i\left(Q(\sigma_k)\right) \begin{Bmatrix} X_k^{M_i} \\ Dir\, X_k^{M_i} \\ h_{ki} \end{Bmatrix}, \qquad \sum_{i=1}^{s} u_i\left(Q(\sigma_k)\right) = 1, \tag{13}$$

где $u_i(Q(\sigma_k))$ – управляющие неотрицательные функции, заданные на множестве ситуаций $\{\sigma_k\}$;

$X_k^{M_i}$, $\mathrm{Dir}\, X_k^{M_i}$ и h_{ki} – точка, направление, исходящее из этой точки, и адаптирующийся шаг поиска, генерированные методом M_i, соответственно;

k – номер итерации.

В качестве гибридиентов M_i для данного варианта гибридного метода оптимизации выбраны следующие модификации методов [5,7]: адаптивный пошаговый спуск, схема Абрамова, овражная модификация, метод параллельных касательных, секущее движение вдоль границы области G. В [5] показано, что гибридный метод может решать широкий класс задач более эффективно, чем каждый из упомянутых гибридиентов.

Спецификой рассматриваемой задачи является отсутствие функции цели. Поэтому оптимальное проектирование в данном случае сводится к процедуре «входа в область»: Эта процедура строит точку $X = \{R, l, h\}$, принадлежащую G, то есть решает систему функциональных неравенств $G_j(X) \geq 0$. Она использует основной алгоритм минимизации [7], но применительно к функции цели вида:

$$\Phi(X) = -\sum_{j \in m} G_j(X), \quad m' = (j \,|\, G_j < 0), \tag{14}$$

то есть представляющей собой сумму нарушенных ограничений. Сходимость метода согласно [7] гарантируется в предположении, что Φ монотонно возрастает вдоль любого направления, которое содержится в выпуклой оболочке множества ограничений и начальной точки. Эта функция является неотрицательной и принимает нулевые значения в допустимой области (12).

Таким образом, задача отстройки цилиндрического резервуара от резонансных частот сводится к решению системы функциональных неравенств

$$\begin{cases} G_1 \geq 0; \\ G_2 \geq 0; \end{cases} \tag{15}$$

Численные результаты

Численно решена задача об отстройке цилиндрического резервуара от резонансных частот.

Для задания ограничений (15) были выбраны следующие значения фигурирующих в них величин:

$$\omega^* = 0.4; \quad V_0 = 20; \tag{16}$$

В качестве начальной точки выбиралась $X_0 = \{1,\ 5,\ 0.01\}$, при этом $V = 15.71\,\text{м}^3$.

Наименьшая частота колебаний оказалась равной

$$\min \overline{\omega} = 0.0347\ . \tag{17}$$

Как видим, оба ограничения оказались нарушенными:

$$G_1 = -0.37; \quad G_2 = -4.29. \tag{18}$$

Если подобрать такой объем, чтобы он совпадал с заданным значением V_0, то первое ограничение все равно остается нарушенным. Методы оптимизации позволяют решить задачу выбора параметров так, чтобы все ограгичения на проектные параметры оказались выполненными. В результате применения гибридного метода оптимизации был получен вектор

$$X_{opt} = \{3, \quad 1, \quad 0.01\}. \tag{19}$$

В этой точке значения ограничений G_1 и G_2 положительны, а именно:

$$G_1 = 0.59; \quad G_2 = 8.273. \tag{20}$$

Таким образом, проведена отстройка от опасной резонансной частоты.

7. Выводы и перспективы дальнейших исследований

Предложенный подход был реализован с помощью компьютерного анализа. Были получены оптимальные параметры цилиндрической оболочки, которые позволили произвести отстройку от резонансной частоты при ограничении на минимальный объем цилиндрического резервуара. В данной задаче третий параметр (толщина оболочки h) оказался неактивным. Это объясняется отсутствием прочностных ограничений на рассматриваемый резервуар.

Разработанный метод предлагается использовать для решения широкого класса задач оптимального проектирования тонкостенных оболочечных конструкций при наличии прочностных и геометрических ограничений.

ЛИТЕРАТУРА

1. Филин А.П. Элементы теории оболочек. – Л.: Стройиздат. Л. отделение, 1975.– 256 с.
2. Годунов С.К. О численном решении краевых задач для систем обыкновенных дифференциальных уравнений / С.К. Годунов // Успехи мат. наук.– 1961.– Вип.3.–С.171–174.
3. Григоренко Я.М. О расчете и выборе рациональных параметров оболочечных конструкций из композиционных материалов / Я.М. Григоренко, А.Т. Василенко // Механика композит. материалов. –1981. –№ 1. –С.64–69.
4. Статика и динамика тонкостенных оболочечных конструкций / А.В. Кармишин, В.А. Лясковец, В.И. Мяченков, А.И. Фролов. –М.: Машиностроение, 1975. –576 с.
5. Шелудько Г.А., Стрельникова Е.А. Гибридный метод оптимизации. /Препр. АН УССР. Ин-т пробл. машиностр.; № 164:–Харьков, 1980.– 64 с.
6. Прочность. Устойчивость. Колебания: Справочник, Т.3. / Под ред. И.А. Биргера, Я.Г.Пановко. – М.:Машиностроение, 1968. – 567 с.
7. Динамика конструкций при воздействии кратковременных нагрузок./ С.С. Кохманюк, А.С. Дмитриев, Г.А. Шелудько, А.Н.Шупиков, В.Г.Титарев, А.Н.Ляхов.– Киев: Наук. думка, 1984. – 198 с.

Надійшла – 22.05.2014

УДК 519.6

Новый метод вычисления базисных функций атомарного обобщенного ряда Тейлора

О. А. Иванова

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

В статье приведен новый метод вычисления базисных функций атомарного обобщенного ряда Тейлора, который основан на разложении базисных функций в ряд Фурье. Были выведены формулы преобразований Фурье базисных функций и записаны разложения с помощью рядов Фурье базисных функций атомарного обобщенного ряда Тейлора порядка от нуля до трех.

Ключевые слова: *атомарная функция, атомарный обобщенный ряд Тейлора, базисная функция, преобразование Фурье, ряд Фурье.*

У статті приведено новий метод обчислення базисних функцій атомарного узагальненого ряду Тейлора, який заснований на розкладанні базисних функцій в ряд Фур'є. Було виведено формули перетворень Фур'є базисних функцій і записано розкладання за допомогою рядів Фур'є базисних функцій атомарного узагальненого ряду Тейлора порядку від нуля до трьох.

Ключові слова: *атомарна функція, атомарний узагальнений ряд Тейлора, базисна функція, перетворення Фур'є, ряд Фур'є.*

The paper presents a new method of basis functions computation for atomic generalized Taylor series. The method is built on Fourier series expansion of basis functions in question. The Fourier transform formulas for the basis functions are derived and Fourier series expansion of the basis functions of atomic generalized Taylor series can be explicitly written for series of order from zero to third.

Key words*: atomic function, atomic generalized Taylor series, the basis function, Fourier transform, Fourier series.*

1. Общая постановка задачи и её актуальность

Атомарный обобщенный ряд Тейлора удобно применять для решения задач с дифференциальными и с интегральными уравнениями в таких областях науки, как в механике, в частности в теории упругости, статической механике, механике сплошных сред, электродинамики, теории антенн и других [1-4].

Атомарный обобщенный ряд Тейлора [5-6] разложения функции f имеет вид:

$$f(x)=\sum_{n=0}^{\infty}\sum_{k\in N_n} f^{(n)}(x_{n,k})baf_{n,k}(x), \qquad (1.1)$$

где

$N_0=\{-1,0,1\}$;

$N_n=\{-2^{n-1},-2^{n-1}+1,...,2^{n-1}-1,2^{n-1}\},\ n\neq 0$;

$x_{n,k}=\dfrac{k}{2^{n-1}},\ n\neq 0,\ k\in N_n,\ x_{0,k}=k,\ k\in N_0$;

$baf_{n,k}(x)\in H_1$ - БАФ атомарного обобщенного ряда Тейлора.

При практическом применении атомарного обобщенного ряда Тейлора, пользоваться ранее известными формулами [3, 8] по нахождению базисных функций атомарного обобщенного ряда Тейлора несколько неудобно, поскольку они выражаются с помощью линейных комбинаций различных сдвигов функции $up(x)$. Привести подобные слагаемые и упростить вычисления весьма затруднительно. Таким образом, возник вопрос о необходимости нахождения таких формул вычисления базисных функций (БАФ) атомарного обобщенного ряда Тейлора, которые упрощали бы вычисления.

В данной статье предлагается новый метод нахождения базисных функций обобщенного атомарного ряда Тейлора с помощью разложения их в ряд Фурье.

2. Преобразование Фурье базисных функций атомарного обобщенного ряда Тейлора.

Для того чтобы записать разложение базисных функций в ряд Фурье, необходимо сначала найти преобразование Фурье для каждой базисной функции.

Пусть прямое преобразование Фурье функции $g(x)$ имеет вид:

$$\int_{-\infty}^{\infty} e^{itx} g(x)dx = G(t). \tag{2.1}$$

Выпишем некоторые формулы, необходимые нам для дальнейшего вычисления.

1. Преобразование Фурье интегрирования имеет вид:

$$\int_{-\infty}^{\infty} e^{itx} \int_{-\infty}^{x} g(\xi)d\xi dx = -\frac{1}{it} G(t). \tag{2.2}$$

Доказательство:

Проделаем интегрирование по частям:

$$\int_{-\infty}^{\infty} e^{itx} \int_{-\infty}^{x} g(\xi)d\xi dx = \left| \begin{matrix} e^{itx}dx = dv \\ \int_{-\infty}^{x} g(\xi)d\xi = u \end{matrix} \right| = \frac{e^{itx}}{it} \int_{-\infty}^{x} g(\xi)d\xi \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{e^{itx}}{it} g(x)dx = -\frac{1}{it} G(t).$$

2. Преобразование Фурье сжатия имеет вид:

$$\int_{-\infty}^{\infty} e^{itx} g(ax)dx = \frac{1}{a} G\left(\frac{t}{a}\right). \tag{2.3}$$

Доказательство:

$$\int_{-\infty}^{\infty} e^{itx} g(ax)dx = \left| \begin{matrix} ax = u \\ dx = \frac{du}{a} \end{matrix} \right| = \frac{1}{a} \int_{-\infty}^{\infty} e^{it\frac{u}{a}} g(u)du = \frac{1}{a} G\left(\frac{t}{a}\right).$$

3. Преобразование Фурье сдвижки имеет вид:

$$\int_{-\infty}^{\infty} e^{itx} g(x-h)dx = e^{ith} G(t)\,. \tag{2.4}$$

Доказательство:

$$\int_{-\infty}^{\infty} e^{itx} g(x-h)dx = \left|x-h=u\right| = \int_{-\infty}^{\infty} e^{it(u+h)} g(u)du = e^{ith} G(t).$$

Преобразование Фурье базисных функций атомарного обобщенного ряда Тейлора вычисляются с помощью рекуррентной формулы.

Поскольку БАФ следующего порядка вычисляется путем интегрирования сжатой функции БАФ предыдущего порядка в нужной точке и вычитаемой сжатой и сдвинутой функции up(x) с нужным коэффициентом:

$$baf_{n,\frac{y}{2}}(x) = \frac{1}{2^{n-1}} \int_{-\infty}^{x} \left[baf_{n-1,y}(2\xi) - \alpha up(2\xi+1) \right] d\xi\,. \tag{2.5}$$

Введем обозначения:

$FTbaf_{n-1,\frac{y}{2}}(t) = \int_{-\infty}^{\infty} e^{itx} baf_{n-1,\frac{y}{2}}(x)dx$ - преобразование Фурье БАФ $baf_{n-1,\frac{y}{2}}(x)$.

$FTup(t) = \int_{-\infty}^{\infty} e^{itx} up(x)dx = \prod_{k=1}^{\infty} \frac{\sin\frac{t}{2^k}}{\frac{t}{2^k}}$ – преобразование Фурье функции $up(x)$.

Преобразование Фурье функции $baf_{n,\frac{y}{2}}(x)$ будет иметь вид:

$$FTbaf_{n,\frac{y}{2}}(t) = \frac{\frac{1}{2^{n-1}} \frac{1}{2} \left[FTbaf_{n-1,y}\left(\frac{t}{2}\right) - \alpha e^{\frac{-it}{2}} FTup\left(\frac{t}{2}\right) \right]}{-it}. \tag{2.6}$$

Чтобы найти коэффициент α, вычислим $\lim_{t\to 0} FTbaf_{n,\frac{y}{2}}(t)$:

$$\lim_{t\to 0} FTbaf_{n,\frac{y}{2}}(t) = \lim_{t\to 0} \frac{\frac{1}{2^{n-1}} \frac{1}{2} \left[FTbaf_{n-1,y}\left(\frac{t}{2}\right) - \alpha e^{\frac{-it}{2}} FTup\left(\frac{t}{2}\right) \right]}{-it}.$$

Очевидно,

$$\alpha = \lim_{t\to 0} \frac{FTbaf_{n-1,y}\left(\frac{t}{2}\right)}{e^{\frac{-it}{2}} FTup\left(\frac{t}{2}\right)}. \tag{2.7}$$

Обозначим

$$H(t)=\frac{1}{2^{n-1}}\frac{1}{2}\left[FTbaf_{n-1,y}\left(\frac{t}{2}\right)-\lim_{t\to 0}\frac{FTbaf_{n-1,y}\left(\frac{t}{2}\right)}{e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)}e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)\right] \quad (2.8)$$

и проделаем это же исследование для БАФ следующего порядка:

$$baf_{n+1,\frac{y}{4}}(x)=\frac{1}{2^n}\int_{-\infty}^{x}\left[baf_{n,\frac{y}{2}}(2\xi)-\beta up(2\xi+1)\right]d\xi. \quad (2.9)$$

$$FTbaf_{n+1,\frac{y}{4}}(t)=\frac{\frac{1}{2^n}\frac{1}{2}\left[FTbaf_{n,\frac{y}{2}}\left(\frac{t}{2}\right)-\beta e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)\right]}{-it}=$$

$$=\frac{\frac{1}{2^n}\frac{1}{2}\left[H\left(\frac{t}{2}\right)-\beta e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)\right]}{-it}.$$

Найдем коэффициент β, вычислив $\lim_{t\to 0}FTbaf_{n+1,\frac{y}{4}}(t)$:

$$\lim_{t\to 0}FTbaf_{n+1,\frac{y}{4}}(t)=\lim_{t\to 0}\frac{\frac{1}{2^n}\frac{1}{2}\left[H\left(\frac{t}{2}\right)-\beta e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)\right]}{-it}.$$

Числитель приравниваем к нулю:

$$\lim_{t\to 0}\left[\frac{H\left(\frac{t}{2}\right)}{-i\frac{t}{2}}-\beta e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)\right]=0.$$

Применив правило Бернулли-Лопиталя к первому слагаемому, получим

$$\beta=\lim_{t\to 0}\frac{\frac{H'\left(\frac{t}{2}\right)}{-\frac{i}{2}}}{e^{\frac{-it}{2}}FTup\left(\frac{t}{2}\right)}. \quad (2.10)$$

Эти умозаключения верны и для БАФ высшего порядка.

3. Ряды Фурье базисных функций атомарного обобщенного ряда Тейлора.

Запишем общий вид ряда Фурье для функции $f(x) \in [-L; L]$:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{\pi n x}{L} + b_n \sin\frac{\pi n x}{L}\right), \tag{3.1}$$

где

$$a_0 = \frac{1}{L}\int_{-L}^{L} f(x)dx,$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{\pi n x}{L}dx,$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{\pi n x}{L}dx.$$

Выпишем разложения базисных функций атомарного обобщенного ряды Тейлора в ряды Фурье.

$$baf_{0,0}(x) = \frac{1}{2} + \sum_{n=1}^{\infty} FTbaf_{0,0}(\pi n)\cos\pi n x. \tag{3.2}$$

Функции нулевого порядка в точках 1 и -1 находят с помощью переноса функции $baf_{0,0}(x)$ соответственно вправо или влево на одну единицу:

$$baf_{0,1}(x) = baf_{0,0}(x-1),\ baf_{0,-1}(x) = baf_{0,0}(x+1).$$

Аналогичными переносами можно найти функции $baf_{1,1}(x)$ и $baf_{1,-1}(x)$, $baf_{2,1}(x)$ и $baf_{2,-1}(x)$, $baf_{3,1}(x)$ и $baf_{3,-1}(x)$ соответствующих им функций $baf_{1,0}(x)$, $baf_{2,0}(x)$ и $baf_{3,0}(x)$:

$$baf_{1,0}(x) = \frac{1}{i}\sum_{n=1}^{\infty} FTbaf_{1,0}(\pi n)\sin\pi n x, \tag{3.3}$$

$$baf_{2,0}(x) = \frac{17}{2304} + \sum_{n=1}^{\infty} FTbaf_{2,0}(\pi n)\cos\pi n x, \tag{3.4}$$

$$baf_{3,0}(x) = \frac{1}{i}\sum_{n=1}^{\infty} FTbaf_{3,0}(\pi n)\sin\pi n x. \tag{3.5}$$

$$baf_{2,-\frac{1}{2}}(x) = -\frac{1}{128} + \sum_{n=1}^{\infty}\left(\left(FTbaf_{2,-\frac{1}{2}}(2\pi n) + FTbaf_{2,-\frac{1}{2}}(-2\pi n)\right)\cos 2\pi n x + \right.$$
$$\left. + \left(FTbaf_{2,-\frac{1}{2}}(2\pi n) + FTbaf_{2,-\frac{1}{2}}(-2\pi n)\right)\frac{\sin 2\pi n x}{i}\right), \tag{3.6}$$

$$baf_{3,-\frac{1}{2}}(x)=\sum_{n=1}^{\infty}\left(\left(FTbaf_{3,-\frac{1}{2}}(2\pi n)+FTbaf_{3,-\frac{1}{2}}(-2\pi n)\right)\cos 2\pi nx+\right.$$
$$\left.+\left(FTbaf_{3,-\frac{1}{2}}(2\pi n)+FTbaf_{3,-\frac{1}{2}}(-2\pi n)\right)\frac{\sin 2\pi nx}{i}\right), \tag{3.7}$$

$$baf_{3,-\frac{1}{4}}(x)=\frac{1}{4096}+\sum_{n=1}^{\infty}\left(\left(FTbaf_{3,-\frac{1}{4}}(2\pi n)+FTbaf_{3,-\frac{1}{4}}(-2\pi n)\right)\cos 2\pi nx+\right.$$
$$\left.+\left(FTbaf_{3,-\frac{1}{4}}(2\pi n)+FTbaf_{3,-\frac{1}{4}}(-2\pi n)\right)\frac{\sin 2\pi nx}{i}\right), \tag{3.8}$$

$$baf_{3,-\frac{3}{4}}(x)=-baf_{3,-\frac{1}{4}}(-x-1). \tag{3.9}$$

Базисные функции для положительных точек легко находятся из соответствующих им симметричных отрицательных точек. Все формулы преобразование Фурье для каждой базисной функции легко находятся по алгоритму, предложенному в разделе 2 данной статьи.

4. Выводы по результатам

Предложен новый метод вычисления базисных функций атомарного обобщенного ряда Тейлора, основанный на разложении непосредственно самих базисных функций в ряд Фурье. Данный метод имеет ряд преимуществ, благодаря компактности формул. Поэтому при практическом применении атомарного обобщенного ряда Тейлора, пользоваться полученными формулами по нахождению базисных функций атомарного обобщенного ряда Тейлора весьма удобно, поскольку возникает возможность привести подобные слагаемые и упростить дальнейшие вычисления.

ЛИТЕРАТУРА

1. Рвачев В.А., Рвачева Т.В., Томилова Е.П. Применение атомарных обобщенных рядов Тейлора к решению интегральных уравнений электродинамики и теории антенн // Радиоэлектр. и компьют. сист. – 2013. – № 1 (60) – С. 7-14.
2. Ivanova O.A. Application generalized Taylor series for solving the Cauchy problem for differential equations of the second order // Всеукраїнськ. Наукова конф. «Сучасні проблеми математичного моделювання та обчислювальних методів»: 22-23 лютого 2013 р.: тез. доп. – Рівне, 2013. – С. 206.
3. Иванова О.А. Новый метод нахождения ядра интегрального уравнения в обратной задаче об определении характеристик вязкоупругих материалов // Вопросы проектир. и произв. конструкц. летат. аппаратов. – 2013. – № 4 (76). – С. 50-55.

4. Рвачев В.А., Рвачева Т.В. О построении мультимодальных многопараметрических экспоненциальных семейств вероятностных законов// Радиоэлектр. и компьют. сист. – 2011. – № 4 (52) – С. 72-76.
5. Рвачев В.А. Обобщенные ряды Тейлора для бесконечно дифференцируемых функций // Мат. методы анализа динамических систем. – 1982. –№ 6. – С. 99-102.
6. Рвачев В.А. Финитные решения функционально-дифференциальных уравнений и их применение // Успехи мат. наук. – 1990. – № 1(271). – С. 77-103.
7. Рвачев В.Л., Рвачев В.А. Об одной финитной функции // ДАН УССР. – 1971. – сер. А. – С. 705-707.
8. Рвачев В.А., Рвачева Т.В. Об Эрмитовой интерполяции с помощью атомарных функций // Радіоелектронні і комп'ютерні системи. – 2010. – № 4(45). – С. 100-104.
9. Рвачев В.Л., Рвачев В.А. Теория приближений и атомарные функции. – М.: Знание, 1978. – 62 с.
10. Рвачева Т.В. О скорости приближения бесконечно дифференцируемых функций частичными суммами обобщенного ряда Тейлора // Вісник ХНУ, сер. «Математика, прикладна математика і механіка». – 2010. – 931. – С. 93–98.
11. Смирнов Д.В. Цифровая обработка сигналов атомарными функциями в радиофизических приложениях: дис. канд. физ.-мат. наук: 01.04.03. – М. – 2005. – 165 с.
12. Колодяжный В.М., Рвачев В.А. Атомарные функции. Обобщения на случай многих переменных и перспективные направления практических приложений // Кибернетика и системный анализ. – 2007. – 43, № 6. – С. 155–177.

Надійшла у першій редакції – 05.09.2014, в останній – 11.09.2014.

УДК 004.056.55

Аналіз колізійних властивостей режиму вироблення імітовставок із вибірковим гамуванням

Д. В. Іваненко[1], О. О. Кузнецов[2], Є. П. Колованова[2]
1 Харківський національний університет радіоелектроніки, Україна
2 Харківський національний університет імені В.Н. Каразіна, Україна

Досліджується режим вироблення імітовставки із вибірковим гамуванням, який призначено для забезпечення цілісності та конфіденційності повідомлень. Розглядаються основні криптографічні перетворення, які застосовуються при реалізації цього режиму, даються теоретичні оцінки ймовірності збігів (колізій) формованих імітовставок. Обгрунтовуються пропозиції щодо вдосконалення дослідженого режиму при застосуванні в інформаційно-комунікаційних системах

Ключові слова: *колізія, гешування, імітовставка, криптографічний захист, блокове симетричне шифрування, цикл, підстановка.*

Исследуется режим выработки имитовставки с выборочным гаммированием, который предназначен для обеспечения целостности и конфиденциальности сообщений. Рассматриваются основные криптографические преобразования, которые используются при реализации этого режима, приводятся теоретические оценки вероятности коллизий сформированных имитовставок. Обосновываются предложения по усовершенствованию исследуемого режима при использовании в информационно-коммуникационных системах

Ключевые слова: *коллизия, хеширование, имитовставка, криптографическая защита, блоковое симметричное шифрование, цикл, подстановка.*

In the paper, we investigate formation of the Galois Message Authentication Code with selective Counter, which is designed to ensure the integrity and confidentiality of communications. The paper describes the basic cryptographic transformations used to implement this mode, and gives the theoretical estimates of collisions probability. We propose and substantiate improvements of this mode when applied to information and communication systems.

Key words:*. collision, hashing, authentication code, cryptographic protection, block symmetric encryption, cycle, permutation.*

1. Вступ

Однією з важливих складових забезпечення інформаційної безпеки є криптографічний захист інформації, тому дослідження сучасних криптоперетворень та обгрунтування перспективних напрямків зі створення надійних національних технологій захисту інформації є важливим та складним науковим завданням.

З метою побудови сучасних механізмів криптографічного захисту інформації широко застосовується блокове симетричне шифрування, яке полягає у перетворенні інформації з використанням ключових даних з метою приховування (відновлення) змісту інформаційного повідомлення, підтвердження його справжності, цілісності, авторства, тощо. При цьому рівень захищеності інформації залежить не лише від властивостей блокового симетричного шифру, але і від режиму шифрування, під яким зазвичай розуміється такий метод його використання, який дозволяє реалізувати

перетворення послідовності блоків відкритих даних в послідовність блоків зашифрованих даних із отриманням певних, наперед визначених криптографічних властивостей [10].

Надійним механізмом забезпечення цілісності та конфіденційності інформації в сучасних інформаційно-комунікаційних системах є режим формування імітовставки із вибірковим гамуванням (Galois/Counter Mode and GMAC), специфікацію якого наведено у міжнародному стандарті NIST SP 800-38D [4]. Цей режим призначено для реалізації швидкого криптоперетворення при забезпеченні послуг безпеки інформації із використанням різних криптографічних примітивів, зокрема поліноміального ґешування, гамування, тощо.

Метою даної роботи є аналіз колізійних властивостей формованих імітовставок нового режиму шифрування Galois/Counter Mode and GMAC (GCM & GMAC) та обґрунтування умов його застосування в сучасних інформаційно-комунікаційних системах.

2. Аналіз криптоперетворень GCM & GMAC

Новий режим шифрування Galois/Counter Mode and GMAC призначено для забезпечення послуг конфіденційності та цілісності даних, перш за все, при реалізації інформаційно-комунікаційних протоколів, зокрема в межах протоколів безпеки IPSec.

Структурна схема режиму вироблення імітовставки із вибірковим гамуванням GCM-AE_K (IV, P, A) наведено на рис. 1, де позначення 0^s визначає рядок довжини s, який складається з бітів ‘0’; len(X) - бітова довжина рядка X; $[x]_s$ повертає бінарне представлення x як рядка бітової довжини s; MSBs(X) повертає s найбільш значущих бітів X; CIPH - затверджений 128-бітний блоковий симетричний шифр. Для забезпечення конфіденційності відкритого тексту P застосовується функція $GCTR_K$ - деяка варіація режиму гамування [1-3], де перший блок лічильника для шифрування відкритого тексту генерується шляхом збільшення (inc_{32}) блоку J_0, сформованого з вектору ініціалізації IV. Для забезпечення цілісності застосовується інший механізм, який засновано на функції ґешування $GHASH_H$. Функцію ґешування використовують для стискання зашифрованих доданих автентифікованих даних A (Additional Authenticated Data – AAD) та шифротексту C в єдиний блок, який далі проходить шифрування для створення коду справжності T (імітовставки).

Зворотне перетворення полягає в перевірці справжності шифртексту C із доданими даними A та реалізується функцією GCM-AD_k (IV, C, A, T), структурну схему якої наведеною на рис. 2. При підтвердженні справжності (знов обчислений код T’ дорівнює отриманому T) виконується розшифрування шифртексту C та формується відкритий текст P.

Таким чином, як видно з рис. 1 та 2, основними перетвореннями нового режиму Galois/Counter Mode and GMAC є ґешування даних із використанням функції $GHASH_H$ та шифрування/розшифрування функцією $GCTR_K$. Розглянемо їх більш детально.

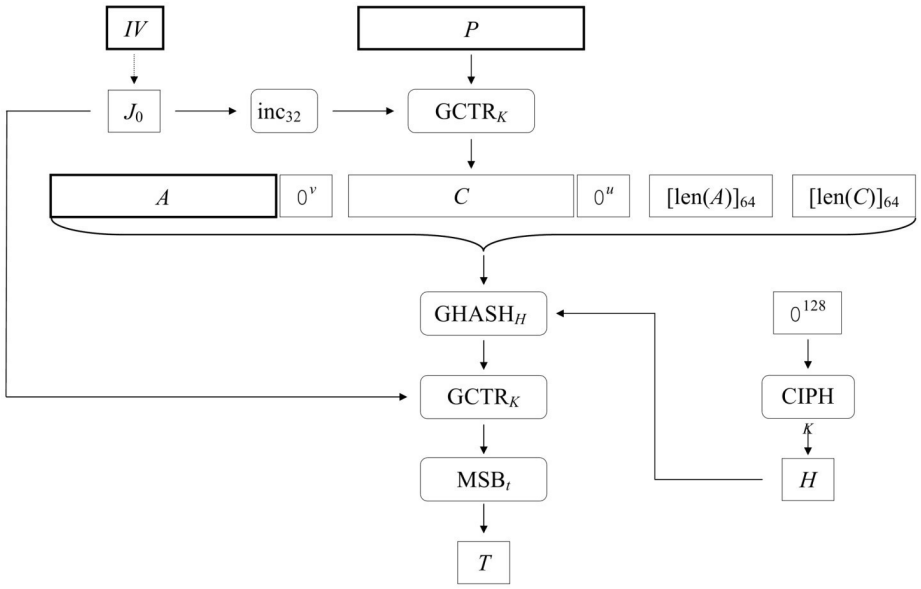


Рис. 1. $\mathrm{GCM-AE_K}\left(\mathrm{IV},\ \mathrm{P},\ \mathrm{A}\right)\ =\ \left(\mathrm{C},\ \mathrm{T}\right).$

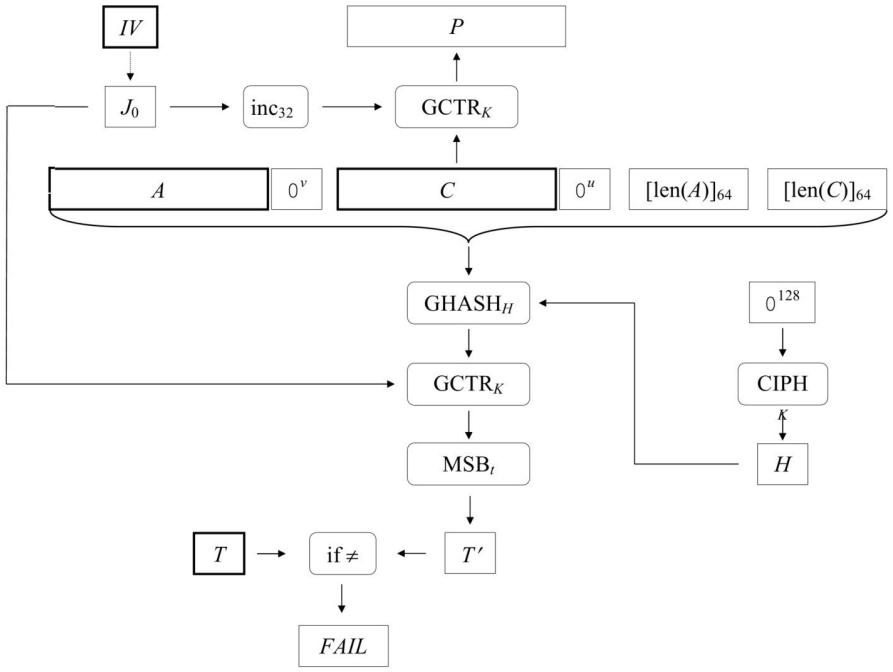


Рис. 2. $\mathrm{GCM-AD_k}\left(\mathrm{IV},\ \mathrm{C},\ \mathrm{A},\ \mathrm{T}\right)\ =\ \mathrm{P}$ *або* $\mathrm{FAIL}.$

На рис. 3 зображено структурну схему функції $\mathrm{GCTR_K}$ для реалізації шифрування/розшифрування, де ICB - початковий блок лічильника; $\mathrm{CB_i}$ – і-ий блок лічильника; inc – функція лічильника. Шифрування відбувається за

допомогою затвердженого блокового симетричного шифру AES зі 128-бітним розміром блоку з використанням ключа шифрування K.

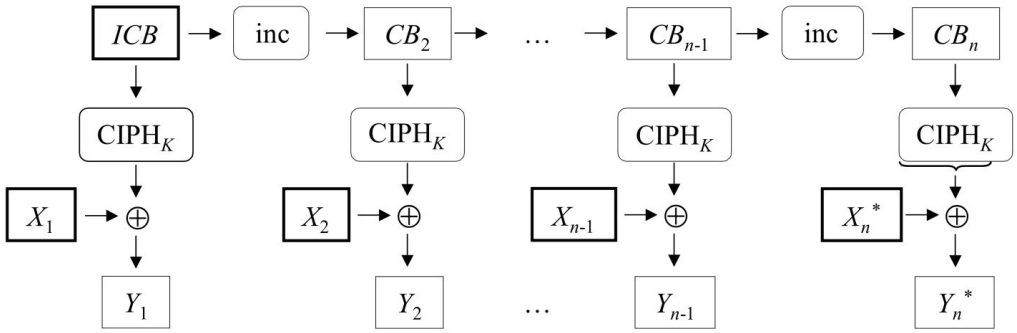


Рис. 3. $\mathrm{GCTR_K}(ICB,\ X_1 \| \mathrm{X}_2 \| ... \| \mathrm{X}_n^*) = Y_1 \| \mathrm{Y}_2 \| ... \| \mathrm{Y}_n^*$.

Для забезпечення цілісності інформації використовується функція ґешування GHASH (див. рис. 1, 2). Цю функцію побудовано на основі поліноміальної схеми та реалізовано за допомогою множення на фіксований параметр – субключ з операціями в двійковому полі Галуа. На вхід функції подається деяка унікальна послідовність блоків довжиною m. Функція ґешування GHASH розраховує ґеш-значенння з використанням схеми Горнера, структуру якої зображено на рис. 4.

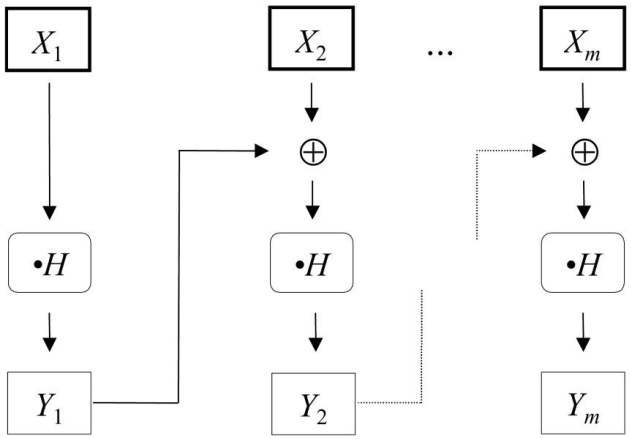


Рис. 4. $\mathrm{GHASH_H}(\mathrm{X}_1 \| \mathrm{X}_2 \| \ ... \ \| \mathrm{X}_\mathrm{m}) = \mathrm{Y}_\mathrm{m}$

Таким чином, проведений аналіз показав, що режим GCM & GMAC, який визначений у NIST Special Publication 800-38D: Galois/Counter Mode (GCM) and GMAC, містить наступні криптоперетворення:

- шифрування/розшифрування відкритого тексту P функцією $\mathrm{GCTR_K}$, яка по суті є деякою варіацією режиму гамування CTR [3, 4];

- обчислення ґеш-значення із використанням поліноміальної функції $GHASH_H$, яку використовують для стискання шифротексту та автентифікованих даних A в єдиний блок;

- шифрування/розшифрування отриманого ґеш-значення, при цьому знов застосовується функція $GCTR_K$.

Отримане зашифроване ґеш-значення і є тим кодом справжності T, який призначено для забезпечення цілісності та автентичності інформації. За вітчизняною термінологією код T є імітовставкою, формування якої призначено для забезпечення захищеності від підміни та/або перекручування даних, здатності протистояти нав'язуванню помилкових повідомлень чи підміні повідомлення з метою зміни його сенсу.

Основним ймовірнісним показником ефективності режиму вироблення імітовставки (коду справжності, ґеш-значення) є ймовірність збігу (колізії) формованих кодів, тобто ймовірність такої події, коли для різних вхідних даних формовані імітовставки співпадають. Тобто ймовірність колізій визначає властивості формованих імітовставок щодо їх співпадіння на повній множині ключів та вхідних даних і є вихідним параметром щодо оцінки імітостійкості [10].

Проведемо дослідження колізійних властивостей розглянутої схеми вироблення імітовставок. При проведенні досліджень зосередимо увагу на вивченні властивостей ґеш-кодів, що сформовано функцією поліноміального ґешування $GHASH_H$, та вихідних кодів справжності (імітовставок), що сформовано за результатом роботи режиму GCM & GMAC загалом.

3. Дослідження колізійних властивостей поліноміального ґешування $GHASH_H$

Проведений аналіз показав, що при формуванні імітовставок у режимі GCM & GMAC використана функція ґешування за поліноміальною схемою Горнера (функція $GHASH_H$). За визначенням вона належить до класу універсальних ґеш-функцій [5].

Ідеєю універсального ґешування є визначення набору ґеш-функцій таким чином, що випадковий вибір функції забезпечить низьку ймовірність того, що будь-які два різних вхідних повідомлення X_a та X_b дадуть колізію, коли їх ґеш-значення розраховані з використанням функції із визначеної множини. Ймовірність виникнення такої колізії можна підрахувати наступним чином [5, 6]:

$$P_k = \delta_y(X_a, X_b) / |H|,$$

де P_k - ймовірність виникнення колізії, $\delta_y(X_a, X_b)$ - кількість співпадінь ґеш-значень, $|H|$ - потужність множини ґеш-функцій (або потужність множини ключів ґешування, бо кожний ключ визначає окрему функцію із визначеної множини).

Для деякої ґеш-функції y та вхідних повідомлень X_a, X_b буде $\delta_y(X_a, X_b) = 1$ якщо $y(X_a) = y(X_b)$, та $\delta_y(X_a, X_b) = 0$ в іншому випадку. Тобто

$\delta_y(\mathrm{X_a},\mathrm{X_b})=1$ лише тоді, коли ґеш-значення від вхідних повідомлень $\mathrm{X_a}$ та $\mathrm{X_b}$ дадуть колізію. Для кінцевої множини H визначимо

$$\delta_H(\mathrm{X_a},\mathrm{X_b})=\sum\nolimits_{y\in Y}\delta_y(\mathrm{X_a},\mathrm{X_b}).$$

Звідси $\delta_H(\mathrm{X_a},\mathrm{X_b})$ підраховує кількість ґеш-функцій (ключів ґешування), які дають колізію для визначених $\mathrm{X_a}$ та $\mathrm{X_b}$.

Згідно з [6] ймовірність виникнення колізії для функції ґешування, яку побудовано на основі поліноміальної схеми, визначається як:

$$P_k=(n-1)/\left|H\right|, \tag{1}$$

де n - довжина вхідного повідомлення (кількість блоків вхідного повідомлення), $\left|H\right|$ - потужність множини ключів (потужність двійкового поля Галуа).

Дійсно, якщо правило ґешування повідомлення $X=(X_1\|X_2\|...\|X_n)$ задається через обчислення у кінцевому полі $GF(2^k)$ значення поліному

$$Y_n=X_1\cdot H_j^{n-1}\oplus X_2\cdot H_j^{n-2}\oplus...\oplus X_n\cdot H_j^0, \tag{2}$$

де $H_j\in H$ - значення ключа ґешування, $X_i,H_j\in GF(2^k)$, тоді колізія (співпадіння) ґеш-значення Y_n із ґеш-кодом Y_n^*, який відповідає іншому повідомленню $X^*=(X_1^*\|X_2^*\|...\|X_n^*)\neq X=(X_1\|X_2\|...\|X_n)$ буде відповідати випадку тотожності:

$$X_1\cdot H_j^{n-1}\oplus X_2\cdot H_j^{n-2}\oplus...\oplus X_n\cdot H_j^0=X_1^*\cdot H_j^{n-1}\oplus X_2^*\cdot H_j^{n-2}\oplus...\oplus X_n^*\cdot H_j^0$$

для будь-якого введеного ключа ґешування H_j.

Останнє рівняння перепишемо у канонічному вигляді:

$$(X_1\oplus X_1^*)\cdot H_j^{n-1}\oplus(X_2\oplus X_2^*)\cdot H_j^{n-2}\oplus...\oplus(X_n\oplus X_n^*)\cdot H_j^0=0, \tag{3}$$

отже колізія ґеш-значень, тобто подія $Y_n=Y_n^*\big|_{X\neq X^*}$ буде виникати лише тоді, коли H_j є коренем рівняння (3).

Однак за основною теоремою алгебри будь який многочлен степеня $n-1$ має точно $n-1$ коренів з врахуванням їхньої кратності, тобто на безлічі значень $H_j\in H$ не більше $n-1$ різних ключів ґешування будуть обертати в нуль ліву

частину рівності (3). Таким чином, при рівноймовірному обранні ключів ґешування $H_j \in H$ ймовірність колізії $P_к = P\left(Y_n = Y_n^* \middle|_{X \neq X^*}\right)$ буде визначатися за виразом (1).

Беручи до уваги зазначене вище, можна зробити висновок, що, обираючи підходяще значення потужності $|H|$ множини ґеш-функцій (ключів) та довжину вхідного повідомлення, можна досягти необхідного рівня ймовірності виникнення колізії. Графіки залежності ймовірності виникнення колізії від довжини повідомлення та потужності множини ключів для поліноміальної схеми представлено на рис. 5.

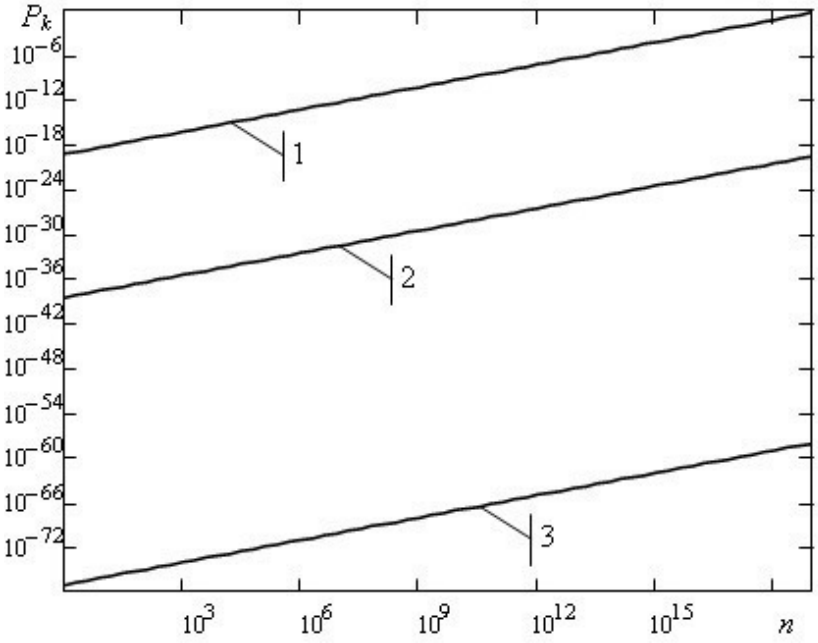


Рис. 5. Залежність ймовірності виникнення колізії від кількості блоків вхідного повідомлення: 1) $|H| = 2^{64}$ *; 2)* $|H| = 2^{128}$ *; 3)* $|H| = 2^{256}$.

Як видно з наведених залежностей ймовірність колізій при застосуванні поліноміальної схеми ґешування значно підвищується із збільшенням довжини вхідного повідомлення. При фіксованій довжині ключів ґешування (та, відповідно, потужності $|H|$) ця залежність накладає додаткові обмеження на довжину повідомлень, які ґешуються. Наприклад, для заданої ймовірності виникнення колізій $P_k = 2^{-32}$ та при довжині ключа ґешування 128 бітів загальна кількість блоків повідомлення повинна задовольняти вимозі $n \le 2^{96}$.

Обчислення ґеш-значення за правилом (2) має певні недоліки. Наприклад, якщо вхідне повідомлення складається лише з одного блоку, тоді формула (2) прийме вигляд $Y_1 = X_1 \cdot H_j^0 = X_1$ і правило ґешування є тотожністю, тобто ґеш-

значення буде дорівнювати блоку повідомлення і ключ ґешування H_j при обчисленні Y_1 зовсім не використовується. Можливо саме тому в специфікації GCM & GMAC [4] застосовується дещо змінена поліноміальна форма обчислення ґеш-коду, а саме (див. рис. 4):

$$Y_m = X_1 \cdot H_j^m \oplus X_2 \cdot H_j^{m-1} \oplus ... \oplus X_m \cdot H_j^1. \quad (4)$$

Випадок колізії буде спостерігатися так само при виконанні рівності:

$$X_1 \cdot H_j^m \oplus X_2 \cdot H_j^{m-1} \oplus ... \oplus X_m \cdot H_j^1 = X_1^* \cdot H_j^m \oplus X_2^* \cdot H_j^{m-1} \oplus ... \oplus X_m^* \cdot H_j^1,$$

яку запишемо у вигляді

$$(X_1 \oplus X_1^*) \cdot H_j^m \oplus (X_2 \oplus X_2^*) \cdot H_j^{m-1} \oplus ... \oplus (X_n \oplus X_n^*) \cdot H_j^1 = 0,$$

що після скорочення на H_j^1 при $n = m-1$ повністю відповідає (3) із оцінкою ймовірності колізій (1).

Таким чином, правило обчислення ґеш-значень (4) є за колізійними властивостями тотожним правилу (2), але навіть у випадку ґешування одного блоку обчислений ґеш-код $Y_1 = X_1 \cdot H_j$ залежить як від X_1, так і від значення ключа ґешування H_j.

Втім, слід відмітити недоліки застосованого правила (4). Якщо ключ ґешування H_j дорівнює нулю, тоді ґеш-код буде також дорівнювати нулю для будь якого вхідного повідомлення. Це накладає додаткові обмеження на схему формування ключових даних схеми ґешування, втім специфікацією режиму GCM & GMAC ніяких обмежень та вказівок з цього приводу не наводиться [4]. Вказано лише, що субключ ґешування H_j формується як зашифрована двійкова послідовність (див. рис. 1, 2).

Розглянемо цей випадок більш докладніше, бо формування нульового ключа ґешування $H_j = 0$, як з'ясувалося, призводить до виродженої роботи поліноміальної схеми $GHASH_H$, і формована імітовставка T у цьому випадку зовсім не буде залежати від ґеш-коду повідомлення, а визначатиметься лише значенням вектору ініціалізації (синхропосилки) IV (див. рис. 1, 2).

Оцінимо ймовірність виникнення нульового субключа ґешування, тобто ймовірність такої події, коли при шифруванні нульового вектору 0^{128} буде отримано значення $H_j = 0$. Для цього скористаємося деякими визначеннями та поняттями теорії підстановок [8].

4. Оцінка ймовірності виникнення нульового субключа гешування $GHASH_H$.

Розглянемо множину всіх бієктивних перетворень множини $Y = \{y_1, y_2, ..., y_n\}$ саму в себе. Ці перетворення, які мають назву підстановок степеня n, утворюють групу відносно операції послідовного виконання перетворень. Така група має назву симетричної групи підстановок степеня n та позначається як S_n [8]. Її потужність визначається потужністю множини всіх підстановок степеня n, тобто дорівнює $n!$.

Кожній підстановці $s \in S_n$ відповідає єдина підстановка $s^{-1} \in S_n$, така, що $s^{-1} \cdot s(y) = s \cdot s^{-1}(y) = e(y)$, $y \in Y$, де $e(y) \in S_n$ - одинична підстановка, тобто $e(y) = y$ для всіх $y \in Y$.

Введемо наступні позначення: $s \cdot s \cdot ... \cdot s = s^k$, $s^{-1} \cdot s^{-1} \cdot ... \cdot s^{-1} = s^{-k}$, де добутки містять k множників. Відповідно маємо $s^k \cdot s^{-k} = s^{-k} \cdot s^k = s^0 = e$.

Множина підстановок степеня n, яка є замкнутою відносно операції множення та обчислення оберненого для $s \in S_n$ елементу $s^{-1} \in S_n$, має назву групи підстановок. Кожна така група є підгрупою симетричної групи S_n [8].

Розглянемо деяку підстановку $s \in S_n$, яка діє на множині Y. Визначимо на множині Y бінарне відношення, при цьому будемо вважати $y \sim y'$ для $y, y' \in Y$ якщо існує таке j, що $y' = s^j(y)$. Це бінарне відношення є рефлексивним, симетричним та транзитивним, тобто є відношенням еквівалентності. Дійсно, відповідно до [8] маємо:

- $y \sim y$, оскільки $y = s^0(y) = e(y)$;

- із умови $y \sim y'$ витікає $y' \sim y$, оскільки із рівності $y' = s^j(y)$ випливає, що $y = s^{-j}(y')$;

- із $y \sim y'$ та $y' \sim y''$ витікає, що $y \sim y''$, бо з рівностей $y' = s^j(y)$ та $y'' = s^i(y')$ випливає, що $y'' = s^i(s^j(y)) = s^{i+j}(y)$.

Цикл s_i підстановки $s \in S_n$ довжини l_i визначається наступним чином:

$$s_i = (y, s_i(y), s_i^2(y), ..., s_i^{l_i - 1}(y)),$$

де $s_i^{l_i}(y) = y$.

Таким чином, довільну підстановку $s \in S_n$ можна розкласти на відповідні цикли [8]:

$$s = (y_1, s_1(y_1), s_1^2(y_1), ..., s_1^{l_1 - 1}(y_1))...(y_k, s_k(y_k), s_k^2(y_k), ..., s_k^{l_k - 1}(y_k)). \quad (5)$$

Наприклад, підстановка s степеня 4 виду

$$s = \begin{pmatrix} y_1 & y_2 & y_3 & y_4 \\ s(y_1) & s(y_2) & s(y_3) & s(y_4) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$$

може бути подана у вигляді розкладу на 3 цикли:

$$s_1 = (y_1) = (1),\ l_1 = 1;$$
$$s_2 = (y_2, s_2(y_2)) = (2,4),\ l_2 = 2;$$
$$s_3 = (y_3) = (3),\ l_3 = 1,$$

тобто маємо наступний розклад:

$$s = (y_1)(y_2, s_2(y_y))(y_3) = (1)(2,4)(3).$$

Загалом, підстановка $s \in S_n$ належить до циклового класу $\{1^{\alpha_1} 2^{\alpha_2} ... n^{\alpha_n}\}$, якщо вона містить α_1 циклів довжини 1, α_2 циклів довжини 2, і так далі, тобто:

$$s = (y_1)(y_2)...(y_{\alpha_1})(y'_1, y''_1)(y'_2, y''_2)...(y'_{\alpha_2}, y''_{\alpha_2})...,$$
$$1\alpha_1 + 2\alpha_2 + ... + n\alpha_n = n.$$

Позначимо через $C(\alpha_1, \alpha_2, ..., \alpha_n)$ число підстановок в цикловому класі $\{1^{\alpha_1} 2^{\alpha_2} ... n^{\alpha_n}\}$, а через $C(n,k)$ - число підстановок степеня n, які мають k циклів. Тоді маємо [8]:

$$C(\alpha_1, \alpha_2, ..., \alpha_n) = \frac{n!}{1^{\alpha_1} 2^{\alpha_2} ... n^{\alpha_n} \alpha_1! \alpha_2! ... \alpha_n!},$$

$$C(n,k) = \sum_{\substack{1\alpha_1 + 2\alpha_2 + ... + n\alpha_n = n \\ \alpha_1 + \alpha_2 + ... + \alpha_n = k,\ \alpha_i \geq 0}} \frac{n!}{1^{\alpha_1} 2^{\alpha_2} ... n^{\alpha_n} \alpha_1! \alpha_2! ... \alpha_n!} = \left| s(n,k) \right|, \qquad (6)$$

де $s(n,k)$ - числа Стірлінга першого роду, які визначаються через співвідношення:

$$x(x-1)...(x-n+1) = \sum_{k=0}^{n} s(n,k) x^k.$$

На множині всіх підстановок степеня n, які утворюють симетричну групу S_n, задамо рівномірний ймовірнісний розподіл, тобто кожній вибраній підстановці $s \in S_n$ поставимо у відповідність ймовірність її обрання, що дорівнює $1/n!$. За сучасними поглядами симетричної криптографії така множина рівноймовірних відображень відповідає уявленню про «ідеальний» шифр, бо якщо обрання окремої підстановки $s \in S_n$ пов'язати із значенням введеного ключа шифрування, тоді отримане перетворення буде відповідати випадковому і рівномірно вибраному шифр тексту для кожного відкритого

тексту при будь-якому ключі, тобто на всіх можливих варіантах відображень відкритого тексту у шифрограму.

За визначенням блочний симетричний шифр є функцією відображення множини текстів і множини ключів в множину шифртекстів: E: K × M → E, де K, M та E – множини ключів, відкритих та шифртекстів, відповідно. Для шифру AES потужність множини ключів $|K| \in \{2^{128}, 2^{192}, 2^{256}\}$, а $|M| = |E| = 2^{128}$. Оскільки при зашифруванні необхідно мати можливість відновити текст за допомогою ключа, потрібно, щоб для всіх ключів $k \in K$ функція зашифрування була перестановкою (підстановкою), тобто відображення E: K × M → E повинно бути бієктивним.

На практиці для довільного n-бітового блокового шифру існує $2^n!$ можливих перестановок відкритого тексту. Практично це означає, що кількість бітів ключа, яку необхідно для отримання всіх можливих перестановок, становить близько $\ln 2^n! \approx n \cdot 2^n$ бітів[1]. Втім розмір ключа більшості блочних шифрів не перевищує невеликого числа, кратного розміру блоку, відповідно такі шифри можуть забезпечити лише невелику частку від повної кількості можливих перестановок. Наприклад, для 128-бітного шифру AES маємо $2^{128}! \approx 2^{128 \cdot 2^{128}}$ можливих перестановок 128-бітових блоків, з яких, в залежності від довжини ключа, використовується тільки 2^{128}, 2^{192} або 2^{256} перетворень. Таким чином, кожен шифр є деяка підмножина повної множини всіх можливих підстановок, що діють на множині блоків оброблюваних даних. Основне припущення, яке приймається при обґрунтуванні стійкості симетричного криптоперетворення полягає саме у збереженні ймовірнісних властивостей випадкової підстановки, тобто припускається, що хоча при шифруванні і застосовується обмежений набір підстановок із S_n, та певні розподіли ймовірностей елементів цієї підмножини відповідають властивостям випадково и рівномірно обраної підстановки із всієї множини S_n.

Проведемо дослідження цих розподілів з метою оцінки ймовірності виникнення нульового шифр тексту при шифруванні нульового відкритого тексту. Для цього розглянемо випадкову величину ξ_n, яка дорівнює числу циклів в випадково вибраній підстановці $s \in S_n$. Оцінимо ймовірність випадкової події $\xi_n = k$, тобто такого випадку, коли у випадково вибраній підстановці буде спостерігатися точно k циклів (див. вираз (5)). З формули (6) безпосередньо випливає вираз для точного розподілу ймовірностей через числа Стірлінга першого роду:

$$P(\xi_n = k) = \frac{C(n,k)}{n!} = \frac{|s(n,k)|}{n!},\ k = 0,1,...,n.$$

В роботах [8, 15] отримано математичне очікування $M\xi_n$ та дисперсію $D\xi_n$ випадкової величини ξ_n:

[1] За формулою Стірлінга $\ln(x!) = x\ln(x) - x - O(\ln(x))$

$$M\xi_n = \sum_{j=1}^{n} \frac{1}{j} = \ln n + C + o(1),\ C = 0,5772...,$$

$$D\xi_n = \sum_{j=1}^{n} \frac{1}{j} - \sum_{j=1}^{n} \frac{1}{j^2} = \ln n + C + o(1),$$

крім того показано, що при $n \to \infty$ випадкова величина $\xi'_n = (\xi_n - \ln n)/(\ln n)$ розподілена асимптотично нормально з параметрами $(0,\ 1)$, тобто

$$\lim_{n\to\infty} P(\xi'_n < u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-y^2/2} dy.$$

В роботах [11-14] досліджено емпіричний розподіл ймовірності виникнення циклу певної довжини у зменшених моделях шифру, встановлено, що цей розподіл дуже близький до розглянутого теоретичного розподілу випадкової підстановки, тобто за цим критерієм можна стверджувати, що шифр за розподілом кількості циклів подібний властивостям випадкової підстановки. В той же час для оцінки ймовірності виникнення нульового субключа ґешування в схемі GCM & GMAC потрібна інша характеристика шифру, а саме розподіл числа циклів заданої довжини. Відповідно до [8], ця характеристика у випадковій підстановці визначається наступним чином.

Позначимо як $\chi_{n,l}$ число циклів довжини l у випадковій рівно ймовірній підстановці степеня n. Тоді розподіл ймовірностей випадкової події $\chi_{n,l} = k$ визначається як:

$$P(\chi_{n,l} = k) = \frac{1}{l^k k!} \sum_{j=0}^{[n/l]-k} \frac{(-1)^j}{l^j j!},\ k = 0,1,...,\lfloor n/l \rfloor. \qquad (7)$$

При $n \to \infty$ випадкова величина $\chi_{n,l}$ має в межі розподіл Пуасона з параметрами $\lambda = 1/l$, тобто

$$\lim_{n\to\infty} P(\chi_{n,l} = k) = \frac{1}{l^k k!} e^{-1/l},\ k = 0,1,....$$

Скористаємося формулою точного розподілу ймовірностей випадкової події $\chi_{n,l} = k$ у вигляді (7). Значення $n!P(\chi_{n,l} = k)$ відповідає кількості підстановок, які містять k циклів довжини l. Нас цікавить кількість підстановок, які обов'язково мають цикли довжини $l = 1$, причому ці цикли повинні переводити фіксований елемент y з множині Y сам у себе. Тобто треба оцінити кількість таких підстановок $s \in S_n$, які для визначеного $y \in Y$ містять цикли $s(y) = y$ довжини $l = 1$. В криптографії при розгляді блокових симетричних

криптоперетвореннь такі випадки прийнято називати фіксованими точками підстановки [11].

Для $l=1$ формула (7) прийме вигляд

$$P(\chi_{n,l=1}=k)=\frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!},\ k=0,1,...,n,$$

причому для $k=1,...,n$ кожен з $n!P(\chi_{n,l=1}=k)$ випадків для фіксованого $y\in Y$ буде спостерігатися

$$\frac{C_{n-1}^{k-1}}{C_n^k}=\frac{(n-1)!}{(k-1)!(n-k)!}\frac{k!(n-k)!}{n!}=\frac{k}{n}$$

разів, тобто кількість фіксованих точок $s(y)=y$ для визначеного $y\in Y$ буде визначатися за формулою:

$$\sum_{k=1}^{n}n!P(\chi_{n,l=1}=k)\frac{C_{n-1}^{k-1}}{C_n^k}=\sum_{k=1}^{n}\left(\frac{(n-1)!}{(k-1)!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!}\right)=(n-1)!,$$

а відповідна ймовірність фіксованої точки у випадковій рівноймовірній підстановці степеня n матиме вигляд:

$$\sum_{k=1}^{n}P(\chi_{n,l=1}=k)\frac{C_{n-1}^{k-1}}{C_n^k}=\frac{(n-1)!}{n!}=\frac{1}{n}. \qquad (8)$$

Останній вираз може бути отриманий значно простіше з тривіальних комбінаторних міркувань. Дійсно, якщо на множині $Y=\{y_1,y_2,...,y_n\}$ зафіксувати m елементів, тоді можливі $(n-m)!$ варіантів перестановок решти елементів. Тобто на всій множині підстановок з S_n при їх випадковому рівноймовірному розподілі ймовірність обрати підстановку з m фіксованими точками буде дорівнювати

$$\frac{(n-m)!}{n!}=\frac{1}{(n-m+1)(n-m+2)...n}=\frac{1}{(n)_m}, \qquad (9)$$

що при $m=1$ співпадає з (8) [2].

Припустимо, що n-бітний блоковий симетричний шифр, який складається з деякої підмножини симетричної групи підстановок степеня 2^n, відповідає ймовірнісним властивостям випадкової підстановки, зокрема приймемо

[2] $(n)_m=n(n-1)...(n-m+1)$ - загальноприйняте позначення спадаючого факторіалу

припущення, що на всій множині ключів шифру, які задають обрання конкретної підстановки s із S_{2^n}, ймовірність обрати підстановку з m фіксованими точками буде визначатися за (9), тобто буде дорівнювати

$$\frac{(2^n - m)!}{2^n!} = \frac{1}{(2^n)_m}.$$

Тоді кількість ключів шифру, які призведуть до появи m фіксованих точок $s(y) = y$ для визначених блоків відкритого тексту $y \in Y$ буде визначатися за формулою $N(m) = \frac{2^k}{(2^n)_m}$, де k - бітова довжина ключа шифру.

Зокрема, для випадку $m = 1$ маємо

$$N(m = 1) = \frac{2^k}{(2^n)_1} = 2^{k-n},$$

тобто:

- якщо $k < n$ з високою ймовірністю можна стверджувати, що шифр не містить фіксованих точок, тобто на всій множині ключів шифрування не знайдеться жодного, який призведе до появи хоча б одного нешифрованого відкритого тексту;

- якщо $k = n$ з високою ймовірністю знайдеться один ключ із $|K| = 2^k$ можливих, який призведе до появи фіксованої точки, тобто отримаємо випадок нешифрованого відкритого тексту;

- якщо $k > n$ кількість ключів, які призводять до появи фіксованої точки, стрімко зростає, їх кількість експоненційно залежить від бітової довжини ключа.

У випадку застосування шифру AES маємо таке:

- при довжині ключа $k = 128$ бітів один ключ буде створювати фіксовану точку, тобто, наприклад, буде спостерігатися випадок, коли при шифруванні нульового вектору 0^{128} буде отримано значення $H_j = 0$;

- при довжині ключа $k = 192$ бітів загалом $2^{k-n} = 2^{64}$ ключів будуть створювати фіксовану точку, тобто випадок із шифруванням нульового вектору 0^{128} в нульове значення $H_j = 0$ буде спостерігатися 2^{64} разів;

- при довжині ключа $k = 256$ бітів вже $2^{k-n} = 2^{128}$ ключів відповідати випадку шифрування нульового вектору 0^{128} в те ж саме значення, тобто в $H_j = 0$.

Отримані оцінки дозволяють стверджувати, що при виконанні зроблених припущень, ймовірність виникнення фіксованої точки шифру не залежить від довжини ключа, вона визначається лише довжиною блоків даних, які обробляє шифр. Цей висновок збігається із загальною інтерпретацією шифру як випадкової підстановки, фіксовані точки в якій – звичайне явище. Однак для розглянутого вище ключового поліноміального ґешування випадки фіксованих

точок для нульового блоку є неприпустимими, бо це повністю спотворює схему формування ґеш-значеннь, які вже не будуть залежати від вихідних даних, колізійні властивості порушуються і не відповідають теоретичним оцінкам. Очевидно, що при збільшенні потужності множини ключів кількість фіксованих точок також буде зростати і число так би мовити «слабких» ключів, які призводять до виродженої роботи функції ґешування збільшується пропорційно потужності ключового простору.

Розглянемо тепер вплив останнього шару перетворень режиму GCM & GMAC, зокрема шифрування/розшифрування отриманого за допомогою функції $GHASH_H$ ґеш-значення, оцінимо ймовірності показники формованих імітовставок.

5. Дослідження колізійних властивостей формованих режимом GCM & GMAC імітовставок.

Останній шар перетворень режиму GCM & GMAC полягає у застосуванні функції $GCTR_K$, яка по суті є деякою варіацією режиму гамування CTR [3, 4]. Тобто до отриманого на попередньому шарі перетворень ґеш-значення додається результат зашифрування деякого вектору J_0, який отримано із вектору ініціалізації IV (див. рис. 1, 2). Тобто фактично, до сформованого ґеш-значення додається зашифрована константа, яка формується встановленим порядком.

Відповідно до специфікації режиму GCM & GMAC на формування вектору ініціалізації (за вітчизняною термінологією - синхропосилка) накладаються такі обмеження:

- ймовірність того, що при обчисленні імітовставок для двох різних вихідних даних будуть застосовані однакові вектори ініціалізації при однакових ключах не повинна бути більшою ніж 2^{-32};

- для кожного заданого ключа повне число формованих імітовставок для будь яких вихідних повідомлень не повинно перевищувати 2^{32}.

Перша умова визначає максимальну ймовірність застосування однакових параметрів криптографічного перетворення для різних вихідних даних. Тобто навіть якщо при введеному ключі та векторі ініціалізації виникне колізія (співпадіння імітовставок для різних вихідних даних), ця подія буде повторена не частіше ніж один раз на 2^{32} [3].

Друга умова визначає максимальну кількість можливих застосувань режиму GCM & GMAC для кожного з введених ключів, тобто ця умова задає обмеження на термін дії ключа. Разом з першою умовою вона визначає, що кожен ключ буде поєднаний із деяким вектором ініціалізації тільки один раз, тобто при обробленні різних вихідних даних одні й ті ж пари «ключ - вектор ініціалізації» застосовуватися двічі не будуть ніколи.

Таким чином можна зробити наступні висновки:

[3] Це твердження не враховує випадку, коли колізія імітовставок відбувається при різних параметрах криптоперетворення

- для кожного введеного ключа при обробленні різних вихідних даних кожен раз будуть застосовуватися різні вектори ініціалізації;
- для кожного введеного вектору ініціалізації при обробленні різних вихідних даних кожен раз будуть застосовуватися різні ключі;
- одні й ті ж самі вихідні дані можуть бути оброблені із застосуванням одного і того ж самого вектору ініціалізації, але ключі кожен раз повинні бути різними;
- одні й ті ж самі вихідні дані можуть бути оброблені із застосуванням одного і того ж самого ключа, але вектори ініціалізації кожен раз повинні бути різними.

Тобто якщо повторюється ключ – тоді не повинні повторюватися вектори ініціалізації, а якщо повторюється вектор ініціалізації – повинні бути різними ключі.

Позначимо через $Y_m = GHASH_H(X_1 \| X_2 \| \ ... \| X_m)$ результат поліноміального ґешування як на рис. 4, а через $T = MSB_t\left(GCTR_K(J_0, Y_m)\right) = MSB_t(Z \oplus Y_m)$ формовану імітовставку як на рис. 1, 2, де $Z = CHIP_K(J_0)$ результат зашифрування J_0 на ключі K.

Очевидно, що колізійні властивості формованих імітовставок залежать як від властивостей ґеш-значень Y_m, так і від результатів зашифрування Z. Колізійні властивості Y_m було оцінено у розділі 3. Оцінимо ймовірність співпадіння Z та кінцевих результатів T.

За визначенням блоковий симетричний шифр є бієктивним відображенням, тому різні вихідні значення J_0 після зашифрування будуть співставленні із різними значеннями Z, тобто колізій виникати не буде. Із умов формування векторів ініціалізації, які розглянуто вище, слідує, що навіть для однакових вихідних повідомлень $\left(X_1 \| X_2 \| ... \| X_m\right)$ і однакових ключів відповідні J_0 будуть різними, тобто ймовірність колізій проміжних значень Z буде завжди дорівнювати нулю.

Розглянемо тепер умови колізій імітовставок, зокрема такий випадок, коли для різних вихідних повідомлень $\left(X_1 \| X_2 \| ... \| X_m\right)$ на однаковому ключі K відповідні результати $Z \oplus Y_m$ будуть тотожними. Така умова формально може бути подана у вигляді:

$$Z \oplus Y_m = Z' \oplus Y'_m, \tag{10}$$

де Z і Y_m є проміжними результатами формування імітовставки для вихідного повідомлення $\left(X_1 \| X_2 \| ... \| X_m\right)$, а Z' і Y'_m є відповідними значеннями для іншого повідомлення $\left(X'_1 \| X'_2 \| ... \| X'_m\right) \neq \left(X_1 \| X_2 \| ... \| X_m\right)$.

Вище показано, що для таких повідомлень завжди виконується нерівність $Z \neq Z'$ і умова (10) буде виконуватися лише при $Y_m \oplus Y'_m = Z \oplus Z' \neq 0^{128}$, тобто лише тоді, коли не буде виникати колізій ґеш-значень Y_m та Y'_m.

Останній висновок найбільш вражаючий, оскільки метою поліноміального ґешування було саме зменшення ймовірності колізій формованих ґеш-значень. Точне значення ймовірності колізій формованих імітовставок буде визначатися комбінаторними властивостями шифру, тобто буде залежати від кількості випадків $Y_m \oplus Y'_m = Z \oplus Z'$. Однак можна з впевненістю стверджувати, що колізійні властивості імітовставок не будуть повторювати відповідні властивості ґеш-значень, що досліджувалися у розділі 3.

6. Висновки.

Проведений аналіз показав, що для використання в режимі шифрування Galois/Counter Mode and GMAC визначеної функції поліноміального ґешування $GHASH_H$ повинні бути застосовані певні обмеження, зокрема не можна використовувати нульовий субключ ґешування, оскільки при цьому значення функція завжди обертається в нуль для будь якого вхідного повідомлення, що створює передумови для зниження рівня імітостійкості. Це є неприпустимим, бо схема формування ґеш-значень спотворюється, формовані імітовставки не будуть залежати від вихідних даних колізійні властивості порушуються і не відповідають теоретичним оцінкам.

В ході досліджень було з'ясовано, що випадок формування нульового субключа поліноміального ґешування $GHASH_H$ виникає при наявності фіксованих точок шифру, тобто таких ключів, які шифрують нульову послідовність саму у себе. Отримані оцінки показують, що ймовірність виникнення фіксованої точки шифру не залежить від довжини ключа, вона визначається лише довжиною блоків даних, які обробляє шифр. При збільшенні потужності множини ключів кількість фіксованих точок також зростає і число «слабких» ключів, які призводять до виродженої роботи функції ґешування збільшується пропорційно потужності ключового простору. У випадку застосування шифру AES при довжині ключа $k = 128$ бітів один ключ буде створювати фіксовану точку; при довжині ключа $k = 192$ бітів загалом $2^{k-n} = 2^{64}$ ключів будуть створювати фіксовану точку; при довжині ключа $k = 256$ бітів $2^{k-n} = 2^{128}$ ключів відповідають випадку шифрування нульового вектору 0^{128} в те ж саме значення, тобто відповідають формуванню нульового субключа ґешування.

Ймовірність колізій формованих імітовставок буде визначатися комбінаторними властивостями шифру, тобто буде залежати від кількості випадків, коли бітова різниця зашифрованих на одному ключі різних констант буде дорівнювати бітовій різниці зформованих на цих же ключах ґеш-значень. Випадок колізій імітовставок можливий лише для різних ґеш-значень. Емпірична оцінка числа колізій та відповідні оцінки ймовірностей із застосуванням зменшених моделей шифрів є перспективним напрямком подальших досліджень. Перспективним також бачиться перевірка зроблених припущень щодо тотожності деяких властивостей випадкової підстановки певним характеристикам застосовуваного шифру, зокрема це безпосередньо відноситься до вивчення ймовірнісних властивостей виникнення фіксованої точки та, що еквівалентно, кількості циклів довжини $l = 1$.

ЛІТЕРАТУРА

1. ГОСТ 28147-89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. – М.: Изд-во стандартов, 1989. – 20 с.
2. ГОСТ Р ИСО/МЭК 10116-93. Информационная технология. Режимы работы для алгоритма n-разрядного блочного шифрования. [Електронний ресурс]. Режим доступу: http://docload.spb.ru
3. ISO/IEC 10116. Information technology – Security techniques – Modes of operation for an n-bit block cipher. [Електронний ресурс]. Режим доступу: http://www.iso.org
4. NIST Special Publication 800-38D. Block Cipher Modes. [Електронний ресурс]. Режим доступу: http://csrc.nist.gov
5. Stinson D.R. Universal hashing and authentication codes // Designs, Codes and Cryptography – 1994. - Volume 4, Issue 3. - pp 369-380.
6. Polynomial hashing: 4,588,985 United States Patent: H 03 M 7/00, field of search 340/347 DD / Carter J. L., Wegman M. N.; International Business Machines Corporation, Armonk, N.Y. – filed Dec. 30, 1983 – May 13, 1986
7. Raphael Chung-Wei Phan. Mini Advanced Encryption Standard (Mini-AES): A testbed for Cryptanalysis Students // Cryptologia, XXVI(4), October 2002, pp. 283-306.
8. Сачков В.Н. Введение в комбинаторные методы дискретной математики. – М.: Наука. Гл. ред. физ.-мат. лит., 1982. – 384 с.
9. Тронин С.Н. Введение в теорию групп. – Казань: Казанский государственный университет, 2006. – 100с.
10. Горбенко І.Д., Горбенко Ю.І. Прикладна криптологія. Теорія. Практика. Застосування: Підручник для вищих навчальних закладів. – Харків: Вид-во «Форт», 2013. – 880с.
11. Долгов В.И., Лисицкая И.В., Руженцев В.И. Анализ циклических свойств блочных шифров // Прикладная радиоэлектроника – 2007. – Т.6, №2 – С. 257-263.
12. Кузнецов А.А., Лисицкая И.В., Исаев С.А. Линейные свойства блочных симметричных шифров, представленных на украинский конкурс // Прикладная радиоэлектроника. – 2011. – Т.10, №2 – С. 135-140.
13. Сорока Л.С., Кузнецов А.А., Московченко И.В., Исаев С.А. Исследование дифференциальных свойств блочно-симметричных шифров. // Системи обробки інформації. – Х: ХУПС. –2010 – Вип. 6(87). – С. 286 – 294.
14. Долгов В.И., Родинко М.Ю. Блочные симметричные шифры – случайные подстановки. Комбинаторные показатели // Прикладная радиоэлектроника – 2013. – Т.12, №2 – С. 236-239.
15. Александров П.С. Введение в теорию групп. – М.: Наука, – 1980. – 145с.

Надійшла - 14.05.2014.

УДК 004.652

Математические основания реляционных баз данных. Часть 2: свойства обобщенных табличных операций

Н. Д. Кахута
Университет экономики и права «КРОК», Украина

Статья посвящена созданию фрагмента теории табличных алгебр, построенных на основе классических реляционных алгебр Кодда. Отличительной особенностью применяемой техники является использование свойств теоретико-множественных конструкций и их перенесение на табличный случай. Такое перенесение свойств возможно ввиду наличия простых представлений сигнатурных операций в терминах указанных теоретико-множественных конструкций. Наличие указанных представлений делает возможным рассмотрение обобщенной табличной алгебры, получаемой снятием требований, во-первых, конечности таблиц и, во-вторых, односхемности строк таблицы.

Ключевые слова: *реляционные алгебры Кодда, полный образ, ограничение, прямое произведение, совместность, коинициальность, обобщенные табличные алгебры.*

Стаття присвячена створенню фрагмента теорії табличних алгебр, побудованих на основі класичних реляційних алгебр Кодда. Особливістю застосовуваної техніки є використання властивостей певних теоретико-множинних конструкцій та їх перенесення на табличний випадок. Таке перенесення властивостей можливо через наявність простих зображень сигнатурних операцій в термінах теоретико-множинних конструкцій. Наявність зазначених зображень робить можливим розгляд узагальненої табличної алгебри, яка одержується зняттям вимог, по-перше, скінченності таблиць і, по-друге, односхемності рядків таблиці.

Ключові слова: *реляційні алгебри Кодда, повний образ, обмеження, прямий добуток, сумісність, коініціальность, узагальнені табличні алгебри.*

The article is devoted to creation of a fragment of the theory for table algebras, which constitute a generalization of classical Codd's relational algebras. A distinctive feature of this technique is that the set-theoretic properties of some specific constructions are used and transferred to the table case. Such transfer is possible because there exist simple representations of signature operations in terms of these set-theoretic constructions. The fact of these representations existence allows to get rid of some requirements to generalized table algebra: firstly, tables have not to be the finite set of rows, and secondly, rows in a table can have different schemas.

Key words: *Codd's relational algebras, a whole image, restriction, Cartesian product, consistency, coinitiality, generalized table algebras.*

1. Введение

Статья продолжает работу [1] и посвящена созданию фрагмента теории табличных алгебр [2-11], построенных на основе классических реляционных алгебр Кодда [12-18]. Отличительной особенностью применяемой техники является использование свойств теоретико-множественных конструкций полного образа, ограничения, обобщенного прямого произведения, бинарных отношений совместности и коинициальности (конфинальности) [19-22] и перенесение свойств указанных конструкций для табличного случая на основе представлений табличных операций, полученных в первой части работы [1].

2. Свойства табличных операций, полурешетка таблиц

В следующих утверждениях приведены свойства операций табличных алгебр, для установления которых явно используются соответствующие свойства указанных выше теоретико-множественных конструкций.

2.1. Свойства проекции, соединения, насыщения и переименования

Предложение 1 (монотонность проекции, соединения, насыщения, переименования; дистрибутивность проекции, соединения, переименования относительно объединений). Операции проекции, соединения, переименования, насыщения монотонны по теоретико-множественному включению:

1) $t_1 \subseteq t_2 \Rightarrow \pi_X(t_1) \subseteq \pi_X(t_2)$ (монотонность проекции);

2) $t_1 \subseteq t_1' \,\&\, t_2 \subseteq t_2' \Rightarrow t_1 \otimes t_2 \subseteq t_1' \otimes t_2'$ (монотонность соединения по совокупности аргументов);

3) $t_1 \subseteq t_2 \Leftrightarrow Rt_{\xi}(t_1) \subseteq Rt_{\xi}(t_2)$, где t_1, t_2 имеют одинаковую ξ-допустимую схему (монотонность переименования);

4) $t_1 \subseteq t_2 \Rightarrow D_{A,t_1} \subseteq D_{A,t_2}$ (монотонность активного домена); $t_1 \subseteq t_2 \Rightarrow C(t_1) \subseteq C(t_2)$ (монотонность насыщения).

Операции проекции, соединения, переименования дистрибутивны относительно теоретико-множественного объединения:

5) $\pi_X(\bigcup\limits_i t_i) = \bigcup\limits_i \pi_X(t_i)$, здесь схемы таблиц t_i одинаковы (дистрибутивность проекции относительно объединений);

6) $t \otimes (\bigcup\limits_i t_i) = \bigcup\limits_i t \otimes t_i$ (дистрибутивность соединения относительно объединений по второму аргументу), здесь схемы таблиц t_i одинаковы (для первого аргумента аналогично);

7) $Rt_{\xi}(\bigcup\limits_i t_i) = \bigcup\limits_{\eta[R]} Rt_{\xi}(t_i)$, где таблицы t_i имеют одинаковую ξ-допустимую схему R (дистрибутивность переименования относительно объединений). □

Доказательство. Используя представления соответствующих операций ([1, предложение 1]), монотонность полного образа относительно объединений ([11, подраздел 1.3, предложение 1.3.1; 19]) и обобщенного прямого произведения ([20], это свойство применяется при доказательстве монотонности насыщения) доказываются пп. 1-3; при доказательстве дистрибутивности операций в пп. 5-7 надо также использовать представления операций и дистрибутивность полного образа относительно объединений ([11, подраздел 1.3, предложение 1.3.1;19]). ▫

Установим монотонность активного домена и насыщения. Из представления насыщения (через активный домен и обобщенное прямое произведение, см. [1, предложение 1, п. 3]) и монотонности обобщенного прямого произведения [20] следует, что из монотонности активного домена вытекает монотонность насыщения; поэтому проверим монотонность активного домена.

Для этого воспользуемся следующим очевидным представлением активного домена

$$D_{A,t} = I_A[t], \tag{2.1}$$

где $I_A : S \tilde{\to} D$, $I_A(s) \simeq s(A)$ для $s \in S$.

Содержательно говоря, частичная операция I_A строке s сопоставляет значение атрибута A в ней. Очевидно, что

$$dom\, I_A = \bigcup_{R:A\in R} S(R). \qquad (2.2)$$

Операция I_A по сути является (обобщенным) селектором.

Из представления (2.1), а также монотонности полного образа и вытекает монотонность активного домена относительно теоретико-множественного включения. ▫□

Установим еще некоторые интересные свойства данных операций. Ниже понятие (первичного, primary key) ключа таблицы понимается стандартно (см. например [11, подраздел 2.4, с. 43; 23]).

Предложение 2 (свойства проекции). Пусть R – схема таблиц t, t_i. Выполняются следующие соотношения:

1) $\pi_X(t) \in T(R \bigcap X)$; причем любая таблица схемы $R \bigcap X$ равна проекции по множеству атрибутов X некоторой таблицы схемы R (т.е.отображение $\pi_X : T(R) \to T(R \bigcap X)$ является сюръективным);

2) $R \subseteq X \Rightarrow \pi_X(t) = t$, причем для непустых таблиц эта импликация переходит в эквивалентность: $t \neq t_\varnothing \Rightarrow (R \subseteq X \Leftrightarrow \pi_X(t) = t)$; в частности, $\pi_R(t) = t$;

3) $\pi_X(t) = t_\varnothing \Leftrightarrow t = t_\varnothing$ (критерий пустоты проекции, сохранение пустой таблицы $t_\varnothing$);

4) $\pi_X(t) = t_\varepsilon \Leftrightarrow t \neq t_\varnothing \;\&\; X \bigcap R = \varnothing$; в частности, $\pi_X(t_\varepsilon) = t_\varepsilon$ (сохранение специальной таблицы t_ε);

5) $\pi_X(\bigcap_i t_i) \subseteq \bigcap_i \pi_X(t_i)$ (верхняя оценка проекции пересечений); X – ключ таблиці $\bigcup_i t_i \Rightarrow \pi_X(\bigcap_i t_i) = \bigcap_i \pi_X(t_i)$ (достаточное условие дистрибутивности проекции относительно пересечений);

6) $\pi_X(t_1 \setminus_R t_2) \supseteq \pi_X(t_1) \setminus \pi_X(t_2)$ (нижняя оценка проекции разности);

7) $\pi_X(\pi_Y(t)) = \pi_{X\bigcap Y}(t)$, или в операторном виде $\pi_X \circ \pi_Y = \pi_Y \circ \pi_X = \pi_{X\bigcap Y}$ (композиция проекций); в частности, $\pi_X(\pi_X(t)) = \pi_X(t)$ (идемпотентность проекции) и $\pi_X(t) = \pi_{X\bigcap R}(t)$.□

Доказательство. При проверке п. 1 надо воспользоваться определением проекции, свойствами ограничения ([11, подраздел 1.3, предложение 1.3.2; 19]): $\pi_1^2(s \mid X) = \pi_1^2 s \bigcap X = R \bigcap X$, где $s \in t$, а $t \in T(R)$. Кроме того, надо учесть, что любая строка схемы $R \bigcap X$ является ограничением по множеству X некоторой строки схемы R. ▫

Рассмотрим п. 2. Пусть $R \subseteq X$. Используя свойство ограничения из [11, подраздел 1.3, предложение 1.3.2; 19], имеем $\pi_X(t) = \{s \mid X \mid s \in t\} = \{s \mid s \in t\} = t$.

Пусть теперь $t \neq t_\varnothing$. Предположим, что $\pi_X(t) = t$ и установим включение $R \subseteq X$. Выберем строку $s_0 \in t$. Тогда $s_0 \mid X \in \pi_X(t)$, т.е. $s_0 \mid X = s'$ для некоторой строки $s' \in t$, следуя допущению. Согласно свойствам ограничения [11, подраздел 1.3, предложение 1.3.2; 19] имеем: $\pi_1^2(s_0 \mid X) = \pi_1^2 s_0 \bigcap X = R \bigcap X = \pi_1^2 s' = R$, откуда $R \bigcap X = R$, т.е. $R \subseteq X$. ▫

Достаточность в п. 3 вытекает из представления проекции через полный образ ([1, предложение 1, п. 1]) и сохранения пустого множества $\varnothing$ полным образом ([11, подраздел 1.3, предложение 1.3.1;19]). Для доказательства необходимости надо учесть те же факты и, кроме того, тотальность ограничения вида $\uparrow X : S \to S$. ▫

Рассмотрим п. 4. Пусть $\pi_X(t) = t_\varepsilon$. Из п. 3 вытекает, что $t \neq t_\varnothing$ (проверяется от противного), покажем, что $X \bigcap R = \varnothing$. Зафиксируем строку $s_0 \in t$. Тогда $s_0 \mid X \in \pi_X(t) = \{\varepsilon\}$, т.е. $s_0 \mid X = \varepsilon$. Возьмем проекции от обеих частей последнего равенства, тогда по свойствам ограничения [11, подраздел 1.3, предложение 1.3.2; 19] имеем $\pi_1^2 s_0 \bigcap X = R \bigcap X = \varnothing$. Обратная импликация устанавливается аналогично. ▫

Включения пп. 5-6 вытекают из представления проекции ([1, предложение 1, п. 1]), верхней оценки полного образа пересечения ([11, подраздел 1.3, предложение 1.3.1; 19]) и нижней оценки полного образа разности ([11, подраздел 1.3, предложение 1.3.1; 19]).

Для доказательства равенства з п. 5 (достаточного условия дистрибутивности проекции относительно пересечений) надо воспользоваться представлением проекции ([1, предложение 1, п. 1]), достаточным условием дистрибутивности полного образа относительно пересечений ([11, подраздел 1.3, предложение 1.3.1; 19]) и следующим очевидным свойством ключа: X – ключ таблицы $t \Leftrightarrow \uparrow X \mid t : t \to S$ – инъекция (т.е. операция $\uparrow X$ инъективна на таблице t). ▫

Рассмотрим п. 7. На основе свойств полного образа и ограничения ([11, подраздел 1.3, предложения 1.3.1, 1.3.2; 19]) получаем цепочку равенств: $\pi_X(\pi_Y(t)) = \uparrow X[\uparrow Y[t]] = \uparrow X \circ \uparrow Y[t] = (\uparrow X \bigcap Y)[t] = \pi_{X \bigcap Y}(t)$. Отсюда и из доказанного п. 2 вытекают равенства $\pi_X(t) = \pi_X(\pi_R(t)) = \pi_{X \bigcap R}(t)$. ▫□

Предложение 3 (свойства активного домена и насыщения). Пусть R – схема таблицы t. Выполняются следующие утверждения:

1) $D_{A,t} = \varnothing \Leftrightarrow t = t_\varnothing \vee A \notin R$ (критерий пустоты активного домена); $C(t) = t_\varnothing \Leftrightarrow t = t_\varnothing$ (критерий пустоты насыщения, сохранение пустой таблицы $t_\varnothing$ насыщением); $C(t) = t_\varepsilon \Leftrightarrow t = t_\varepsilon$ (сохранение специальной таблицы t_ε насыщением);

2) $t \subseteq C(t)$ (возрастание насыщения по теоретико-множественному включению). □

Доказательство. Проверим п. 1 и начнем с критерия пустоты активного домена. Для этого воспользуемся представлением активного домен (2.1), видом области определенности $dom\, I_A$ (2.2) и критерием пустоты полного образа ([11,

подраздел 1.3, предложение 1.3.1; 19]), имеем цепочку эквивалентностей: $D_{A,t} = \varnothing \Leftrightarrow I_A[t] = \varnothing \Leftrightarrow t \cap domI_A = \varnothing \Leftrightarrow t \cap \bigcup_{\widetilde{R}:A\in\widetilde{R}} S(\widetilde{R}) = \varnothing$.

Поскольку по предположению таблица t имеет схему R, то $t \subseteq S(R)$ и остается учесть следующую кусочную схему, корректность которой проверяется непосредственно:

$$t \cap \bigcup_{\widetilde{R}:A\in\widetilde{R}} S(\widetilde{R}) = \begin{cases} t, \text{если } A \in R, \\ \varnothing, \text{если } A \notin R. \end{cases}$$

Проверим критерий пустоты насыщения. Достаточность вытекает из представления насыщения [1,предложение 1, п. 3], критерия пустоты обобщенного прямого произведения [11, подраздел 1.3, предложение 1.3.5; 20] и доказанного критерия пустоты активного домена. ▫

Необходимость, т.е. импликацию $C(t) = t_\varnothing \Rightarrow t = t_\varnothing$, докажем от противного.

Итак, пусть $t \neq t_\varnothing$. Очевидно, что $t \neq t_\varepsilon$ (действительно, это проверяется от противного с учетом фактов $C(t_\varepsilon) = t_\varepsilon, t_\varepsilon \neq t_\varnothing$). Значит, схема R таблицы не пуста. Согласно представлению насыщения имеем равенство $C(t) = \prod_{A\in R} D_{A,t}$; поскольку таблица t непуста, то в силу критерия пустоты активного домена все множества $D_{A,t}$, $A \in R$, непусты. Но отсюда по критерию пустоты обобщенного прямого произведения вытекает, что и произведение $\prod_{A\in R} D_{A,t}$ непусто; пришли к противоречию. ▫

Остается проверить эквивалентность $C(t) = t_\varepsilon \Leftrightarrow t = t_\varepsilon$. Достаточность вытекает прямо из определения насыщения; установим необходимость. Пусть $C(t) = t_\varepsilon$, т.е. $C(t)$ имеет пустую схему. Поскольку по определению насыщения эта операция схему не меняет, то и таблица t имеет пустую схему. Но существуют всего две таблицы пустой схемы – $t_\varnothing$ и t_ε, причем случай $t = t_\varnothing$ невозможен (поскольку тогда бы выполнялось равенство $C(t) = t_\varnothing$ согласно доказанному критерию пустоты насыщения). ▫

Рассмотрим п. 2. Случай таблиц пустой схемы $t \in T(\varnothing)$ проверяется непосредственно; пусть далее $t \notin T(\varnothing)$, т.е. $R \neq \varnothing$. Пусть $s \in t$, тогда по свойствам обобщенного прямого произведения [11, подраздел 1.3, предложение 1.3.5; 20] имеем принадлежность $s \in \prod_{A\in R} \{s(A)\}$; поскольку для всех $A \in R$ выполняется включение $\{s(A)\} \subseteq D_{A,t}$, то вследствие монотонности оператора $\prod$ ([11, подраздел 1.3, предложение 1.3.5; 20]) имеем принадлежность $s \in \prod_{A\in R} D_{A,t} = C(t)$. ▫□

2.2. Полугруппа и нижняя полурешетка таблиц по соединению

Ниже приведены основные свойства бинарной операции соединения.

Предложение 4 (свойства соединения). Пусть таблицы t_1, t_2 имеют схемы R_1, R_2 соответственно. Тогда выполняются следующие соотношения:

1) $t_1 \otimes t_2 \in T(R_1 \bigcup R_2)$;
2) $t_1 \otimes t_2 = t_2 \otimes t_1$ (коммутативность соединения);
3) $t_1 \otimes (t_2 \otimes t_3) = (t_1 \otimes t_2) \otimes t_3$ (ассоциативность соединения);
4) $t \otimes t_\varnothing = t_\varnothing \otimes t = t_\varnothing$ (сохранение пустой таблицы $t_\varnothing$);
5) $t \otimes t_\varepsilon = t_\varepsilon \otimes t = t$ (специальная таблица t_ε – правая и левая единица);
6) $R_2 \subseteq R_1 \Rightarrow t_1 \otimes t_2 \subseteq t_1$;
7) $t \otimes t = t$ (идемпотентность соединения). □

Доказательство. Принадлежность п. 1 вытекает непосредственно из определения соединения. Содержательная интерпретация следующая: схема соединения таблиц получается теоретико-множественным объединением схем таблиц-аргументов. ▫

Пп. 2, 3 вытекают из коммутативности и ассоциативности операции $\bar{\bigcup}$ [11, подраздел 2.7, лемма 2.7.2], утверждении о наследовании коммутативности и ассоциативности операции при распространении с элементов на множества элементов (с помощью полного образа) [19], а также из представления соединения [1,предложение 1, п. 2].▫

Доказательство п. 4. вытекает из представления операции соединения и сохранения пустого множества декартовым произведением и полным образом [11, подраздел 1.3, предложение 1.3.1; 19]. ▫

Равенства п. 5 вытекают из определения операции соединения и свойств отношения совместности ([11, подраздел 2.7, лемма 2.7.2]; напомним, что специальная строка ε представляет собой пустое множество): $t \otimes t_\varepsilon = \bar{\bigcup}[t \times \{\varepsilon\}] = \bar{\bigcup}[\{< s, \varepsilon > | s \in t\}] = \{s \bigcup \varepsilon | s \in t\} = \{s | s \in t\} = t$.▫

Рассмотрим п. 6. Пусть $s \in t_1 \otimes t_2$, тогда $s = s_1 \bigcup s_2$ для подходящих строк $s_1 \in t_1$, $s_2 \in t_2$, причем $s_1 \approx s_2$. Так как $R_2 \subseteq R_1$, то по свойствам ограничения [11, подраздел 1.3, предложение 1.3.2; 19] имеем $s_2 = s_2 | R_2 \subseteq s_2 | R_1$. Согласно свойств отношения совместности [11, подраздел 2.7, лемма 2.7.1] выполняется неравенство $s_2 | R_1 \subseteq s_1$. Отсюда вытекает включение $s_2 \subseteq s_1$, т.е. $s = s_1 \bigcup s_2 = s_1 \in t_1$. Таким образом, $t_1 \otimes t_2 \subseteq t_1$. ▫

Переходим к последнему п. 7. Из предыдущего пункта вытекает включение $t \otimes t \subseteq t$, поэтому остается проверить обратное включение. Пусть $s \in t$, тогда $s = s \bigcup s$ (идемпотентность объединения функций), причем $s \approx s$ (рефлексивность отношения совместности). Отсюда по определению соединения вытекает включение $s \in t \otimes t$ ▫□

Для введения частичных порядков на носителе табличных алгебр установим дискретность таблиц относительно теоретико-множественного включения на

множестве строк, т.е. дискретность частично упорядоченного множества (ч. у. м.) вида $\langle t, \subseteq \rangle$; дискретность понимается стандартно (см., например, [24]: $\forall s_1 \forall s_2 (s_1, s_2 \in t \Rightarrow (s_1 \subseteq s_2 \Rightarrow s_1 = s_2))$. Этот факт непосредственно вытекает из свойств ограничения [11, подраздел 1.3, предложение 1.3.2; 19], здесь существенно, что строки одной таблицы имеют одинаковые схемы.

Введем на множестве всех таблиц T бинарное отношение $\lhd$:

$$t_1 \lhd t_2 \overset{def}{\Leftrightarrow} \forall s_1 (s_1 \in t_1 \Rightarrow \exists s_2 (s_2 \in t_2 \,\&\, s_2 \subseteq s_1)). \qquad (2.3)$$

Это отношение допускает содержательную интерпретацию: если $t_1 \lhd t_2$, то информация, представленная в таблице t_1, продолжает часть информации, представленной в таблице t_2.

Предложение 5 (устройство множества таблиц). $< T, \lhd >$ является частично упорядоченным множеством с наименьшим элементом $t_\varnothing$ и наибольшим элементом t_ε. □

Доказательство. Очевидно, что бинарное отношение $\lhd$, заданное в (2.3), является отношение коинициальности, индуцированным теоретико-множественным отношением включения строк. Вследствие дискретности таблиц и упорядочения семейства дискретных множеств отношением коинициальности [22], получаем, что $< T, \lhd >$ является ч. у. м. с наименьшим элементом $t_\varnothing$. ▫

То, что t_ε – наибольший элемент, вытекает из определения отношения $\lhd$ и того очевидного факта, что пустая строка ε – наименьший элемент ч. у. м. $< S, \subseteq >$. С содержательной точки зрения, особая таблица t_ε вообще не содержит информации, поэтому и является наибольшим элементом (отметим, что пустая таблица тоже не содержит информации).▫ □

Далее через $\prod$ обозначается точная нижняя грань (инфимум) множества.

Лемма 1 (связь между частичным порядком $\lhd$ и соединением). Выполняются следующие утверждения:

1) $t_1 \lhd t_1' \wedge t_2 \lhd t_2' \Rightarrow t_1 \otimes t_2 \lhd t_1' \otimes t_2'$ (монотонность соединения относительно порядка $\lhd$ по совокупности аргументов);

2) $t_1 \otimes t_2 = \prod\limits_{\lhd} \{t_1, t_2\}$ (представление соединения через порядок $\lhd$);

3) $t_1 \lhd t_2 \Leftrightarrow t_1 = t_1 \otimes t_2$ (представление порядка $\lhd$ через соединение). □

Доказательство. Начнем с п. 1. Пусть $t_1 \lhd t_1'$ и $t_2 \lhd t_2'$, покажем, что тогда $t_1 \otimes t_2 \lhd t_1' \otimes t_2'$. Случай $t_1 \otimes t_2 = t_\varnothing$ тривиальный, поскольку пустая таблица $t_\varnothing$ есть наименьший элемент ч. у. м. $< T, \lhd >$.

Поэтому пусть далее $t_1 \otimes t_2 \neq t_\varnothing$, рассмотрим произвольную строку $s \in t_1 \otimes t_2$. По определению соединения существуют строки $s_1 \in t_1$, $s_2 \in t_2$, такие, что $s = s_1 \bigcup s_2$, причем $s_1 \approx s_2$. Так как $t_1 \lhd t_1'$ и $t_2 \lhd t_2'$ по предположению, то существуют строки $s_1' \in t_1'$ и $s_2' \in t_2'$, такие, что $s_1' \subseteq s_1$, $s_2' \subseteq s_2$. Отсюда вытекает, что $s_1' \bigcup s_2' \subseteq s_1 \bigcup s_2$, т.е. $s_1' \bigcup s_2'$ является функцией (как подмножество

функции); по [11, подраздел 2.7, лемма 2.7.1] отсюда следует, что $s'_1 \approx s'_2$. Таким образом, по определению соединения имеем $s'_1 \bigcup s'_2 \in t'_1 \otimes t'_2$. Остается учесть уже указанное включение $s'_1 \bigcup s'_2 \subseteq s_1 \bigcup s_2$.

Итак, для произвольной строки $s \in t_1 \otimes t_2$ нашлась строка из соединения $t'_1 \otimes t'_2$ (а именно $s'_1 \bigcup s'_2$), меньшая s (по теоретико-множественному включению). Поэтому по определению отношения коинициальности имеем требуемое неравенство $t_1 \otimes t_2 \lhd t'_1 \otimes t'_2$. ▫

Рассмотрим п. 2. Из определения отношения коинициальности вытекает, что $t_1 \otimes t_2 \lhd t_1$ и $t_1 \otimes t_2 \lhd t_2$; таким образом, $t_1 \otimes t_2 \in \{t_1, t_2\}^{\nabla}$ (в левой части последней принадлежности записан нижний конус, следуя обозначениям [24]) и остается показать, что $t_1 \otimes t_2$ является наибольшим элементом указанного нижнего конуса. Действительно, пусть $t \in \{t_1, t_2\}^{\nabla}$, т.е. $t \lhd t_1$ и $t \lhd t_2$. Из предыдущего доказанного пункта (о монотонности соединения) вытекает неравенство $t \otimes t \lhd t_1 \otimes t_2$; остается учесть идемпотентность соединения (предложение 4, п. 7). ▫

П. 3 следует непосредственно из предыдущего п. 2. ▫ □

Из предложения 4 вытекает, что группоид $< T, \otimes >$ является идемпотентной коммутативной полугруппой. Используя хорошо известную связь между идемпотентными коммутативными полугруппами и полурешетками (см., например, [24; 25]), полугруппу $< T, \otimes >$ можно преобразовать в, например, нижнюю полурешетку, вводя порядок стандартно:

$$t_1 \lhd_{\otimes} t_2 \overset{def}{\Leftrightarrow} t_1 = t_1 \otimes t_2 . \qquad (2.4)$$

Следствие 1. Отношение коинициальности на множестве таблиц T совпадает с порядком нижней полурешетки (коммутативной идемпотентной) полугруппы $< T, \otimes >$. □

Доказательство вытекает из определения (2.4) и п. 3 леммы 1. □

3. Обобщенная табличная алгебра

При построении таблиц в работе [1] (первая часть цикла) на них накладывались следующие ограничения: во-первых, строки имели конечные схемы («горизонтальная» конечность); во-вторых, таблицы были конечными множествами строк («вертикальная» конечность) и, в-третьих, строки одной таблицы имели одну схему.

Отметим тот факт, что указанные ограничения, как правило, не использовались в приведенных доказательствах (исключением является подраздел 2.2, где использовалась односхемной строк таблиц, что существенно для обеспечения дискретности таблиц по теоретико-множественному включению).

Снимая эти три ограничения, получим обобщенную табличную алгебру $< \widetilde{T}, \widetilde{\Omega}_{P,\Xi} >$, где $\widetilde{\Omega}_{P,\Xi} \overset{def}{=} \{\bigcup, \bigcap, \setminus, \sigma_p, \pi_X, \otimes, \div_{R_2}^{R_1}, Rt_{\xi}, \sim\}_{X,R_1,R_2 \subseteq \boldsymbol{A}}^{p \in P, \xi \in \Xi}$, (чтобы не

усложнять обозначения, за операциями обобщенной табличной алгебры сохраняем те же обозначения, что и для табличной алгебры).

В этой алгебре *схемой* называется произвольное множество атрибутов (в частности, бесконечное) $R \subseteq \boldsymbol{A}$, а *таблицей* называется произвольное множество строк (в частности, схемы строк одной таблицы могут буть различными, а количество строк таблицы может быть бесконечным).

Все сигнатурные операции обобщенной табличной алгебры определяются по аналогии с одноименными операциями табличной алгебры.

Самый сложный случай – это случай операции (обобщенного) соединения (обобщенных) таблиц, но представление соединения, указанное в п. 2 предложения 1 из [1], позволяет корректно ввести соответствующее определение.

Что касается деления, то его можно ввести формально как производную операцию, основываясь на хорошо известном представлении (см., например, [11, подраздел 2.13, предложение 2.13.1, п. 10; 26]):

$$t_1 \div_{R_2}^{R_1} t_2 \overset{def}{=} \pi_{R'}(t_1) \setminus_{R'} \pi_{R'}(\pi_{R'}(t_1) \otimes t_2 \setminus_{R_1} t_1), \text{ где } R' = R_1 \setminus R_2 .$$

Правда остается открытым вопрос о области определенности деления; положим, что она равна $T(R_1) \times T(R_2)$.

Кроме того, для введения (обобщенного) переименования надо модифицировать определение множества таблиц ξ-допустимых схем T_ξ: это множество состоит из таблиц, схемы (вообще говоря, разные) всех строк которых должны быть ξ-допустимыми:

$$T_\xi \overset{def}{=} \{t \mid \forall s(s \in t \Rightarrow \xi[\pi_1^2(s)] \cap (\pi_1^2(s) \setminus dom\xi) = \varnothing)\} .$$

Рассматривая две введенные в работе алгебры таблиц, имеем очевидную связь между ними.

Теорема 1. Табличная алгебра является собственной подалгеброй обобщенной табличной алгебры. □

Доказательство следует непосредственно из построения указанных алгебр. □

Теорема 2. Все утверждения предложений 1-4 выполняются в обобщенной табличной алгебре. □

Доказательство полностью повторяет соответствующие доказательства для табличной алгебры (поскольку в доказательствах не оиспользовались требования конечности таблиц и односхемности строк одной таблицы). □

4. Основные результаты

1. Показана монотонность операций проекции, соединения, насыщения, переименования относительно теоретико-множественного включения (предложение 1, пп. 1-4).

2. Показана дистрибутивность относительно объединений операций проекции, соединения, переименования (предложение 1, пп. 5-7); приведено естественное достаточное условие в терминах первичного ключа дистрибутивности проекции относительно пересечений (предложение 2, п. 5).

3. Установлено сохранение специальных таблиц $t_\varnothing, t_\varepsilon$ проекцией (предложение 2, пп. 3, 4); дана верхняя оценка проекции пересечений (предложение 2, п. 5) и нижняя оценка проекции разности (предложение 2, п. 6); рассмотрен случай композиции проекций, показана идемпотентность проекции (предложение 2, п. 7).

4. Установлено сохранение специальных таблиц $t_\varnothing, t_\varepsilon$ насыщением (предложение 3, п. 1); показано, что насыщение является возрастающим оператором по теоретико-множественному включению (предложение 3, п. 2).

5. Установлена коммутативность и ассоциативность соединения (предложение 4, пп. 2, 3); показано, что соединение сохраняет пустую таблицу $t_\varnothing$ (предложение 4, п. 4), а специальная таблица t_ε является единицей (правой и левой) по соединению; установлена идемпотентность соединения (предложение 4, п. 7).

6. На множестве всех таблиц по логической схеме отношения коинициальности введено бинарное отношение $\lhd$ и показано, что $<T, \lhd>$ является частично упорядоченным множеством с наименьшим элементом $t_\varnothing$ и наибольшим элементом t_ε (предложение 5); показано, что порядок $\lhd$ совпадает с порядком нижней полурешетки, соответствующей коммутативной идемпотентной полугруппе $< T, \otimes >$ (следствие 1).

7. Снятием ограничений финитности схем таблиц и самих таблиц, а также требования односхемности строк таблиц построена обобщенная табличная алгебра, которая содержит табличную алгебру в качестве подалгебры (теорема 1).

8. Показано, что все утверждения предложений 1-4 выполняются и для обобщенно табличной алгебры.

5. Заключение

В работе построен содержательный фрагмент теории табличных алгебр. Основная особенность используемой техники заключается, во-первых, в установлении естественных представлений сигнатурных операций этих алгебр в терминах теоретико-множественных конструкций полного образа, ограничения, обобщенного прямого соединения, коинициальности, отношения совместности (все это было сделано в работе [1], являющейся первой часть данного цикла) и, во-вторых, в переносе свойств указанных конструкций на табличный случай.

Это демонстрирует мощность используемого аппарата и, по мнению автора, указанный теоретико-множественный аппарат может успешно применяться и в других областях.

Применение указанных представлений дало возможность не только исследовать операции табличных алгебр, но и построить обобщенную табличную алгебру снятием требований конечности схем, таблиц и ограничения односхемности строк таблиц. Установленный результат позволит подойти к построению моделей данных для NoSQL баз данных [27-29].

Отметим, что вопрос о введении естественного порядка на носителе обобщенной табличной алгебры пока остается открытым.

Отметим также, что многие аспекты реальных табличных систем управления базами данных (например, наличие дубликатов строк в таблицах, агрегатные функции, внешние соединения) не нашли своего отражения в табличных алгебрах, поэтому сама модель требует расширения. Это будет предметом рассмотрения в следующих работах.

ЛИТЕРАТУРА

1. Кахута Н.Д. Математические основания современных реляционных систем управления базами данных. Часть 1: теоретико-множественные представления основных табличных операций // Вестник Харковского национального университета. Серия "Математическое моделирование. Информационные технологии. Автоматизированные системы управления".– 2014. – Вып. ??. – С. ??-??.
2. Брона Ю.Й. Основні співвідношення в табличних алгебрах // Вісник Київського університету. Сер.: фіз.-мат. науки. – 1997. – Вип. 3. – С. 143-148.
3. Брона Ю.Й. Оптимізація обчислення запитів у реляційних базах даних // Питання оптимізації обчислень: міжнародна конференція, 6-8 жовтня 1997 р., Київ, ІК ім. В.М. Глушкова НАНУ: праці. – 1997. – С. 45-49.
4. Буй Д.Б., Брона Ю.Й. Операторы замыкания в теории реляционных баз данных // Тезисы докладов XI Международной конференции по проблемам теоретической кибернетики. Под ред. С.В. Яблонского. – Ульяновск: Изд-во СВНЦ. – 1996. – С. 29-30.
5. Буй Д.Б., Брона Ю.Й. Теоретико-множинні конструкції в теорії реляційних баз даних // Вісник Київського університету. Сер.: фіз.-мат. науки. – 1996. – Вип. 1. – С. 216-224.
6. Буй Д.Б. Теорія програмних алгебр композиційного типу та її застосування: дисертація доктора фізико-математичних наук: 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем / Буй Дмитро Борисович. – Київ, 2002. – 365 с.
7. Редько В.Н., Брона Ю.Й., Буй Д.Б. Информационный аспект Case-технологий: основные соотношения в табличных алгебрах // Проблемы программирования. – 1997. – Вып. 1. – С. 5-11.
8. Редько В.Н., Буй Д.Б. К основаниям теории реляционных моделей баз данных // Кибернетика и системный анализ. – 1996. – № 4. – С. 3-12.
9. Редько В.Н. Реляционные алгебры: операции деления и переименования / В.Н. Редько, Ю.И. Брона, Д.Б. Буй // Кибернетика и системный анализ. – 1997. – № 5. – С. 3-15.
10. Редько В.Н., Брона Ю.Й., Буй Д.Б. Реляционные алгебры: операции проекции и соединения // Кибернетика и системный анализ. – 1997. – № 4. – С. 89-100.
11. Реляційні бази даних: табличні алгебри та SQL-подібні мови / В.Н. Редько, Ю.Й. Брона, Д.Б. Буй, С.А. Поляков. – Київ: Видавничий дім „Академперіодика”, 2001. – 198 с.
12. Codd E.F. A Relational Model of Data for Large Shared Data Banks // Communications of the ACM. – 1970. – Vol. 13, № 6. – P. 377-387.

13. Codd E.F. A Data Base Sublanguage Founded on the Relational Calculus // ACM-SIGFIDET Workshop on Data Description, Access and Control: international conference, November 11-12, 1971, San Diego, California: proceedings. – 1971. – P. 35-68.
14. Codd E.F. Normalized Data Base Structure: A Brief Tutorial // ACM-SIGFIDET Workshop on Data Description, Access and Control: international conference, November 11-12, 1971, San Diego, California: proceedings. – 1971. – P. 1-17.
15. Codd E.F. Further Normalization of Data Base Relational Model // Data Base Systems. – 1972. – P. 33-64.
16. Codd E.F. Relational Completeness of Data Base Sublanguages // Data Base Systems. – 1972. – P. 65-93.
17. Codd E.F. Relational Database: A Practical Foundation for Productivity // Communications of the ACM. – 1982. – Vol. 25, № 2. – P. 109-117.
18. Codd E.F. The Relational Model for Database Management [2-nd edition]. – Pearson:Addison-Wesley, 1990. – 538 p.
19. Буй Д.Б., Кахута Н.Д. Властивості теоретико-множинних конструкцій повного образу та обмеження // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2005. – Вип. 2. – С. 232-240.
20. Кахута Н.Д. Відношення сумісності, узагальнене з'єднання та узагальнений прямий добуток // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2007. – Вип. 4. – С. 167-173.
21. Кахута Н.Д. Критерії ін'єктивності бінарних відношень // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2007. – Вип. 3. – С. 141-146.
22. Буй Д.Б., Кахута Н.Д. Властивості відношення конфінальності та устрій множини часткових функцій // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2006. – Вип. 2. – С. 125-135.
23. Elmasri R., Navathe S. Fundamentals of database systems: [4-th edition]. – Pearson: Addison-Wesley, 2004. – 1030 p.
24. Скорняков Л.А. Элементы теории структур. – Москва: Наука, 1982. – 158 с.
25. Мальцев А.И. Алгебраические системы. – Москва: Наука, 1970. – 392 с.
26. Ульман Дж. Основы систем баз данных. – М.: Фин. и стат., 1983. – 334 с.
27. Strozzi C. NoSQL: a relational database management system. – [Электронный ресурс]. – Режим доступа: http://www.strozzi.it/cgi-bin/CSA/tw7/I/en.
28. Pokorny J. NoSQL databases: a step to database scalability in web environment [Текст] // Proc. of the 13th International Conference on Information Integration and Web-Based Applications and Services. – 2011. – P. 278-283.
29. Strauch C. NoSQL databases. – [Электронный ресурс]. – Режим доступа: http://www.christof-strauch/de/nosqldbs.pdf.

Надійшла у першій редакції 09.07.2014, в останній - 17.07.2014.

UDC 534, 517.928

Analytical-numerical approach to analyze forced and parametric vibrations of some pendulum systems

A. A. Klimenko, Yu. V. Mikhlin
National Technical University «Kharkov Polytechnic Institute»

The parametric oscillations of physical pendulum and forced vibrations of a system with pendulum absorber are analyzed using the approach based on combined application of the concept of nonlinear normal vibration modes, the Rauscher method, and numerical procedures. Frequency responses are obtained.

Key words: *pendulum systems, nonlinear normal vibrationsя, Rauscher method.*

Проведен анализ параметрических колебаний физического маятника и вынужденных колебаний системы с маятниковым гасителем колебаний с применением единого подхода, базирующегося на совместном использовании метода Раушера, метода нелинейных нормальных форм колебаний и численных процедур. Построены амплитудно-частотные характеристики.

Ключевые слова: *маятниковые системы, нелинейные нормальные колебания, метод Раушера.*

Проведено аналіз параметричних коливань фізичного маятника та вимушених коливань системи з маятниковим гасителем коливань із застосуванням єдиного підходу, що базується на спільному використанні методу Раушера, методу нелінійних нормальних форм коливань та чисельних процедур. Побудовано амплітудно-частотні характеристики.

Ключові слова: *маятникові системи, нелінійні нормальні коливання, метод Раушера.*

1. Introduction

It is well-known that resonance forced vibrations of a single-degree-of freedom (DOF) nonlinear systems under small periodic perturbations in the region of main resonance are close to natural vibrations of unperturbed conservative system. This result can be transferred to finite-DOF systems. In the last case the resonance forced vibrations are close to nonlinear normal vibration modes of corresponding finite-DOF conservative systems. Thus, it is appropriate to use the nonlinear normal vibration modes of conservative systems to construct forced resonance vibrations. It permits to consider vibrations with essential amplitudes.

Origins of the nonlinear normal vibrations theory can be found in works by Lyapunov [1] on systems with the first analytical integral. Concept of nonlinear normal vibration modes (NNMs), which is based on construction of trajectories in the system configuration space, is developed in works by Kauderer [2] and Rosenberg [3]. Approach of construction of curvilinear trajectories of NNMs is proposed in publications [4,5]. Principal aspects of the NNMs theory by Kauderer-Rosenberg are presented in books [5,6] and in review [7]. Approach which combines the concept of nonlinear normal vibration and the Rauscher method is used to construct forced resonance vibrations of systems having few degrees of freedom in [8,5,6]. (Initially the Rauscher method was proposed for a nonlinear conservative single-DOF system [9]) Note that the same approach can be used also in construction of parametric vibrations.

Here the approach based on combined use of the concept of NNMs, the Rauscher method and numerical procedures, is used in problem of parametric vibrations of the spring pendulum and in problem of forced resonance vibrations of the system containing a pendulum absorber.

2. Parametric vibrations of a spring pendulum

The model of two-DOF spring pendulum is presented in Fig. 1. Vibrations of the mass m on the linear spring of the length l in unstressed state are considered. Dynamics of the system is described by two generalized coordinates ρ and φ.

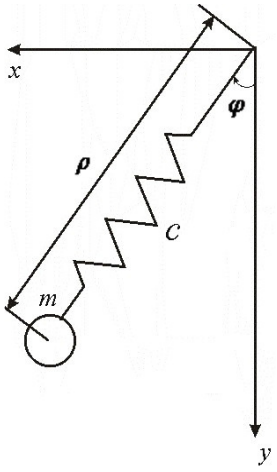


Fig. 1. Spring pendulum

Let the small periodic perturbation applied in vertical direction is considered. Equations of parametric vibrations of the system can be wrote as the following:

$$\begin{cases} \ddot{z}+z=\mu\left(\rho_0\dot{\varphi}^2+\dfrac{(g+\mu F\cos\Omega t)}{2}\varphi^2\right)+\mu^2\dot{\varphi}^2 z; \\ \rho_0^2\ddot{\varphi}+(g+\mu F\cos\Omega t)\rho_0\varphi=\mu\left(-2\rho_0\ddot{\varphi}-2\rho_0\dot{\varphi}\dot{z}-(g+\mu F\cos\Omega t)z\varphi\right)+ \\ +\mu^2\left(-z^2\ddot{\varphi}-2z\dot{z}\dot{\varphi}+\dfrac{(g+\mu F\cos\Omega t)\rho_0}{6}\varphi^3\right)+\mu^3\dfrac{(g+\mu F\cos\Omega t)}{6}z\varphi^3, \end{cases} \quad (1)$$

where $\mu z=\rho-\rho_0$, $\rho_0=l+\dfrac{gm}{c}$ is an extension of the spring in the equilibrium state; μ is a formal small parameter.

Combination of the NNMs approach and the Rauscher method is used to construct normal modes of parametric vibrations. One considers, first of all, the autonomous system obtained from equations (1) in the zero approximation by the small parameter ($\mu=0$). In regime of nonlinear normal mode one has $z=z(\varphi)$; and the system under consideration is reduced to the single-DOF system with respect to the variable φ. The nonlinear normal mode can be obtained by power series [4,5]:

$$z = z(\varphi) = z_0 + \mu z_1 + ..., \tag{2}$$

$$\begin{aligned} z_0 &= a_0 + a_1\varphi + a_2\varphi^2 + a_3\varphi^3 + a_4\varphi^4 + ...; \\ z_1 &= b_0 + b_1\varphi + b_2\varphi^2 + b_3\varphi^3 + b_4\varphi^4 + ... \end{aligned} \tag{3}$$

One uses a representation of the generalized coordinate φ in the regime of the NNM as Fourier series, namely, $\varphi = A_0 + A_2\cos(2\Omega t) + A_4\cos(4\Omega t) + ...$ We introduce the Fourier series to the obtained previously single-DOF system, saving only three first terms of the series and then using the harmonic balance. One obtains, as a result, a system of three nonlinear algebraic equations with respect to four unknowns (A_0, A_2, A_4, Ω). During the next numerical calculations a value A_2 is given with some step. The system of nonlinear algebraic equations is solved with respect to unknown quantities (A_0, A_2, A_4, Ω) for each value of A_2. One has from here the required parameters (A_0, A_4, Ω). As a result, the first coefficient of the Fourier series will be determined.

One transforms now the Fourier series using known trigonometric formulae, as

$$\begin{aligned} &\varphi = A_0 + A_2\cos(2\Omega t) + A_4\cos(4\Omega t) + ... = \\ &= A_0 + A_2\left(2\cos^2(\Omega t) - 1\right) + A_4\left(8\cos^4(\Omega t) - 8\cos^2(\Omega t) + 1\right) + ... = \\ &= \left(A_0 - A_2 + A_4\right) + \left(2A_2 - 8A_4\right)\cos^2(\Omega t) + 8A_4\cos^4(\Omega t) + ... \end{aligned} \tag{4}$$

The following expansion can be obtained from the relation (4):

$$\varphi = \left(A_0 - A_2 + A_4\right) + \left(2A_2 - 8A_4\right)\cos^2(\Omega t) + 8A_4\cos^4(\Omega t) + ... \tag{5}$$

Then, some algebraic transformations permit to invert the expansion (5) and to obtain the following relation:

$$\cos\Omega t = \alpha_0 + \alpha_1\varphi + \alpha_2\varphi^2 + ... \tag{6}$$

On has the external periodic excitation is presented as a function of the generalized coordinate φ in zero approximation by the small parameter. Introducing the expansion (6) to the initial non-autonomous dynamical system (1), one obtains so-called «pseudo-autonomous» dynamical system. Such adduction of the non-autonomous system to the autonomous one corresponds to main idea of the Rauscher method. In obtained autonomous dynamical system the nonlinear normal vibration mode can be anew obtained. It permits to make more precise expansions (2) - (3), and to realize again a transfer from the initial non-autonomous system to the «pseudo-autonomous» one. Thus, the iteration *analytical-numerical procedure* can be used for a construction of forced resonance vibrations, which permits to obtain a solution with good exactness.

Two nonlinear normal vibration modes can be selected in this system: a) localized vibration mode close to longitudinal vibration mode of the pendulum system without the external excitation; in the localized mode amplitudes of rotations are small;
b) coupled vibration mode, when amplitudes of longitudinal vibrations and rotations are comparable.

Trajectories of the localized vibration mode of parametric vibrations in the system configuration space are shown in Fig. 4a; trajectories of the mode of coupled vibrations are shown in Fig. 4b. Calculations are made for the next values of the system parameters: $g = 9.8$, $l = 0.5$, $m = 1$, $c = 3$, $\mu = 0.1$, $f = 0.3$, $\varphi_0 = 0.01$ (for the localized mode) and $g = 9.8$, $l = 0.5$, $m = 0.1$, $c = 2$, $\mu = 0.1$, $f = 3$, $\varphi_0 = 0.1$ (for the mode pf coupled vibrations). Here red lines correspond to the analytical solution, and blue lines correspond to checking numerical simulation by the Runge-Kutta method, which is made for initial solutions obtained from analytical solution. Numerical calculations confirm good exactness of the analytical results.

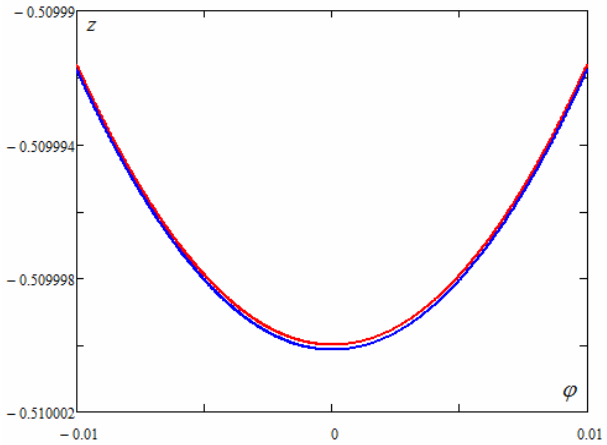


Fig. 2.a

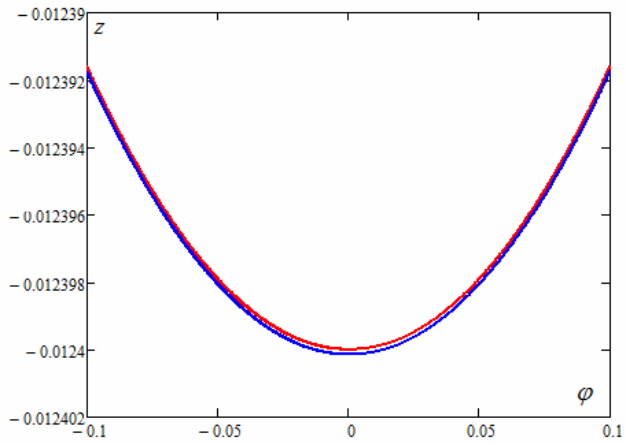


Fig.2.b

Fig. 2. Trajectories of the localized vibration mode of parametric vibrations (Flg.2a) and of the mode of coupled vibrations (Fig.2b). Comparison of analytical and checking numerical solutions.

3. Forced vibrations of the system, which contains a pendulum absorber

The second model under consideration is presented in Fig. 3. Here mechanical subsystem which vibrations must be extinguished is presented as oscillator of the mass m_1 with anchor spring which rigidity coefficient is equal to k. The pendulum absorber of the mass m_2 and of the length l is attached to the linear oscillator.

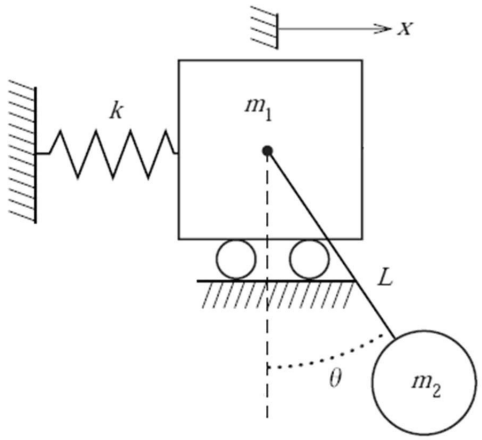


Fig. 3. System containing the pendulum absorber

Motions of the system are described by two generalized coordinates x (displacement of the linear subsystem) and θ (angle of the pendulum absorber).

Equations of motion of the model in the presence of the small external periodic action are the following:

$$\begin{cases} \left(m_1+\varepsilon m_2\right)\ddot{x}+\varepsilon m_2 l\ddot{\theta}\left(1-\dfrac{\theta^2}{2}\right)-\varepsilon m_2 l\dot{\theta}^2\left(\theta-\dfrac{\theta^3}{6}\right)+kx=\varepsilon F\cos\left(\Omega t\right); \\ \ddot{x}\left(1-\dfrac{\theta^2}{2}\right)+l\ddot{\theta}+g\left(\theta-\dfrac{\theta^3}{6}\right)=0. \end{cases} \tag{7}$$

Here ε is the formal small parameter.

Two nonlinear normal vibration modes can be selected in this system: a) localized vibration mode when vibration amplitudes of the linear subsystem are essentially smaller than amplitudes of the pendulum; b) coupled vibration mode, when amplitudes of the linear oscillator and of the pendulum are comparable. The first vibration mode is appropriate for absorption of vibrations of the linear subsystem.

Trajectories of the localized mode of forced vibrations (Fig.4a) and of the mode of coupled vibrations (Fig. 4b) are constructed using approach described in the preceding Section. This approach joints the nonlinear normal mode concept and the Rauscher method. Calculations are made for the following parameters of the system: $m_1=1$, $m_2=0.1$, $l=1$, $k=5$, $\varepsilon=0.1$, $f=0.1$. Here red lines correspond to the analytical solution, and the blue lines correspond to checking numerical simulation by the

Runge-Kutta method, which is made for initial solutions obtained from analytical solution. Numerical calculations confirm good exactness of the analytical results.

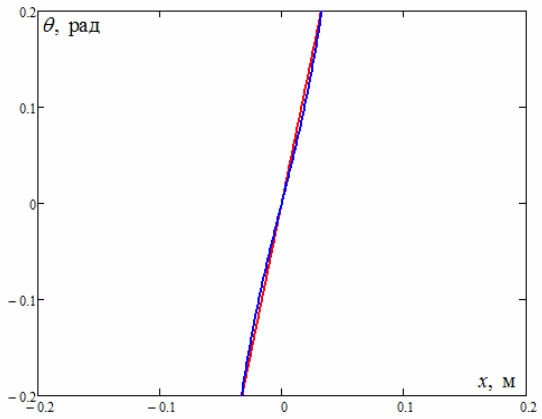


Fig. 4a.

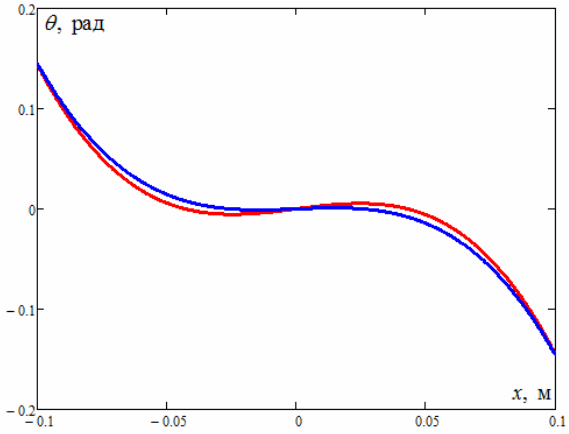


Fig. 4b.

Fig. 4. Trajectories of the localized mode of forced vibrations (Fig.4a) and of the mode of coupled vibrations (Fig.4b) for the system having the pendulum absorber.

Construction of frequency responses is made by the harmonic balance method. In correspondence with this method the variables x and θ are presented in the form of the following sum of harmonics: $x = A_1 \cos(\Omega t) + A_2 \sin(\Omega t)$; $\theta = B_1 \cos(\Omega t) + B_2 \sin(\Omega t)$. The frequency responses are shown in Fig. 5 for the vibration mode of coupled vibrations, and in Fig. 6 for the localized vibration mode. It can see that in regime of the localized vibration mode vibrations of the linear subsystem are essentially smaller than ones of the pendulum absorber; the vibration energy concentrates in the absorber. So, this regime is appropriate for a quenching of vibrations of the linear subsystem.

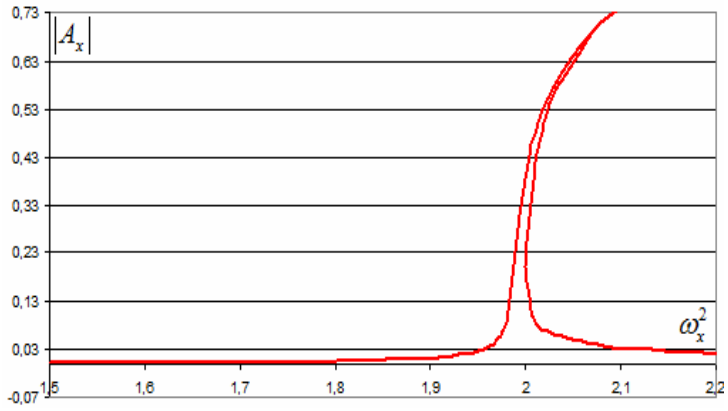


Fig. 5a.

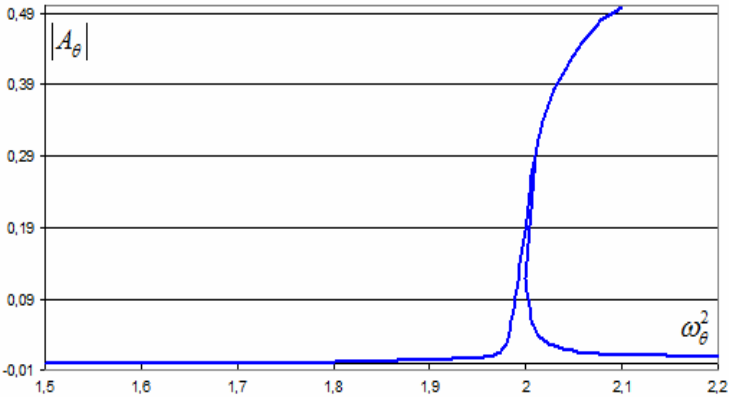


Fig. 5b.

Fig. 5. Frequency responses of the mode of coupled vibrations for the linear subsystem (Fig. 5a) and for the pendulum absorber (Fig. 5b)

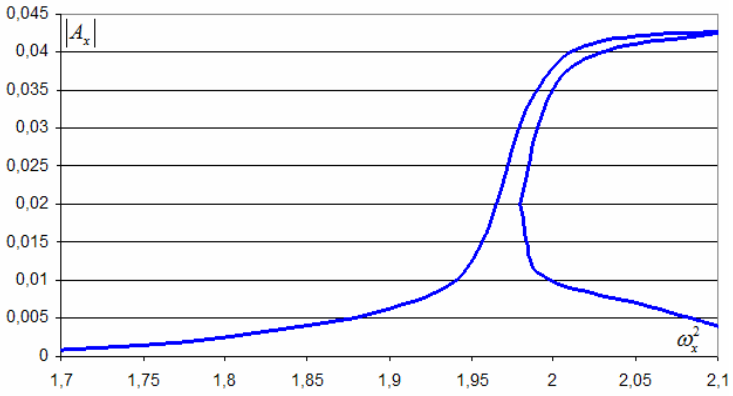


Fig. 6a

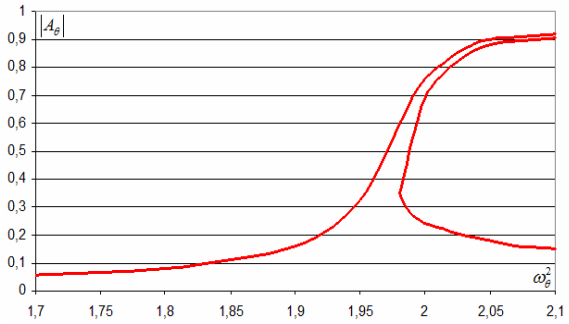


Fig. 6b

Fig. 6. Frequency responses of the localized vibration mode for the linear subsystem (Fig. 6a) and for the pendulum absorber (Fig. 6b)

4. Conclusions

New analytical-numerical approach to analyze parametric and forced vibrations of some pendulum systems is proposed. This approach is based on concept of nonlinear normal vibration modes, the Rauscher method and some numerical procedure. The proposed approach permits to construct trajectories of parametric and forced vibration modes in the system configuration space. Frequency responses are constructed too.

REFERENCES

1. Lyapunov A. M. The general problem of the stability of motion. Princeton University Press, Princeton. – 1947.
2. Kauderer H. Nichtlineare Mechanik. Springer-Verlag, Berlin. – 1958.
3. Rosenberg R. Nonlinear vibrations of systems with many degrees of freedom // Advances of Applied Mechanics. – 1966. - Vol.9. - pp. 156–243.
4. Manevich L., Mikhlin Yu. Periodic solutions close to rectilinear normal vibration modes // Prikladnaja Matematica i Mekhanica (PMM USSR). – 1972. – v. 36. – pp. 1051–1058 (in Russian).
5. Manevich L., Mikhlin Yu., Pilipchuk V. The method of normal oscillation for essentially nonlinear systems. Nauka, Moscow. – 1989 (in Russian).
6. Avramov K., Mikhlin Yu. Nonlinear Dynamics of Elastic Systems // Models, Methods and Approaches. Scientific Centre “Regular and Chaotic Dynamics”. - Moscow-Izhevsk, 2010. – 704 p. (in Russian).
7. Mikhlin Yu., Avramov K. Nonlinear normal modes for vibrating mechanical systems // Review of theoretical developments. Applied Mechanics Review. – 2010. – v. 63. - pp. 4-20.
8. Rauscher M. Steady oscillations of system with nonlinear and unsymmetrical elasticity // Journal of Applied Mechanics. - 1938. - v. 5.
9. Yu. Mikhlin. Resonance modes of near-conservative nonlinear systems // Prikladnaja Matematica I Mekhanica (PMM USSR). – 1974. – v. 38. – pp. 425–429 (in Russian).

Надійшла у першій редакції – 23.09.2014.

УДК 004.412:004.052

Метрики трудности в оценке надёжности инструментальных библиотек и фреймворков

В. О. Мищенко

Харьковский национальный университет имени В. Н. Каразина, Украина

К надёжности инструментальных средств программирования предъявляются повышенные требования. При этом всегда актуально развитие экономичных методов оценки надёжности, применимых на всех этапах разработки библиотек и фреймворков. В данной работе развивается статический подход к оценке трудности понимания компилируемых модулей программ. Трудность или легкость такого понимания является важным объективным фактором успешности тестирования и модификации модулей. Метод статьи построен на идеях М. Холстеда, энергетическом анализе программ и приёме фильтрации модулей по трудности, связанном с известными проектами. Наш метод испытан применительно к оценке зрелости модулей системы Matreshka (из арсенала Ада индустрии), которая для контроля сравнивается с библиотекой OAE.

Ключевые слова: *надёжность программного обеспечения, статика, метрика, М. Холстед, трудность, энергетический анализ, инструментальная библиотека, фреймворк.*

До надійності інструментальних засобів програмування висуваються підвищені вимоги. При цьому завжди актуальним є розвиток економічних методів оцінки надійності, здатних до застосування на всіх етапах розробки бібліотек і фреймворків. У даній роботі розвинуто статичний підхід до оцінки труднощів розуміння модулів програм, що компілюються. Труднощі або легкість такого розуміння є важливим об'єктивним фактором успішності тестування та модифікації модулів. Метод статті побудовано на ідеях М. Холстеда, енергетичному аналізі програм і прийомі фільтрації модулів за трудністю, що пов'язаний з відомими проектами. Наш метод випробувано стосовно оцінки зрілості модулів системи Matreshka (з арсеналу Ада індустрії), яка порівнюється з бібліотекою OAE.

Ключові слова: *надійність програмного забезпечення, статичні методи, метрика, М. Холстед, тудність, енергетичний аналіз, інструментальна бібліотека, фреймворк.*

The reliability of software must meet raised requirements. Development of reliability assessment methods, which are economic and applicable at all stages of libraries and frameworks creation, stays always actual. In this paper, we develop the static approach to estimation of difficulties in understanding of compilation units. Is this unit clear or difficult to understand, that is an important objective factor affecting the success in the unit testing and modification. This article method is built on the M. Halstead's ideas, energy analysis of program and technique of units filtration by their difficulty, that is associated with well-known projects. Our method have been tested with respect to the assessment of maturity of Matreshka software units (from the Ada industry toolkit), which is compared with the OAE framework.

Key words: *software reliability, static methods, metrics, M. Halstead, difficulty, energy analysis, software toolkit, framework.*

1 Введение

Статические метрики качества программ, которые здесь обсуждаются, используют хорошо известную систему примитивных характеристик (примитивов), которую, некогда изобрёл М. Холстед. Позднее часть этой системы послужила исходным пунктом в развитии энергетического анализа

программ [1]. Первыми из числа холстедовских примитивов должны быть упомянуты словарь программного модуля η, т.е. число разных программных символов, из которых составлен исходный текст модуля, и длина программы N – общее число программных символов в модуле. Ввести определение программных символов для конкретного формального языка не сложно, но ради сравнимости оценок это целесообразно делать по аналогии примерами М. Холстеда. В [2] перечислены (и аргументированы) принципы, позволяющие это сделать. Т.н. научные метрики Холстеда [3] используют также дополнительную дифференциацию программных символов на операторы и операнды:

η_1 и N_1 – словарь и общее число операторов;

η_2 где N_2– словарь и общее число операндов ($\eta_2 = \eta - \eta_1$, $N_2 = N - N_1$).

В современных языках программирования способы различения операторов и операндов не очевидны в силу контекстных зависимостей. Более того, работа [4] показала, что такие способы по существу субъективны и, следовательно, не могут быть формализованы однозначно (например, одно и то же имя операции f при своём описании играет роль операнда, а в выражениях – оператора, и неясно, какую точку зрения предпочесть). Показательно, что несмотря на это, общий словарь и общее число программных символов практически не зависят от выбора способа различения операторов и операндов [5]. С проблемой неопределённости оценки примитивов η_1 и N_1 отчасти сталкивался и сам Холстед. Он же предложил простейший выход – для модулей, относимых к какой-то общей категории, при отсутствии точной информации полагать

$$\eta_1 : \eta_2 \approx N_1 : N_2 = \theta , \tag{1}$$

где θ - среднее по предварительно оценённой выборке или, ещё грубее, = 1/2.

В канонизированном подмножестве холстедовских метрик [3] роль операторов и операндов де факто сводится, к определению «метрики трудности»

$$\hat{D} = \frac{\eta_1}{2} \cdot \frac{N_2}{\eta_2} , \tag{2}$$

значения которой фиксируются на шкале безразмерных вещественных чисел, не меньших 1 (считается, что число η_1 разных символов-операторов в программе не может быть меньше 2, а $N_2 \geq \eta_2$ в силу определения). Точное название метрики (2) в полной системе холстедовских метрик это *оценка трудности*, поскольку Холстед дал и другой подход к характеризации трудности модуля, полагая, что «приемлема любая из двух интерпретаций» [6, С. 39]. Этот другой подход определяет *трудность D*, используя понятие потенциального объёма, введенного в [6] без точного определения. При этом подходе (2) является оценкой *D*. Большинство прикладников от использования понятия потенциального объёма отказались, и метрикой трудности *D* считают (2).

На сегодня, энергетический анализ программ строго реализует понятие потенциального объёма и определяет метрику трудности альтернативным образом по отношению к (2). Такой подход более адекватен применительно к современным программным системам, и мы его принимаем. Пусть рассматривается модуль программы или, если необходимо, его внутренний блок

(понятие, обобщающее понятия функций, процедур языков программирования [2,1]). Следуя идее Холстеда, введём потенциальный объём формулой:

$$V^* = \eta^* \log_2 \eta^* \ . \tag{3}$$

где η_2 - число формальних параметров модуля (блока). Для определения этой величины мы используем новые примитивы, допускающие точную интерпретацию в языках программирования:

$$\eta_2^* = p_1 + p_2 + j \ , \tag{4}$$

где

p_1 - число параметров шаблона или типа (если блок это шаблон или тип);

p_2 - число явных формальных параметров блока по его интерфейсу;

j - число квазиобьектов ввода-вывода в теле данного блока, полагая

$$j = \sqrt{j_1 j_2} \ , \tag{5}$$

где j_1 - число файлов, с которыми работает блок;

j_2 - число операторов ввода-вывода в эти файлы в исходном тексте тела блока.

Модули сложной внутренней архитектуры представляются [1] в виде объединения непересекающихся блоков, а их потенциальные объёмы полагаются тогда суммой потенциальных объёмов всех блоков представления.

Теперь становится корректной формула метрики трудности любого модуля

$$D = \frac{V}{V^*} \ , \tag{6}$$

где V^* потенциальный (3), а V – действительный объём данного модуля (M), который выражается через примитивы длины и словаря:

$$V = N \log_2 \eta = V(M) = V(M; N, \eta) \ . \tag{7}$$

Отметим ради истории, что в известной книге [6] фигурируют обратные к (2) и (1) величины, называемые «уровнем программы» и её «аппроксимацией».

По поводу интерпретации наблюдённых значений $\hat{D}$ документ [3, С.84] лаконично констатирует, что «…для средней программы PL/I, трудность должна быть не менее 115. Трудность больше, чем 160, указывает на серьезные проблемы. Данные о трудности доступны также и для других языков». Учитывая беглые замечания некоторых других авторов, публиковавшихся на рубеже 70-80 гг., эту информацию можно понимать как следующую рекомендацию: На основе статистических данных априори разделять модули большой программной системы с точки зрения риска ошибочной реализации спецификаций на 3 класса (один из которых содержит модули, не внушающие подозрений). При этом использовать 2 рубежа трудности:

$$D_1 = E\left(\hat{D}(M_1), .. \hat{D}(M_N)\right), \qquad D_2 = E + S \qquad , \tag{8}$$

где E – среднее значений трудности (2) по базовой выборке модулей M_i, которые разрабатывались в данной системе программирования на данном языке;

S – стандартное отклонение в той же выборке.

Более точная информация содержится в малодоступных технических отчётах фирмы IBM. Однако основные моменты находим в фундаментальном обзоре [7], фактически подбившем итог эпохи интенсивных исследований по направлениям холстедовской науки о программах. Из наиболее важного для нас, процитируем следующее место [7, С.160], относящееся к отчёту С. П. Смита из лаборатории указанной фирмы в Санта Терезе: «… последнее исследование предложило учредить пороговые значения $\hat{D}$, которые могут быть использованы программистами в качестве руководства для разработки программных продуктов. Например, для программ PL/S, было установлено, что среднее значение η_1 должно быть 46, а отношение N_2/η_2 – меньше 5 ... Два пороговых значения для сложности метрики могут быть определены следующим образом:

$$\hat{D}_1 = (46/2) \times 5 = 115 \ , \qquad (22)$$

$$\hat{D}_2 = (46 + 18/2) \times 5 = 160 \ . \qquad (23)$$

(Значение 18, использованное в (23), является стандартным отклонением η_1) Если для определенного модуля $115 < \hat{D} < 160$, то программисту рекомендуется пересмотреть свой код с точки зрения наличия неудачных приёмов программирования, таких как использование слишком большого числа операторов GO TO, неоправданного обилия встроенного кода на ассемблере или необоснованного использования дополнительных операндов. Если $\hat{D} > 160$, то рекомендуются более радикальные меры, такие как экспертиза в группе. Исследования IBM позволяют заключить, что подобные пороговые значения могут быть установлены для других языков программирования высокого уровня».

Подводя итог, подчеркнем вывод о том, что метрику оценки трудности можно, основываясь на опытных данных, использовать для предварительной классификации модулей программной системы по степени их понятности для исполнителей при тестировании, устранении дефектов или модификации системы. При этом в силу психофизической подоплёки дела является очевидным, что определение соответствующих порогов с высокой точностью, даже, если они хорошо выражены, не критично для успешности метода. Представляется, что опубликованные значения порогов для конкретной системы программирования могли бы служить ориентиром и непосредственно использоваться до накопления собственного опыта.

Те статические методы оценки качества программ, которые требуют использования алгоритмов только линейной или близкой к ней сложности (таковы, например, алгоритмы оценки метрик трудности!), всегда привлекательны в силу малости затрат при использовании. В частности, задачи оперативного сравнения больших систем, например, новых инструментальных или прикладных библиотек, с общих позиций сложности, рисков остаточных ошибок и т.п. могут решаться с помощью научных или энергетических метрик,

что подтверждается систематическим появлением публикаций на такие темы (из недавних [8-11]). Однако техника фильтрации, основанная на (2), с эпохи 80-х годов не обновлялась, а возможность использования в этом приёме метрики (6) вместо (2) является чистой гипотезой (хотя и правдоподобной).

Актуальность развития в современных условиях методов использования метрик трудности раздельно компилируемых модулей программ наглядно подтверждается следующим фактом. При модернизации документа [3] холстедовские метрики были в [12] исключены из перечня рекомендованных для стандартного использования инструментов контроля надёжности критических систем со следующим объяснением причин: «n_1, n_2, N_1, N_2 оказались полезными в Спейс-Шаттл в роли дискриминантов качества. Тем не менее, эти метрики трудно внедрить, потому что пороги, которые отличают низкое качество от высокого, должны быть оценены статистически. Этот процесс выходит за пределы навыков большинства практиков» [12, С. 24] (в цитируемом документе n_1, n_2 означают то, что выше обозначалось как η_1, η_2).

Таким образом, не ставя под сомнение полезность метрики $\hat{D}$, основанной на примитивах η_1, η_2, N_1, N_2 , авторы документа констатировали невозможность их механического использования сегодня с помощью прежней техники. Необходимы новые исследования для уточнения и восстановления доверия к данной метрике. Что же касается метрики D (6), то она в прикладном аспекте не исследована и, если упоминается, то в обзорном порядке.

Целью работы является создание и обоснование решающих правил, необходимых для реализации идеи фильтрации модулей программ на основе степени трудности, с испытанием на примере реальной программной системы .

2 Постановка задач данного исследования

Возможность применения метрики оценки трудности (2) к современным проектам мы рассмотрим на примере сравнения степени зрелости инструментальных библиотек Ада программирования [13,14]. При этом принимается во внимание, что уже в проектах с числом модулей порядка 10^3, и, тем более, в больших проектах, ни менеджеры качества, ни группа сопровождения обычно не располагают возможностью тщательного просмотра всего программного кода. Выборочные методы контроля не снимают этой экономической проблемы полностью, поскольку из соображений репрезентативности, выборки должны насчитывать, порядка 10^2 модулей и больше. Только после удачной классификации модулей можно сосредоточиться на тех из них, которые наиболее уязвимы с точки зрения рисков дефектности кода, трудности исправления выявленных ошибок или неудобства модификаций. Нашей первой задачей является *разработать современный вариант фильтрации модулей по классической метрике оценки трудности кода*. При этом в своих экспериментах с кодом, написанным на языке Ада, мы примем огрубляющее предположение (1), по-видимому, неизбежное для языков достаточно высокого уровня. Оно не лишает оценку (2) содержательности, а там, где возможны «точные» правила различения операторов и операндов выводы из решающих правил слабо чувствительны к таким огрублениям, как отмечал еще Холстед [6].

Вторая задача могла бы состоять в том, чтобы предложить аналогичную классификацию с использованием метрики трудности (6). Однако в результате корректного определения потенциального объёма, которое удалось достичь в энергетическом анализе [1,2], отношение (6) теряет однородность. В его числителе осталась бы величина лексической природы, а в знаменателе теперь появился бы числовой атрибут характеристики, отражающей наличие внутренних и внешних интерфейсов модуля. Кстати говоря, там, где его можно в духе Холстеда как-то оценить, потенциальный объём занижается, так как не учитывает внутренней архитектуры и всех возможных параметров. Поэтому в энергетическом анализе было естественным (в частности, с точки зрения противодействия систематическому смещению оценок в сравнении с наукой о программах) определять трудность на основе более адекватной (кстати, всегда не меньшей объёма(7)) величины *объёма разработки* модуля, полагая

$$D = W / V^* \qquad (\textit{трудность модуля} \text{ программы}) \tag{9}$$

где объём разработки W прибавляет к (7) вклады отношений между модулями (отношения статические, и, отчасти хронологические):

$$W(M) = V(M; N, \eta') + \sum_i V(M_i) \tag{10}$$

где M_i –интерфейсные модули системы, от которых зависит данный модуль, и которые были созданы раньше его;

V – холстедовский объем модуля (7);

η' – полный словарь нашего модуля M ($\eta' \geq \eta$), определяемый так, чтобы учитывать контекст этого модуля в программной системе (подробнее в [1-2]). В частном случае изолированного модуля получается $W = V$, и D (9) совпадает с трудностью в смысле Холстеда [6].

В итоге, ставится задача *разработки метода классификации модулей по обновлённой метрике трудности из энергетического анализа программ* (9).

Третья частная задача исследования - это *уточнение вопроса о связи между метриками трудности и оценки трудности*. М. Холстед в [6] обосновывал точку зрения о том, что (2) и (6) могут быть на практике отождествлены, но в дальнейшем исследователи пришли к другому выводу: это существенно разные метрики [15]. В работах [9-11] высказаны гипотезы о том, что, будучи разными, метрика (2) и метрика (9), обобщающая (6), могут быть взаимозаменяемы в функции фильтрации модулей по их трудности. Это было бы неплохо потому, что нередко одну из них технически трудно использовать, а другая доступна.

Наконец, в прикладном плане мы попытались на основе формального анализа с помощью подходящей выборки охарактеризовать трудность модулей системы фреймворков Matreshka [13]. Это не малая инструментальная система программных модулей, предназначенных для профессионального использования в проектах на языке Ада, которая быстро развивается. Её можно сравнивать с априори более зрелой разработкой инструментальной библиотеки OEM [14], которая, продолжая развиваться, в значительной части утилизирует прошедшие испытание временем разработки 1999-2004 гг из проекта Gnatcom.

3 Материалы исследования, относящиеся к метрике оценки трудности

По сути, необходимо проверить, что классификация модулей в соответствии с величиной оценки трудности позволяет оставить в одном полярном классе, соизмеримом по величине со всей выборкой, «лёгкие» модули, а в другом, меньшем (скажем, на порядок), чем выборка, трудные модули. Оговорим, что часть Ада модулей проекта [13] генерировалась автоматически, для них понятие трудности лишено обычного смысла, и они не рассматривались.

Из модулей разработанных программистами-людьми была с участием В. М. Годунко сделана репрезентативная выборка объёма 74. Проведена оценка научных и энергетических метрик с использованием программного анализатора SS_Ada_Scanner [2], который разрабатывался для стандарта Ада-95. Часть этой работы выполнила А. И. Паньковская при прохождении магистратуры в соответствии с планом дипломного проектирования. Поскольку система Matreshka создавалась на языке стандарта Ада-2005 с переходом на Ада-2012, то изредка новые языковые конструкции воспринимались анализатором в отдельных модулях как синтаксические ошибки с плохой локализацией места «ошибки». Это затруднение преодолевалось нами вручную – путем локальной «микромодификации» исходного текста без изменения результата оценки или с его искажением на какие-то доли процента.

Переходя к классификации, будем исходить из полезности порогов трудности (8), наличие которых отражает ограниченность комбинаторных возможностей ума. Как мы отмечали, они были подсчитаны на примере PL/S программирования, но определялись, по сути, психофизическими нормами, а потому нельзя ожидать их кардинального изменения с изменением языка программирования.

Очевидно, однако, что значения $\hat{D}$ в современных программных системах будут варьироваться намного заметнее с учётом разнообразия технологий программирования и возросшей автоматизации этого процесса в компьютерных средах разработки. Поэтому имеет смысл сразу ввести дополнительные классы трудности модулей, воспользовавшись изложенным в [7] опытом. Действительно, применим известное эвристическое положение В. И. Романовского о практической универсальности «правила трёх сигм». Тогда для большинства модулей M той выборки, с помощью которой были образованы «стандартные» пороги, должно выполняться неравенство:

$$\eta_1(M) \le E(\eta_1) + 3\sigma(\eta_1)\,, \tag{11}$$

где E – среднее по выборке, а σ – выборочное стандартное отклонение.

Следовательно, имеет смысл нанести на шкалу метрики оценки трудности дополнительный ориентир:

$$\hat{D}_3 = 0.5 \cdot (E(\eta_1) + 3\sigma(\eta_1)) \cdot \min N_2/\eta_2) = 0.5(46 + 3 \cdot 18) \cdot 5 = 250\ . \tag{12}$$

Имеет смысл ввести также аналогичные пороги

$$\hat{D}_4 = 0.5 \cdot (E(\eta_1) + 5\sigma(\eta_1)) \cdot \min N_2/\eta_2) = 340\ , \tag{13}$$

$$\hat{D}_5 = 0.5 \cdot (E(\eta_1) + 7\sigma(\eta_1)) \cdot \min N_2/\eta_2) = 430\ . \tag{14}$$

Если за пятым порогом $\hat{D}_5$ окажется число модулей, явно превосходящее численность соседних добавочных классов, то, очевидно, либо классификация оказалась узка, не современна, либо при разработке кода изучаемой системы не стремились соблюсти необходимое условие понятности модулей – их обозримость. При этом полярные классы (с наименьшими и наибольшими значениями $\hat{D}$) будут играть ту же роль, что 1-й и 3-й в классификации [6], а промежуточные классы распределят между собой роль среднего класса стандартной классификации.

Сказанное послужило основанием для следующих определений, где, однако, интерпретация трудности в форме требований к самоконтролю и контролю дана в качестве примера, и в зависимости от условий разработки и степени её критичности может конкретизироваться или корректироваться.

Определение 1. Классы трудности по метрике $\hat{D}$ определяются указанными выше пороговыми значениями:

0 класс, $\hat{D} \le D_1$ – модули, не требующие особого внимания, кроме обычного контроля разработчика («лёгкие»);

1 класс, $D_1 < \hat{D} \le D_2$ – модули, требующие дополнительного тестирования со стороны самого разработчика («умеренно трудные»);

2 класс, $D_2 < \hat{D} \le D_3$ – модули, требующие повышенного внимания со стороны разработчика, дополнительного тестирования другим исполнителем («довольно трудные»);

3 класс, $D_3 < \hat{D} \le D_4$ – модули, требующие со стороны разработчика повышенного внимания, тестирования другими исполнителями с сохранением отчётов о тестировании, возможен пересмотр кода для его упрощения («трудные»);

4 класс, $D_4 < \hat{D} \le D_5$ – модули, требующие выяснения причин высокого показателя трудности, создания предварительного плана тестирования, обсуждения в группе кода и/или тестов и/или результатов тестирования, показан пересмотр кода в смысле улучшения стиля («очень трудные»);

5 класс, $D_5 < \hat{D}$ – модули, требующие особого внимания, выяснения причин ненормально высокого показателя трудности, обязательного улучшения кода всюду, где нарушен хороший стиль, и передачи его на экспертизу в другую группу или/и декомпозиции (модификация) в соответствии со смыслом с целью кардинального снижения трудности («ненормально трудные»).

С неформальной точки зрения, в зрелой системе, как минимум, половина компонент не должны быть трудными для понимания. Численность класса максимальной трудности не должна превышать примерно 10%. Формализуем эти требования с разумной поправкой на число предусмотренных классов (у нас их всегда будет 6, но [6] рассматривалось 3, и можно представить другие числа в других классификациях).

Решающее ***правило*** **1**. Пусть имеется метрика модулей программ, которая имеет интерпретацию трудности понимания исходного кода и шкалу для классификации модулей по значениям этой метрики, определяя K классов, $3 \le K \le 9$. Программную систему будем считать компонентно зрелой по данной

метрике тогда, когда при $K \geq 5$ не менее половины всех модулей программы лежат в классе наименьшей трудности (0-м классе), а при $K<5$ там находится не меньше $\{(2K\text{-}3)/(4K\text{-}8)\}\cdot 100\%$, и в любом случае в классе максимальной трудности содержится не более $3/2(K+9)$ модулей.

При $K = 6$, класс максимальной трудности по этому определению не должен включать более десятой части всех рассмотренных модулей.

Это правило всё еще не вполне формально по двум причинам. Во-первых, системы могут быть компонентно неоднородными (одни файлы содержат исходные тексты на основном языке программирования, другие – ресурсные и т.п.). Во-вторых, редко обследуется система целиком, чаще – выборка модулей. Если она образована с использованием случайного выбора, то численности классов трудности – случайные величины, и требуется уточнить способ их сравнения с детерминированными значениями. Однако обе эти причины, как правило, связаны с уточнениями «на втором-третьем знаках после запятой». С другой стороны, если их принимать во внимание, то нужно знать конкретные обстоятельства обследования, в общем случае непредсказуемые. Ограничимся одним общим уточнением: контролировать репрезентативность используемых выборок. В частности, как принято в практической статистике, желательно, чтобы объёмы выборок были не меньше 50.

Табл. 1 показывает предварительное распределение 74 модулей обследуемой системы по значениям метрики оценки трудности. Оно имеет «длинный хвост».

Табл. 1 Распределение оценок трудности модулей в исследуемой выборке из Matreshka

$\hat{D} \leq 10^2$	$10^2 \leq \hat{D} \leq 2\cdot 10^2$	$200 \leq \hat{D} \leq 4\cdot 10^2$	$400 \leq \hat{D} \leq 8\cdot 10^2$	$800 \leq \hat{D} \leq 16\cdot 10^2$	$1600 \leq \hat{D} \leq 32\cdot 10^2$	$3200 \leq \hat{D} \leq 64\cdot 10^2$	$6400 \leq \hat{D} \leq 128\cdot 10^2$	$128\cdot 10^2 \leq \hat{D}$
$n =$ 36	$n =$ 11	$n =$ 6	$n =$ 5	$n =$ 9	$n =$ 1	$n =$ 2	$n =$ 0	$n =$ 4

По классам рассмотренные модули распределились так, как показано в табл. 2. Распределение 70% модулей по классам 0-4 демонстрирует выраженную компонентную зрелость. Однако попадание оставшихся 27% в класс ненормальной трудности указывает на вероятность того, что некий пласт модулей в исследованной версии системы формальному критерию компонентной зрелости не удовлетворяет. К выяснению причин мы вернёмся.

Табл. 2 Частоты и частости попадания модулей Matreshka в классы трудности по $\hat{D}$

Класс	0	1	2	3	4	5
Частота	39	4	6	2	2	21
Частость	0.527	0.054	0.081	0.027	0.027	0.284

4 Данные по системе, используемой для сравнения

Многие нужды пользователей, которые обеспечиваются мультиплатформенной системой Matreshka, могут быть под Windows обеспечены также библиотекой OEM [14], в большей степени ориентированной на бизнес-приложения. История её разработки и более широкий круг пользователей позволяют неформально характеризовать её как зрелую

программную систему, и ожидать выполнения критерия компонентной зрелости. При этом нужно иметь в виду асимметрию этого сравнения по смыслу. Поскольку изучаемой проблемой является степень зрелости системы Matreshka, то выборка её модулей формировалась с использованием рандомизации ради обеспечения репрезентативности. Однако для сравнительной оценки на основе научных и энергетических метрик желательна сопоставимость выборок. Для этого выборку Matreshka разделили на 3 части – большие, средние и малые по объёму модули с тем, чтобы такие модули в той же примерно пропорции были представлены в контрольной выборке из OEM. Выбор их был случаен, но после исчерпания квоты малых модулей, такие в дальнейшем не включались. По выполнении квоты для средних, пропускались и они, пока не набралось нужное число больших модулей. Разделять на большее число частей нельзя, поскольку большие модули из OEM в среднем намного меньше, чем большие из Matreshka.

Поскольку объём и оценка трудности между собой часто (положительно) коррелируют, то в случае обнаружения признаков одинаковой компонентной зрелости, напрашивающийся вывод имел бы, в силу способа формирования выборки, сомнительную значимость. Однако при обнаружении признаков различия, напротив, вывод о различии можно в данном случае делать с большей уверенностью, чем, если бы выборка из OEM была совершенно случайной.

Из 326 модулей ядра OEM была описанным способом образована выборка из 62 модулей (не взяли 74, поскольку не хватило крупных модулей). Их распределение по метрике оценки трудности показывает табл. 3. В отличие от Matreshka, где наблюдался сильно вытянутый хвост, модули из OEM ограничены рубежом, примерно равным 3000, причём в 87% случаев не переступают и рубеж 400.

Табл. 3 Распределение оценок трудности модулей в контрольной выборке из OEM

$\hat{D} \le 10^2$	$10^2 \le \hat{D} \le 2 \cdot 10^2$	$200 \le \hat{D} \le 4 \cdot 10^2$	$400 \le \hat{D} \le 8 \cdot 10^2$	$800 \le \hat{D} \le 16 \cdot 10^2$	$1600 \le \hat{D} \le 32 \cdot 10^2$	$3200 \le \hat{D} \le 64 \cdot 10^2$	$6400 \le \hat{D} \le 128 \cdot 10^2$	$128 \cdot 10^2 \le \hat{D}$
$n =$ 33	$n =$ 10	$n =$ 11	$n =$ 5	$n =$ 2	$n =$ 1	$n =$ 0	$n =$ 0	$n =$ 0

Распределение по классам трудности представлено в табл. 4. Модули OEM в 53% случаев лёгкие, и менее чем в 10% случаев попадают в класс ненормально трудных модулей. Критерий компонентной зрелости относительно $\hat{D}$ в OEM можно считать выполненным. На самом деле он выполнен ещё более рельефно, ведь по способу образования контрольной выборки, в неё вошли все наиболее крупные модули, а меньшие и мелкие только в 1/5 от их числа в системе.

Табл. 4 Частоты и частости попадания OEM модулей в классы трудности по $\hat{D}$

Класс	0	1	2	3	4	5
Частота	33	7	8	5	3	6
Частость	0.532	0.113	0.129	0.081	0.048	0.097

5 Новый метод, относящийся к метрике трудности

Замещение теоретически обоснованной метрики трудности метрикой оценки трудности (которую часто именуют просто «метрикой трудности», как в [3]) предполагало их приближенное равенство. Однако в некоторых случаях значения трудности D на порядки отличаются от значений оценки трудности $\hat{D}$.

При этом метрика D обычно даёт больший разброс по модулям большого проекта, чем $\hat{D}$. При моделировании её значений случайными величинами рискованно предполагать, что они имеют конечные средние (математические ожидания)! Поэтому нормативы, рассчитанные для $\hat{D}$ на основе средних, следует использовать при анализе распределения D на модулях проекта с крайней осторожностью (например, в только начальном приближении или при свидетельствах усиленной корреляции), как в [9].

Мы преодолеем указанную трудность, используя разделение на классы по самой выборке. Это можно делать, поскольку, в отличие от $\hat{D}$, трудность D никак не связывается с психофизическими нормативами. Она безразмерна и является, как в (7), так и в (9), грубо говоря, коэффициентом усложнения исходного текста, рассматриваемого как реализация спецификаций на данном языке программирования в сравнении с воображаемым эталонным представлением тех же спецификаций. Сравнимость классов трудности, выделенных в разных проектах, будет обеспечиваться их образованием на основе квантилей наблюдённых распределений. Введём 6 классов трудности для анализа на основе метрики D, имея в виду попытаться их сопоставить классам трудности, основанным на $\hat{D}$. Рекомендации действий снова условны.

Определение 2.

0* класс образуют все те модули (системы или исследуемой выборки), для которых трудность D не превосходит медианы значений этой метрики на рассматриваемых модулях; предполагается, что такие модули не требуют иного внимания, кроме обычного контроля разработчика;

классы с 1* по 4* определяются как соответствующие децили (с 6-го по 9-й) распределения трудности D на рассматриваемых модулях; при этом 1*-й класс содержит модули, также не требующие, по предположению, особого внимания, если на противное не укажет другой используемый индикатор трудности; модули 2*-го класса, предположительно, нуждаются в дополнительном тестировании со стороны самого разработчика, если только другой используемый индикатор трудности не позволит от этого отказаться; модули 3*-го класса рассматриваются, как безусловно требующие дополнительного интенсивного тестирования со стороны разработчика; в 4*-м классе находятся модули, требующие повышенного внимания и дополнительного тестирования другим исполнителем (или обсуждения в группе);

5* класс образуется как 10-й дециль распределения трудности D на рассматриваемых модулях; его представители требуют особого внимания, выяснения причин ненормально высокого показателя трудности (в частности, в сравнении с индикацией по другим критериям), показана их передача на

экспертизу в другую группу или декомпозиция (модификация) с целью кардинального снижения трудности.

Для определённости, если число модулей $m = 10k+2l+s$, $0 \le l \le 4$, $0 \le s \le 1$, то 1* класс полагаем содержащим $5k+l+s$ модулей, 5* класс – k, а из остальных первые l классов – по $k+1$ модулю, оставшиеся (если будут) – по k модулей.

Замечание 1. Особенность определения 2 состоит в том, что по принятому выше решающему правилу 1 всякая система модулей, отфильтрованных в классы трудности D, будет «считаться» компонентно зрелой по метрике трудности D. Дело в том, что данная классификация не нацелена на проверку зрелости систем в силу отсутствия у метрики D собственных абсолютных ориентиров приемлемой и/или неприемлемой трудности. Однако определение 2 согласуется с правилом 1 в предположении, что классификация будет применяться, как правило, к зрелым системам.

Вернёмся к рассмотренной в разделе 4 выборке модулей из безусловно зрелой системы OEM. Учитывая, что $m = 62 = 10 \cdot 6+2 \cdot 1$, и, отыскав децили, мы можем показать рубежи между классами (беря полу-суммы граничных элементов соседних классов или граничный элемент более избыточного класса):

$$D_1^* = 76{,}5\ , \quad D_2^* = 120{,}3 \quad D_3^* = 183{,}3 \quad D_4^* = 409{,}7\ , \quad D_5^* = 1152{,}0\ . \quad (15)$$

Позволят ли наблюдения за большим числом различных проектов подобрать на основе таких рубежей собственные пороги для метрики трудности D, сомнительно в силу сказанного выше. Однако, такие рубежи дают очевидный способ сравнения разных проектов по трудности их модулей.

Проведём такое сравнение с помощью уже использованной в разделе 4 выборки 74 модулей из системы Matreshka, которую мы исследуем на наличие разных признаков зрелости. Для неё получились следующие межклассовые рубежи:

$$D_1^* = 125{,}5, \quad D_2^* = 283{,}9 \quad D_3^* = 389{,}0 \quad D_4^* = 1068{,}0\ , \quad D_5^* = 5472{,}6\ . \quad (16)$$

Сравнение (16) и (15) показывает, что модули Matreshka имеют существенно большую трудность в смысле метрики D, чем модули OEM. А именно, налицо «опережение на класс»: первый рубеж трудности Matreshka (D_1^*) больше, чем второй (D_2^*) для OEM, второй – больше 3-го, третий примерно равен 4-му (точность – 0,5%), а четвёртый – 5-му (1,5%). Только один, самый трудный модуль из OEM выборки, имеет значение трудности (6461,1), соответствующее наивышему классу трудности Matreshka (который насчитывает 7 модулей).

Несмотря на то, что метрика энергетическог анализа D позволяет сравнивать трудность разных систем, но не зрелость, зрелость систем и даже отдельных модулей в энергетическом анализе также оценивается. К этому и перейдём.

6 Сравнение с результатом применения энергетического критерия

Оставаясь в рамках сравнения статических характеристик, к вопросу зрелости программного кода можно подойти иначе, чем в разделе 3. Согласно

математической модели энергетического анализа программ [1,2] спецификационная энергия и работа программирования связаны между собой законом сохранения, который учитывает также ментальные энергетические затраты разработчика:

$$E = A + Q\,, \tag{17}$$

где E – спецификационная энергия рассматриваемой подсистемы S;

A – работа, затраченная в процессе разработки программного кода S;

Q – суммарное интеллектуальное тепло, полученное от разработчика в этом процессе.

В концептуальном плане данная модель и её уравнение (17) аналогичны хорошо известной модели термодинамики и уравнению её первого начала. Подобно тому, как это имеет место при изучении термодинамических процессов, уравнение сохранения энергии в анализе программ делает возможной количественную оценку эффектов, которые непосредственно не измеряются (в физике это движения молекулярного уровня, а в программировании – творческая составляющая работы человека над кодом). При этом соотношение между достигаемым в процессе разработки значением спецификационной энергии смыслового модуля программы и затраченной на это работой,

$$A = k \cdot E\ , \tag{18}$$

определяется типом процесса [1,2], а он, в свою очередь, – использованной технологией программирования (а её выбор может зависеть от типа модуля). Ход процесса обычно недоступен для измерений, и тогда используется итоговая характеристика процесса, причём обычно вместо Q удобно рассматривать безразмерную величину нормированного информационного тепла [1,2]:

$$\mathrm{q = Q/max(E,A)} \tag{19}$$

где работа программирования отдельного модуля определяется с помощью характеристик (5), (7) формулой:

$$A = {W^2}/{V^*}\ , \tag{20}$$

а спецификационная энергия интерфейсного модуля, если он составлен из некоторого числа не сгруппированных блоков, имеет оценку через энергию его свободной группы (в исследованных модулях других групп с заметной энергией не встречалось):

$$E_{free} = \begin{cases} \lambda^{-2}(V^*)^3 & \Leftarrow m < 5 \\ \lambda^{-2}\dfrac{(V^*)^3}{m/9 + 0.5} & \Leftarrow m \geq 5 \end{cases}, \tag{21}$$

где m – количество блоков в модуле,

λ– уровень языка программирования (в наших расчетах λ=1.66).

Вынужденно опустим здесь имеющие отношение к (18) - (21) многочисленные и важные варианты, подробности, обоснования, предположения, упрощения и оговорки, которые можно найти в [3,1]. Ограничимся изложением смысла и техники проведенных оценок. Системы, разработанные на языке Ада, физически состоят из интерфейсных модулей в форме спецификаций библиотечных пакетов, тел таких пакетов (при этом не все спецификации нуждаются в телах), а также библиотечных подпрограмм, субмодулей и др. Однако рассматриваемые в статье системы составлены, в основном, из пар спецификация-тело, которые определяют

$$E_i = E(M_i), \quad A_i = A(M_i) + A(bM_i), \quad Q_i = E_i - A_i, \tag{22}$$

где M_i - описание i-го библиотечного пакета (самостоятельный компилируемый модуль, который, как правило, находится в отдельном файле);

bM_i - тело i-го библиотечного пакета (самостоятельный компилируемый модуль, который, как правило, находится в отдельном файле).

Для программной системы в целом или её подсистемы (в т.ч. репрезентативной выборки модулей) рассматриваемые метрики определяются суммированием:

$$E = \sum_i E_i, \quad A = \sum_i A_i, \quad Q = \sum_i Q_i = E - A. \tag{23}$$

Необходимо подчеркнуть, что величина A в (17), будучи прямым обобщением холстедовского усилия, может трактоваться как нормализованное время программирования [6,3], тогда как соответствующая величина (20) – это метод вычисления метрики программного кода. Связь между ними состоит в том, что метрика характеризует затраты на создание кода в неких идеализированных стандартных условиях (информация, необходимая для оценки реальных затрат [2], обычно отсутствует). В связи со сказанным, будем дальше трактовать (18), как соотношение между метриками кода. Теоретические соображения и многочисленные наблюдения за исходными кодами (причём на разных, вообще говоря, языках программирования, в частности, Ада) свидетельствуют о наличии тенденции к тому [2,1], что в зрелой программе величина k по порядку величины равна 1, а, значит, Q мало в сравнении с абсолютными величинами E и A, а, значит, q просто мало. Шкала для q уточняет это обстоятельство, а также отражает ту реальность, что необходимые оценки, в особенности, уровня языка программирования, нередко имеют погрешность на уровне 10%, так что оценки при k = 1 и, скажем, при k = 1.25 или 0.8, как правило, не имеют принципиального различия. Добавим, что значению k = 10 соответствует q = - 0.9, k = 2 соответствует q = - 0.5, k = 0.5 - q =+0.5, а k = 0.1 - q =+0.9. Если $k \to \infty$, то $q \to$ - 1, а если $k \to 0$, то $q \to +1$.

Решающее ***правило 2***. Программную систему или подсистему (в частности, отдельный модуль или логический модуль, объединяющий интерфейсный модуль и его тело вместе, если есть, со всеми субмодулями тела) считать *энергетически сбалансированной*, если

$$-0.9 \le q \le 0.9 \tag{24}$$

и *хорошо* (*энергетически*) *сбалансированной*, если

$$-0.5 \le q \le 0.5 \qquad (25)$$

Соответственно подсистема (энергетически) *не сбалансирована* при $|q|>0.9$.

В силу сказанного выше, для завершенных в разработке, эксплуатируемых или проходящих испытания систем энергетическая сбалансированность интерпретируется как показатель зрелости этой системы. Такой признак зрелости назовём *энергетической зрелостью*. При этом сбалансированность более показательна для системы в целом или её представительной части, чем для отдельного модуля и малых подсистем. Для них знак и величина q, очевидно, зависят от специализации модулей в системе, так что даже в хорошо сбалансированной системе процент сбалансированных логических модулей может быть довольно мал. Этот процент представляет статистический интерес при анализе, но в суждении о зрелости системы его роль второстепенна.

Применим правило 2 к рассматриваемым программным системам. Результаты отражены в табл. 5. Несбалансированность для системы модулей Matreshka, так же как сбалансированность для OEM хорошо различаются по численным значениям метрики нормированного информационного тепла. Процент сбалансированных логических модулей в OEM также выше, но сам по себе не высок, что, как пояснялось, не является значимой особенностью.

Табл. 5 Показатели энергетической сбалансированности в системах Matreshka и OEM

Сбалансированность модулей :	выборка в целом	хорошо сбалансированных	всех сбалансированных	несбалансированных
Matreshka (43 лог.модулей)	нет, q= -0.9999995	2.3% (1)	20.9% (9)	79.1% (33)
OEM (31 лог.молуль)	есть q = 0.81	3.2% (1)	32.3% (10)	65.0% (20)

7 Анализ результатов и выводы

Формула (2) для $\hat{D}$ вполне содержательна, и первоначально не возникает подозрений по поводу приведенной нами цитаты (С. 130) из документа [12] относительно трудности оценки пороговых значений для данной метрики. Можно, однако, рассуждать так. В грубом приближении оценка $\hat{D}$ (2) примерно пропорциональна длине программы, а, точнее, близка к значению $N/4$. В упрощающем предположении (1), которым широко пользовался сам Холстед, это очевидно. Но и в общем случае, по нашим наблюдениям, «более точная» оценки редко отклоняется от четверти длины больше, чем на 25-30%. Подоплёка в том, что символ-оператор, как правило, не выступает без операндов, а операнды не накапливаются ни в каком месте программы без разделителей, т.е. операторов. В итоге, попадание модулей в классы трудности зависит в главном от размеров модулей. Не прозрачен ли в той же степени вопрос о порогах между классами, использованными в разделе 3?

Во все времена эпохи массового программирования при разработке исходных кодов используются персональные терминалы и распечатка текстов на бумагу. Наиболее надёжно человек анализирует то, что может охватить одним взглядом, т.е. то, что умещается на странице (или в одном-двух окнах на экране монитора). В разные времена и при разных форматах это могло составлять примерно от 16 до 60 строк (например, в тексте этой статьи страница вмещает до 45 строк, а в GPS [17] - популярной среде подготовки Ада программ – окно редактора позволяет обычно видеть до 30 строк одновременно). Положим по максимуму - 60. Многолетний опыт энергетического анализа показывает, то в исходных текстах программ на ЯВУ (за вычетом из них пустых строк и многострочных комментариев) среднее число программных символов на одной строке редко выходит за диапазон 3 - 10, а чаще колеблется на уровне 5 - 7 (например, для Matreshka этот показатель составляет 5.04±1.13 при медиане 5.16). Берём 7. Получим оценку максимального размера текста программы, удобного для просмотра, как $N/4 \approx 60 \cdot 7/4 = 105.0$, т.е., близко к $D_1=115$ [3,7]. Если взять предельную оценку числа программных символов на строке, то получим $60 \cdot 10/4 = 150.0$, т.е., близко к $D_2=160$ [3,7]. Следовательно, *проводя анализ иначе*, чем это делали авторы, предложившие пороги трудности для исходных текстов программ лет 30-35 тому назад, и, исходя из других, но тоже вполне достоверных (без претензий на высокую точность) оценок психофизичеких возможностей человека, *мы приходим к аналогичным оценкам этих порогов*!

Стоит ли тогда подвергать ревизии эти пороги в современной программометрии? Приведенная выше фраза [12] о том, что «пороги, которые отличают низкое качество от высокого, должны быть оценены статистически» не вызывает возражений в принципе, но она очевидно, отражает негативный опыт попыток извлечь уточнённые значения порогов трудности из современных программ прежними методами. Это в действительности невозможно! Табл. 1 хорошо иллюстрирует ситуацию с современным программированием, когда наряду с тенденцией к ограничению размеров отдельно компилируемых модулей ради упрощения процесса (около 63.5% модулей имеют оценку <200) в современных средах программирования создаются также и сколь угодно большие модули (около 16% модулей нашего примера имеют оценку трудности около $1\text{-}5 \cdot 10^3$, а 4 – оценку >12800). В таком случае среднее (у нас 4559) и стандарт (25001) оказываются слишком велики для того, чтобы без большого риска служить ориентирами (в нашем примере 94.6% модулей попали бы в 0-й класс легких, и, не потребовали бы к себе внимания, хотя многие имеют $\hat{D}$, раз в 10 превосходящее классические пороги!). Ясно, что в других системах такие оценки давали бы совсем иной результат (например, для OEM первый подобный «порог» в 22 раза, а верхний в 51 раз меньше!). *Правильный выход из положения состоит в увеличении числа классов трудности с сохранением неизменного порога для лёгких модулей*. Это решение имеет прозрачный смысл. Почему сегодня можно успешно работать с текстами значительных размеров? Благодаря удобной навигации и гипертекстовому подходу языковых процессоров. Например, если неясен смысл элемента программы, то мы одним нажатием курсора открываем дополнительное окно с тем участком программы (или

документации), где этот элемент определён. Если модуль занимает несколько страниц, такие переходы не очень-то напрягают программиста, но при необходимости открывать много, да еще вложенных окон, ориентация разработчика снова под угрозой. Поэтому *число дополнительных классов должно быть ограничено* (знаменитый закон психологии «7±2» подсказывает не вводить более 9 классов, что и предусмотрено решающим правилом 1).

Вернёмся к вопросу о связи между метриками трудности D и оценки трудности $\hat{D}$. Коэффициент выборочной корреляции Пирсона на примере выборки 46 модулей Matreshka (результат предоставленный А. И. Паньковской) оказался для них на уровне 0.82. Заметим, что практику интересуют не только, а иногда не столько абсолютные значения трудности модулей. Так D классификация основана на порядковых статистиках явно, а с увеличением числа $\hat{D}$ классов и для этой метрики роль абсолютной величины межклассовых порогов стирается. В виду этого мы предпочли рассматривать ранговую корреляцию.

На модулях рассмотренной выше выборки 74 модулей системы Matreshka коэффициент ранговой корреляции Спирмена ρ между оценкой трудности $\hat{D}$ и трудностью D составляет $\rho = 0.336$ (S= 44842). Поэтому наличие положительной ранговой корреляции между такими оценками может быть принято в данном проекте с вероятностью ошибки первого рода (в соответствии с [16, С. 98]), равной

$$\alpha \approx 2(1 - \Phi(0.336 \cdot 8.544)) \approx 0.006 \qquad (26)$$

(примерно на 1%-м уровне значимости). При всей кажущейся определённости вывода, коэффициенты корреляции дают некую осреднённую оценку связи, ничего не гарантируя для отдельного модуля. Важно посмотреть насколько соответствуют друг другу (или, напротив, смешиваются) классы трудности разных метрик. Весь 5* класс (наиболее трудных в смысле D модулей данной выборки), 71% модулей 4* класса и 57% модулей 3* класса подтверждают свой статус тем, что попадают в 5-й класс - модули с наибольшей оценкой трудности $\hat{D}$. Обратное неверно, но тенденция есть: 80% модулей 5-го класса распределяются в трудных 3*-5* классах, причём 65% - в 4*-5* классах. С другой стороны, из 28 наиболее лёгких в смысле D модулей 71% являются также наиболее лёгкими в смысле $\hat{D}$, но 11% - наиболее трудными (попадают в 5-й класс). В свою очередь, только 7% наиболее лёгких модулей в смысле $\hat{D}$ лежат вне лёгких 0*-2* классов по метрике D, причём только 1 самый трудный из них смог «забраться» в 4* класс. Таким образом, *корреляция между признаками «трудности» и «оценки трудности» реализуется в форме удовлетворительного согласия между соответствующими классификациями. Это не удивительно потому, что Matreshka содержит заметное число явно трудных модулей, на выявление которых в первую очередь нацелены обе метрики*.

Для библиотеки OEM, которая используется в нашей работе в качестве источника контрольных примеров, коэффициент ранговой корреляции Спирмена между оценкой трудности $\hat{D}$ и трудностью D оказывается на порядок меньшим, чем для Matreshka, и равным $\rho = 0.033$ (S= 38399). Теперь нулевую

гипотезу о независимости этих признаков не удаётся отвергнуть ни на каком разумном уровне значимости. Это подтверждается и сравнением классов трудности. Все 6 модулей из класса 5* ненормально трудных по D классификации попадают в самый лёгкий 0-й класс по $\hat{D}$. Само по себе это противоречит только наличию положительной корреляции. Однако распределение лёгких модулей 0-го класса (табл. 5) исключает отрицательную корреляцию. Перемешивание классов напоминает «игру случая» и вполне *подтверждает формальный вывод о независимости рассматриваемых признаков трудности в «контрольной» заведомо зрелой библиотеке (OEM).*

Табл. 5 Распределение модулей 0-го класса метрики $\hat{D}$ по классам D метрики (OEM)

D класс	0*	1*	2*	3*	4*	5*
Частота	16	5	2	1	5	6
Частость	0.457	0.143	0.057	0.029	0.143	0.171

Итоговый вывод состоит в том, что статистическая *связь между двумя рассматриваемыми метриками* трудности модулей *может как отсутствовать* (вероятнее, в зрелых проектах), *так и присутствовать* (вероятно, при достаточном числе трудных модулей, поскольку на них нацелены обе метрики).

Фильтрация модулей по классам трудности связана с необходимостью определения модулей, требующих повышенного внимания на завершающем этапе разработки, при поиске дефектов или планировании модификации системы. Поместим все наиболее «подозрительные» модули из рассмотренных выборок в табл. 6,7, руководствуясь подходящими метриками.

Табл. 6 Модули выборки из Matreshka, имеющие признаки максимальной трудности

Модули высших рангов трудности:	D класс	$\hat{D}$ класс (ранг)	q ранг	**причина ненормальной трудности**
1. matreshka-internals-unicode-ucd-colls.ads	5*	5 (1)	2 несб.	128 громоздких константных агрегатов
2. matreshka-internals-text_codecs-iana_registry.ads	5*	5 (2)	3 несб.	858 конст. агрегатов и константный массив – 858 элементов
3. matreshka-internals-unicode-ucd-core.ads	5*	5 (5)	1 несб.	константный массив – 1354 элемента
4. matreshka-internals-unicode-ucd-core_0000.ads	5*	5 (12)	4 несб.	константный агрегат, имеющий 110 внутренних
5. matreshka-internals-unicode-ucd-cases.ads	5*	5 (3)	9 несб.	22 громоздких константных агрегата
6. matreshka-internals-unicode-ucd-core_000a.ads	5*	5(15)	5 несб.	константный агрегат, имеющий 88 внутренних
7. matreshka-internals-unicode-ucd-norms.ads	5*	5 (2)	8 несб.	50 разных константных агрегатов
A. matreshka-internals-unicode-ucd-indexes.ads	4*	5 (7)	15 несб.	константный массив – 1326 элементов

Табл. 6 составлена на основе класса ненормально трудных модулей по метрике D. Все эти 7 модулей оказались также ненормально трудными и по метрике $\hat{D}$. Поскольку ненормально трудных по $\hat{D}$ значительно больше (21), в таблице (в скобках) уточняются их рейтинги: 1 – самый трудный в смысле $\hat{D}$, 2 – следующий по трудности и т.д. Все эти модули также оказались энергетически несбалансированными (или принадлежат несбалансированной паре, состоящей из описания и тела одного и того же библиотечного модуля). Показан также (под индексом «А») самый трудный модуль из класса 4*: у него и только у него параметры мало отличаются от модулей из «топ 7» труднейших по метрике D.

Все модули таблицы при содержательном рассмотрении обнаруживают специфичесные громоздкие структуры, которые, несмотря на признаки квазирегулярности, при своих размерах «в ручную» *не контролируемы*. Например, модуль под индексом 1 имеет размер свыше 3 Мб (почти 76 тыс. строк). Если такой модуль по какой-то причине будет подлежать модификации или возникнет подозрение на дефект, это потребует особого внимания.

В случае OEM при отборе модулей, для которых прогнозируются риски, связанные с трудностью их понимания, мы не можем взять за основу один, наиболее строгий признак в силу установленной выше некоррелированности

Табл. 7 Модули выборки из OEM с максимальными рисками по причине трудности

Модули с наибольшим классовым весом:	D класс	$\hat{D}$ класс	q класс	**причина относитель-но высок.трудности**
1. oem-com-types.ads	3*	5	1	377 типов/подтипов, 145 сонстант,25групп
2. oem-com-variant.adb	3*	5	1	939 строк, 2-й по размерам, 21 операция
3. oem-com-com_interface.adb	3*	5	1	716 строк, из крупнейших, 38 операций
4. oem-com-errors.adb	4*	2	2	268 строк, 14 операций
5. oem-com-errors.ads	5*	0	2	находится в контексте oem-com-types. ads
6. oem-com-events-event_object.ads	5*	0	2	находится в контексте oem-com-types. ads и др.
7. oem-com-initialize.adb	5*	0	2	находится в контексте oem-com-types. ads
8. oem-sockets-naming.adb	2*	4	1	545 строк, 39 операций
9. oem-sockets-thin.ads	0*	5	2	469 строк, 23 типа, 49 операций
10. oem-sockets-utils.adb	5*	0	2	находится в контексте oem-sockets-thin. ads
11. oem-com-bstr.ads	5*	0	2	находится в контексте oem-com-types. ads
12. oem-com-create-gc_interface.adb	3*	2	2	234 строки, 13 операций
13. oem-gwindows-events.ads	5*	0	2	находится в контексте oem.gwindows.base. ads

рассматриваемых метрик. При отборе для табл. 7 использовался классовый вес

$$C = C_D + C_{\hat{D}} + C_q \ , \qquad (27)$$

где C_D - номер D класса трудности (напр., $C_D = 5$ для класса 5*);

$C_{\hat{D}}$ - номер $\hat{D}$ класса трудности (напр., $C_{\hat{D}} = 5$ для класса 5-го класса);

$C_q = 2$ для класса несбалансированных, $C_q = 1$ для только лишь сбалансированных и $C_q = 0$ для хорошо сбалансированных модулей.

Оказалось, что максимальный вес, равный 9 имеют только 3 модуля, один – вес 8 и девять – вес 7. Содержательный анализ исходных текстов показал, что риски, связанные с трудностью понимания, можно действительно усмотреть для модулей в позициях 1-3 табл. 7, в меньшей мере для модулей в позициях 9,8 и в ещё меньшей – 4-12. На остальные 6 отфильтрованных модулей метрики реагировали тоже «законно»: эти тексты разрабатывались с использованием описаний, заимствованных из крупных пакетов, и понимание данных модулей испытывает риск, поскольку эти заимствования потенциально могут быть весьма многочисленны. Только рассмотрение по смыслу позволило в данном случае убедиться, что заимствования малочисленны, и связанный с этим риск невелик. *В табл. 7 попали все модули выборки, которые можно заподозрить в повышенной трудности*. Следовательно, и в данном случае *фильтрация по нашему методу себя оправдала*. Отметим, что при этом в разряд подозрительных, в конечном итоге, попали все ненормально трудные модули в смысле метрики энергетического анализа D, тогда как в смысле $\hat{D}$ – две трети.

8 Заключение

Обоснована целесообразность сохранения в решающих правилах известных порогов для классической метрики оценки трудности Холстеда при условии введения дополнительных классов трудности.

Для метрики трудности энергетического анализа, более современной и реализующей понятие трудности альтернативным способом, разработана новая (относительная) классификация модулей программы. Она позволяет сравнивать модули одной программной системы и, при определённых условиях, трудность разных систем.

Уточнена природа связи между признаками трудности на основе рассмотренных метрик. В новых проектах, где велик процент трудных модулей, на большинство из них указывают значения обеих метрик, обеспечивая удовлетворение критериев наличия статистической связи. Тогда для того, чтобы «вылавливать» безусловно трудные модули, можно использовать любую из них. Чем более зрелой является изучаемая система, чем меньше там трудных модулей и ниже степень их трудности. В этом случае проявляется независимость метрик (вытекающая из различия их природы). Тогда для выделения трудных модулей целесообразно использовать обе метрики и привлекать для контроля другие (в данной работе в этой роли была полезеной метрика нормированного информационного тепла).

В анализе двух конкретных систем обнаружилось, что большую избирательность в отношении трудных модулей имеет метрика трудности энергетического анализа. Важно проверить это предположение на материале других программных систем.

В дальнейших исследованиях разработанные методы и правила целесообразно испытать в масштабах полной программной системы.

Подводя итоги, подчеркнём, что при всей важности метрических оценок тех или иных сторон качества программ, особенно в автоматизированных системах поддержки управления качеством, нельзя забывать о границах их применимости. Мы можем с большой вероятностью предсказать, какие модули программы потребуют больших затрат при своей модификации в виду большей трудности. Однако какие именно модули с высокой вероятностью придётся модифицировать при развитии системы – изученные метрики выявить не помогут. Аналогично сравнительная прогностическая оценка зрелости систем сама по себе ещё не позволяет сравнивать их пользовательское качество. В связи с этим отметим, что представленное здесь исследование является частью процесса комплексного изучения и внедрения систем фреймворков Ада программирования на кафедре моделирования систем и технологий Харьковского национального университета имени В. Н. Каразина. Планируется сопоставлять рассмотренные здесь и другие выводы, основанные на метрических характеристиках с применениями исследуемых систем в учебном процессе.

В заключние автор выражает благодарность коллегам по сообществу Ада программистов В. М. Годунко, С. И. Киркорову, А. И. Паньковской, от которых он получал содействие своей работе.

ЛИТЕРАТУРА

1. Мищенко В. О. CASE–оценка критических программных систем. Том 1. Оценка качества. / В. О.Мищенко, О. В.Поморова, Т. А. Говорущенко ; под ред. Харченко В. С. – Х : Нац. аэрокосмический ун–т «Харьк. авиац. ин–т», 2012. – 201 с.
2. Мищенко В. О. Энергетический анализ программного обеспечения с примерами реализации для Ада-программ / Виктор Олегович Мищенко. – Х.: ХНУ имени В. Н. Каразина, 2007. – 129 с.
3. 982.2-1988 - IEEE Guide for the Use of IEEE Standard Dictionary of Measures to Produce Reliable Software. - Institute of Electrical and Electronics Engineers, 1989.
4. Shaw Wade H. A Software Science Model of Compile Time / Wade H. Shaw Jr., James W. Howatt, Robert S. Maness, Dennis M. Miller // IEEE Transactions on Software Engineering. – 1989. – Vol. 15, № 5. – P. 543–549.
5. Mishchenko V. O. Does The Different Definitions Of Ada Program Tokens Have Significant Difference? / V. O. Mishchenko // Радиоэлектронные и компьютерные системы. – 2008. – № 7 (34) – С. 103–106.

6. Холстед М. Х. Начала науки о программах / М. Х. Холстед; пер. с англ. В. М. Юфа. – М.: Финансы и статистика, 1981. – 128 с.
7. Shen V. Y. Software Science Revisited: A Critical Analysis of the Theory and Its Empirical Support / V. Y. Shen, S. D. Conte, H. E. Dunsmore // IEEE Transactions on Software Engineering. – 1983. – Vol. SE-9, № 2. - P. 155-165.
8. Генералов К. А. Анализ эффективности использования генетических алгоритмов в задачах при использовании различных языков программирования // Вопросы современной науки и практики. Университет им. В.И.Вернадского. №2(12). 2008. Том 2. С. 148-154.
9. Годунко В. М. Качество транслятора шаблонов динамических html страниц для Ada WEB серверов / В. М. Годунко, В. О. Мищенко, М. М. Резник, Д. В. Штефан // Радіоелектронні і комп'ютерні системи. – 2012. – № 5. – С. 225-229.
10. Боровинский А. В. Сложность реализации интеллектуальных графических интерфейсов для приложений, основанных на МДО / А. В. Боровинский, В. О. Мищенко // Радіоелектронні і комп'ютерні системи. – 2012. – № 7. – С. 260–265.
11. Литвинов Д. Н. Применение энергетических метрик для оценки использования ASIS и в других подобных задачах / Д. Н. Литвинов, В. О. Мищенко // Радіоелектронні і комп'ютерні системи. – 2012. – № 7. – С. 301-306.
12. 982.1-2005 - IEEE Standard Dictionary of Measures of the Software Aspects of Dependability. - Institute of Electrical and Electronics Engineers, 2006. – 34 p.
13. Matreshka [Электронный ресурс] / Vadim Godunko, IE //01.04.2014. – Режим доступа: http://forge.ada-ru.org/matreshka
14. Ada Software – Free DownLoad Ada OEM-2009/2010/2011/2012 library [Электронный ресурс] / MediaScan.by // 15.10.2014. – Режим доступа http://www.mediascan.by/index.files/Page695.html .
15. Hamer Peter G. M.H. Halstead's Software Science – a critical examination / Peter G. Hamer, Gillian D. Frewin // International Conference on Software Engineering. Proceedings of the 6th international conference on Software engineering. - Tokyo, 1982. – P. 197-206.
16. Большев Л.Н. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – М.: Наука. Главная редакция физико-математической литературы, 1983. – 416 с.
17. GNAT Programming Studio - Википедия Matreshka [Электронный ресурс] http://ru.wikipedia.org/wiki/GNAT_Programming_Studio // 15.10.2014. – Режим доступа http://www.mediascan.by/index.files/Page695.html

Надійшла у першій редакції 23.10.2014, в останній - 03.11.2014.

УДК 519.6

Метод решения возмущённых краевых задач, способных моделировать деформированные состояния замкнутых торсовых оболочек

В. И. Олевский

Украинский государственный химико-технологический университет, Украина

Излагается разработанный автором вычислительный метод решения задач теории замкнутых гибких торсовых оболочек, имеющих малые возмущения края в плоскости, перпендикулярной оси оболочки. Метод носит комплексный характер и, в частности, отличается формой введения параметра для получения решения в виде двойного асимптотического ряда, который удаётся просуммировать с заданной точностью двумерными дробно-рациональными приближениями. Это позволяет обосновать сходимость к точному решению в данной модели, а численные эксперименты показывают, что метод эффективен при существенно больших амплитудах возмущения, чем у других авторов.

Ключевые слова: *модель, оболочка, возмущения, асимптотический ряд, двумерное суммирование, дробно-рациональные приближения.*

Викладається розроблений автором обчислювальний метод розв'язання крайових задач теорії замкнутих гнучких торсових оболонок, що мають малі збурення краю в площині, перпендикулярній осі оболонки. Метод має комплексний характер та, зокрема, відрізняється формою введення параметра для отримання рішення у вигляді подвійного асимптотичного ряду, який вдається підсумувати із заданою точністю двовимірними дробово-раціональними наближеннями. Це дозволяє обґрунтувати збіжність до точного рішення моделі, що задана, а чисельні експерименти показують, що метод ефективний при істотно більших амплітудах збурення, ніж у інших авторів.

Ключові слова: *модель, оболонка, збурення, асимптотичний ряд, двовимірне підсумовування, дробово-раціональні наближення.*

The author describes his own computational method of solving boundary value problems in the theory of closed flexible torso shells with small perturbations of edges in the plane, which is perpendicular to the shell axis. The method is of complex character. The feature of the method is, particularly, a technique of enabling of parameter, which allows to obtain solution in the form of a double asymptotic series that might be summed with given accuracy using two-dimensional fractional rational approximations. This allows to prove convergence to the exact solution in the present model; and numerical experiments show that the method is effective for significantly larger perturbation amplitudes than that of other authors.

Key words: *model, shell, perturbations, asymptotic series, two-dimensional summation, fractional rational approximation.*

1 Введение

Необходимость учета малых отклонений различного рода в начальных и граничных условиях, а также в разрешающих уравнениях возникла с самого начала использования математических моделей в механике. Особую актуальность задачи моделирования технических устройств с отклонениями приобрели с использованием высокоточного оборудования, машин, облегченных машиностроительных деталей и строительных конструкций [1]. Оказалось, что в целом ряде задач отклонения основных (первичных)

параметров или изменение уровня заданных нестрого (вторичных) параметров оказывают существенное влияние на показатели прочности и механической устойчивости оборудования, и даже могут привести к его разрушению и значительным материальным потерям [2, 3].

Несмотря на активную разработку в течение последних десятилетий вычисленных методов для задач механики оболочек [4, 5], прогресс в методах расчета возмущенных состояний торсовых оболочек (цилиндрических, конических) тормозился. Не в последнюю очередь это связано с отсутствием устоявшейся классификации и математического описания отклонений. Главное, однако, состоит в том, что учет малых отклонений путем использования популярных методов дискретизации расчетной области приводит к существенному увеличению объема вычислений и, что особенно важно, к появлению в разрешающих системах членов существенно различной величины. В частности, для сеточных методов размер ячейки должен быть меньше исследуемого возмущения. Для метода конечных элементов дробное разбиение часто приводит к изменению характера конечных элементов и необходимости использования более сложных их форм: например, оболочечные элементы необходимо заменять объемными и согласовывать их деформирование. Это приводит к сильной обусловленности разрешающих систем уравнений, потере точности расчета и, как следствие, к некорректным результатам моделирования. Сложная форма отклонений часто делает непригодным стандартные методы генерирования дискретных подобластей, что еще более усложняет задачу. Использование сглаживающих функций требует на порядок более высокой точности приближения, позволяющей учитывать величину малых отклонений. Получаемые при этом результаты носят вид массивов чисел, что требует дополнительно проведения их интерпретации, визуализации и сглаживания. Расчет для каждого конкретного объекта в этих случаях необходимо производить индивидуально, начиная с разбиения области и аппроксимации нагрузок. Это значительно усложняет процесс расчета и сужает набор пригодных для использования программных средств.

В принципе, хорошо известен и другой путь моделирования состояний систем с отклонениями – использование метода возмущения по параметру, который существует в различных формах [6–8]. Этот подход нацелен на получение приемлемых приближений для малых отклонений и, соответственно, малых значений параметра возмущения. При этом можно выяснять качественную природу решений и закономерности поведения моделируемой системы. Однако, по крайней мере, для интересующего нас класса задач, получение достаточной для практического использования точности вычислений при реальных по величине отклонениях часто влечёт недопустимо большой объём вычислительной работы. Это характерно для т. н. сингулярных задач, содержащих малый параметр при старшей производной и для нелинейных задач. Именно такие задачи наиболее характерны для гибких торсовых оболочек имеющих отклонения формы [9]. Решение задачи в этом случае становится, громоздким из-за большого количества требуемых аналитических преобразований сложных функций.

Важным этапом на пути решения этой проблемы оказалось привлечение к стандартной схеме использования возмущения по параметру методов обобщенного суммирования [10]. Благодаря этому появилась возможность получения достаточно точного решения для сравнительно больших значений параметра при малом числе приближений. Однако характер и закономерности приближений данного подхода оставались в сколько-нибудь общем случае не изученными. Исследователи лишь констатировали или достижение нужной точности, или расходимость приближений в ходе решения частных задач (довольно полный обзор публикаций данного направления см. в [11]).

При решении частных краевых задач механики упругих гибких торсовых оболочек [12,13,1], автору, используя подход асимптотического характера, удалось фактически создать новый комплексный вычислительный метод, позволяющий производить адекватные реальности приближенные расчёты для значений параметра возмущения области, существенно превосходящих интервал адекватности известных ранее методов. Природа этого эффекта, как сейчас стало окончательно понятно, связана с тем, что успешный путь использования асимптотического подхода замаскирован напрашивающимися аналогиями со схемами, применяемыми для одномерных рядов. Однако возникающие двумерные ряды не являются, как правило, безусловно сходящимися. Поэтому их усечение, использующее произвольный (удобный для записи, но необоснованный) порядок суммирования, не позволяет по объективной причине получать ни оценок сходимости, ни адекватных экспериментальным данным численных результатов при амплитудах возмущений, которые интересуют создателей технических систем. Напротив, в работах [12,13,1] структуре асимптотического ряда и удержанию подходящего числа его членов уделено первостепенное внимание с тем, чтобы удержанные члены позволяли провести обобщенное суммирование ряда дробно-рациональными приближениями. В настоящей работе используется соответствующая специальная техника построения двумерных аппроксимант Падé–типа, которая находится в согласии в с одним фундаментальным результатом В.В. Вавилова [14] и может рассматриваться как развитие вычислительной схемы, применявшейся ранее для одномерных рядов в другой области исследований [6].

2 Цель работы и объяснение метода на примере решения модельной задачи

Рассматривается задача построения приближенных решений возмущённых периодических краевых задач для систем уравнений в частных производных определённого класса, включающего известные задачи теории упругих гибких торсовых оболочек. Целью работы является разработка и обоснование схемы комплексного вычислительного метода для расчета таких решений, который обеспечивает их сходимость и получение приемлемого для приложений результата при наиболее широких условиях, допустимых в рамках асимптотического подхода к использованию рядов с параметрами возмущения.

Для того, чтобы особенности разработанного метода не потерялись в громоздких формулах и прочих технических подробностях, мы продемонстрируем этот метод на специально подобранном для этой цели

модельном примере, который принадлежит к рассматриваемому далее классу, охватывающему в качестве частных случаев, модели замкнутых торсовых оболочек и краевые задачи теории погранслоя в гидродинамике. Важной особенностью примера является наличие точного решения в аналитической форме, что позволяет иллюстрировать свойства приближенных решений относительно точного решения. Упрощает детали также и невысокий второй порядок уравнения (тогда как в механике оболочек порядки уравнений упругого состояния всегда выше).

Итак, рассмотрим граничную задачу вида

$$\varepsilon^2 \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,\ 0 < \varepsilon < 1,\ u(x,0) = u(x,\pi) = 0,\ u(0,y) = \sin y,\ u(a,y) = e^{-a/\varepsilon} \sin y,$$

которая имеет точное решение $u(a,y) = e^{-x/\varepsilon} \sin y$. Разложение точного решения в ряд по степеням x

$$u(a,y) = \sin y \left(1 - x/\varepsilon + (x/\varepsilon)^2 - (x/\varepsilon)^3 + \ldots + (-1)^{n-1} (x/\varepsilon)^n + \ldots\right)$$

сходится неравномерно из-за присутствия малого параметра ε в знаменателе общего члена ряда, поэтому для приближенного решения такой задачи при помощи возмущения по параметру требуется использование специальных методов [16, 17].

Найдем приближенное решение этой задачи для возмущенной границы $a = a_0 + \delta a$ по предлагаемому методу [12] следующим образом. Сначала произведем возмущение формы границы по искусственному параметру $\tilde{\varepsilon}$. Будем искать решение в виде ряда $U = \sum_{m=1}^{\infty} U_m \tilde{\varepsilon}^m$. Приравнивая нулю коэффициенты при каждой степени параметра, получим последовательность т. н. предельных задач. Это последовательность рекуррентных граничных задач на прямоугольнике с границей a_0 для определения U_m вида

$$\tilde{\varepsilon}^0: \quad \varepsilon^2 \frac{\partial^2 U_0}{\partial x^2} + \frac{\partial^2 U_0}{\partial y^2} = 0,$$

$$U_0(x,0) = U_0(x,\pi) = 0,\ U_0(0,y) = \sin y,\ U_0(a_0,y) = e^{-(a_0+\delta a)/\varepsilon} \sin y,$$

$$\tilde{\varepsilon}^1: \quad \varepsilon^2 \frac{\partial^2 U_1}{\partial x^2} + \frac{\partial^2 U_1}{\partial y^2} = 0,$$

$$U_1(x,0) = U_1(x,\pi) = 0,\ U_1(0,y) = 0,\ U_1(a_0,y) = -\frac{\partial U_0}{\partial x} \delta a,$$

$$\tilde{\varepsilon}^2: \quad \frac{\partial^2 U_2}{\partial x^2} = -\frac{\varepsilon_1}{\varepsilon^2} \frac{\partial^2 U_2}{\partial y^2},$$

$$U_2(x,0) = U_2(x,\pi) = 0,\ U_1(0,y) = 0,\ U_2(a_0,y) = -\frac{\partial U_1}{\partial x} \delta a - \frac{1}{2} \frac{\partial^2 U_0}{\partial x^2} (\delta a)^2,$$

$$\ldots.$$

Для регуляризации разрешим уравнение относительно старшей производной, что эквивалентно приведению к нормальному виду, и введем искусственный

параметр ε_1 в так, что при $\varepsilon_1 = 0$ получим упрощенные задачи на прямоугольнике, а для $\varepsilon_1 = 1$ – исходные:

$$\tilde{\varepsilon}^0: \quad \frac{\partial^2 U_0}{\partial x^2} = -\frac{\varepsilon_1}{\varepsilon^2}\frac{\partial^2 U_0}{\partial y^2},$$

$$U_0(x,0) = U_0(x,\pi) = 0,\ U_0(0,y) = \sin y,\ U_0(a_0,y) = e^{-(a_0+\delta a)/\varepsilon}\sin y,$$

$$\tilde{\varepsilon}^1: \quad \frac{\partial^2 U_1}{\partial x^2} = -\frac{\varepsilon_1}{\varepsilon^2}\frac{\partial^2 U_1}{\partial y^2},$$

$$U_1(x,0) = U_1(x,\pi) = 0,\ U_1(0,y) = 0,\ U_1(a_0,y) = -\frac{\partial U_0}{\partial x}\delta a,$$

$$\tilde{\varepsilon}^2: \quad \frac{\partial^2 U_2}{\partial x^2} = -\frac{\varepsilon_1}{\varepsilon^2}\frac{\partial^2 U_2}{\partial y^2},$$

$$U_2(x,0) = U_2(x,\pi) = 0,\ U_1(0,y) = 0,\ U_2(a_0,y) = -\frac{\partial U_1}{\partial x}\delta a - \frac{1}{2}\frac{\partial^2 U_0}{\partial x^2}(\delta a)^2,$$

$$\dots.$$

Решение будем искать в виде асимптотических рядов $\tilde{U}_m = \sum_{i=0}^{\infty}\varepsilon_1^i u_{mi}(x)$, $U_m = \tilde{U}_m \sin y$. Для каждого $\tilde{U}_m$ получим последовательность предельных граничных задач для обыкновенных дифференциальных уравнений (ОДУ) вида

$$\tilde{\varepsilon}^0: \quad \varepsilon_1^0: u_{00}'' = 0,\ u_{00}(0) = 1,\ u_{00}(a_0) = e^{-(a_0+\delta a)/\varepsilon},$$

$$\varepsilon_i^1: u_{0i}'' = u_{0(i-1)}/\varepsilon^2,\ u_{0(i-1)}(0) = u_{0(i-1)}(a_0) = 0,\ i = \overline{2,\infty},$$

$$\tilde{\varepsilon}^1: \quad \varepsilon_1^0: u_{10}'' = 0,\ u_{10}(0) = 0,\ u_{10}(a_0) = -\tilde{U}_0'(a_0)\delta a,$$

$$\varepsilon_i^1: u_{1i}'' = u_{1(i-1)}/\varepsilon^2,\ u_{1(i-1)}(0) = u_{1(i-1)}(a_0) = 0,\ i = \overline{2,\infty},$$

$$\dots$$

$$\tilde{\varepsilon}^m: \quad \varepsilon_1^0: u_{20}'' = 0,\ u_{m0}(0) = 0,\ u_{m0}(a_0) = \varphi_m,$$

$$\varepsilon_i^1: u_{mi}'' = u_{m(i-1)}/\varepsilon^2,\ u_{m(i-1)}(0) = u_{m(i-1)}(a_0) = 0,\ i = \overline{2,\infty},$$

$$\dots,$$

где φ_m – значение на границе, определяемое предыдущими приближениями.

Последовательное решение этих задач позволяет получить u_{mi} в виде полиномов по степеням переменной x, порядок которых возрастает с увеличением порядка приближения:

$$\tilde{\varepsilon}^0: \quad u_{00} = \left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)x/a_0 + 1,$$

$$u_{01} = \frac{1}{\varepsilon^2}\left[\left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)x^3/(6a_0) + x^2/2 - x\left(\left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)a_0^2/6 + a_0/2\right)\right],$$

$$u_{02} = \frac{1}{\varepsilon^4}\left[\left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)x^5/(120a_0) + x^4/24 - x^3\left(\left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)a_0^2/36 + a_0/12\right) - \right.$$

$$\left. - x\left(\left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)a_0^3/120 + a_0^3/24 - a_0^2\left(\left(e^{-(a_0+\delta a)/\varepsilon} - 1\right)a_0^2/36 + a_0/12\right)\right)\right],$$

$$\dots \quad ,$$

$$\tilde{\varepsilon}^m: \qquad u_{m0} = x\varphi_m / a_0 \,,\; u_{m1} = x^3 \varphi_m \Big/ \left(6\varepsilon^2 a_0\right) - x\varphi_m a_0 \Big/ \left(6\varepsilon^2\right),$$

$$u_{m2} = x^5 \varphi_m \Big/ \left(120\varepsilon^4 a_0\right) - x^3 \varphi_m a_0 \Big/ \left(36\varepsilon^4\right) - x\left(x^3 \varphi_m \Big/ \left(120\varepsilon^4\right) - x^2 \varphi_m a_0 \Big/ \left(36\varepsilon^4\right)\right),$$

$$\dots \quad .$$

Анализ вида коэффициентов $\tilde{U}_m$ показывает, что радиус их сходимости зависит от величины малого параметра и сходимость будет неравномерной из-за присутствия малого параметра ε в знаменателе. Для ускорения сходимости и регуляризации ряда используем обобщенное суммирование дробно-рациональными приближениями [12, 13].

Полученные ряды являются двойными – по степеням искусственного параметра и переменной интегрирования:

$$\tilde{U}_m = \sum\nolimits_{j=0}^{\infty} \sum\nolimits_{i=0}^{\infty} u_{mij} \varepsilon_1^i x^j \,.$$

Дробно-рациональное преобразование рядов можно провести различными путями. Первый путь – суммирование только по степеням параметра ε_1, как предусмотрено в существовавших ранее схемах. При этом ряд приближается дробно-рациональной функцией вида

$$PA_{\varepsilon[n_1,n_2]}\left(\tilde{U}_m\right) = \sum\nolimits_{i=0}^{n_1} a_i(x)\varepsilon_1^i \Big/ \left(1 + \sum\nolimits_{i=1}^{n_2} b_i(x)\varepsilon_1^i\right).$$

Результат суммирования в этом случае не имеет теоретического обоснования сходимости к истинному решению.

Другим способом, который может быть обоснован, является приближение только по степеням переменной интегрирования x:

$$PA_{x[m_1,m_2]}\left(\tilde{U}_m\right) = \sum\nolimits_{i=0}^{m_1} a_i(\varepsilon_1) x^i \Big/ \left(1 + \sum\nolimits_{i=1}^{m_2} b_i(\varepsilon_1) x^i\right).$$

Этот подход не всегда обеспечивает сходимость решения для $\varepsilon_1 = 1$.

Наиболее приемлемым является двумерное суммирование по ε_1 и x с использованием результатов [14], которые описывают условия сходимости к точному решению приближения вида

$$PA_{[n_1,n_2][m_1,m_2]}\left(\tilde{U}_m\right) = \sum\nolimits_{j=0}^{m_1} \sum\nolimits_{i=0}^{n_1} a_{ij} x^j \varepsilon_1^i \Big/ \sum\nolimits_{j=0}^{m_2} \sum\nolimits_{i=0}^{n_2} b_{ij} x^j \varepsilon_1^i \,,\; b_{00} \equiv 1 \,.$$

Практическое поведение различных приближений приведено на рис. 1.2 для $\varepsilon = 0{,}5$, $a = 1$, $\delta a = 0{,}5$ при $y = \pi/2$, где тонкой сплошной линией показано точное решение, точками – сумма двух первых членов приближения по параметру, пунктиром – дробно-рациональное преобразование приближения только по параметру, штрихпунктиром – только по переменной интегрирования, и толстой линией – двумерная дробно-рациональная модель.

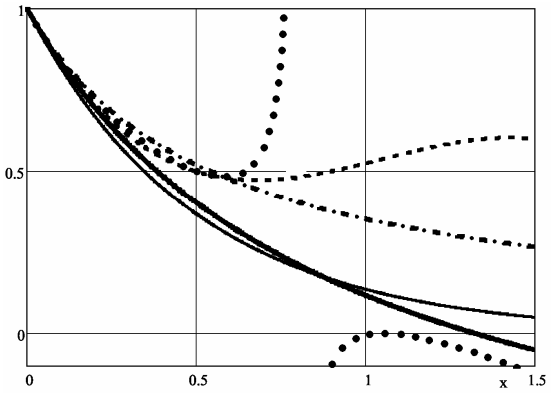


Рис. 1.2. Сравнение точного решения и моделей

Видно, что двумерная дробно-рациональная модель существенно расширяет радиус сходимости по переменной интегрирования и дает более точный результат даже при малом числе приближений.

3 Постановка задач в общем случае

В евклидовом пространстве $\mathbb{E}^2$ двух переменных $\{\eta,\xi\}\in\mathbb{E}^2$ рассмотрим замкнутую односвязную криволинейную трапецию Ω, близкую к прямоугольнику Ω_0, вида

$$\{\Omega:-\pi<\eta<\pi,-f_1(\eta)<\xi<1+f_2(\eta)\},\{\Omega_0:-\pi<\eta<\pi,0<\xi<1\},$$
$$\Omega_0\subset\Omega,\ f_i(\eta)\geq 0,\ \max_{\eta\in[-\pi,\pi]} f_i(\eta)\ll 1,\ i=1,2, \qquad (3.1)$$

где $f_i(\eta)$ – непрерывные функции на отрезке $[-\pi,\pi]$.

На область Ω могут быть отображена, в частности, срединная поверхность замкнутых оболочек нулевой гауссовой кривизны, имеющих гладкое возмущение торцов в плоскости, перпендикулярной оси оболочки [1]. Граница трапеции $\partial\Omega$ состоит из четырех участков вида

$$\partial\Omega_1=\partial\Omega\big|_{\eta=-\pi},\ \partial\Omega_2=\partial\Omega\big|_{\eta=\pi},\ \partial\Omega_3=\partial\Omega\big|_{\xi=f_1(\eta)},\ \partial\Omega_4=\partial\Omega\big|_{\xi=1+f_2(\eta)},$$

а граница прямоугольника $\partial\Omega_0$ – из четырех участков вида

$$\partial\Omega_{01}=\partial\Omega\big|_{\eta=-\pi},\ \partial\Omega_{02}=\partial\Omega\big|_{\eta=\pi},\ \partial\Omega_{03}=\partial\Omega_0\big|_{\xi=0},\ \partial\Omega_{04}=\partial\Omega_0\big|_{\xi=1}\ .$$

На замыкании $\overline{\Omega}=\Omega\cup\partial\Omega$ введем множество равномерно ограниченных функций

$$\{U:U_i\in U,i=\overline{1,N}\}, \qquad (3.2)$$

имеющих на $\overline{\Omega}$ равномерно ограниченные непрерывные частные производные степени не ниже n_i для каждой функции u_i, вида

$$\left\{U^{(n)}:U_{ikp}\in U^{(n)},U_{ikp}=\frac{\partial^{k+p}U_i}{\partial\eta^k\partial\xi^p},0\leq k+p\leq n_i,\ k,p=\overline{1,n_i},\ i=\overline{1,N}\right\}, \qquad (3.3)$$

множество непрерывных частных производных не ниже n_i-1 степени вида

$$\left\{U^{(n-1)} : U_{ikp} \in U^{(n)},\ 0 \le k+p \le n_i - 1,\ k,p = \overline{1,n_i - 1},\ i = \overline{1,N}\right\} \tag{3.4}$$

и множество частных производных старшего порядка

$$\left\{U^{(\max)} : U_{ikp} \in U^{(\max)}, U_{ikp} \in U^{(n)},\ k+p = n_i,\ k,p = \overline{1,n_i},\ i = \overline{1,N}\right\}. \tag{3.5}$$

При этом $U_i = U_{i00},\ i = \overline{1,N}$.

Пусть дана система N нелинейных дифференциальных уравнений в частных производных вида

$$\Lambda_j\left(\eta,\xi,U^{(n)}\right) = \Phi_j\left(\eta,\xi,U^{(n-1)}\right),\ j = \overline{1,N}, \tag{3.6}$$

где Λ_j, Φ_j – алгебраические аналитические функции, равномерно ограниченные на $\overline{\Omega}$ и представимые на нем своими рядами Маклорена относительно η,ξ,U_{ikp} (все компоненты воспринимаются как независимые переменные), Λ_j линейны относительно компонентов $U^{(\max)}$.

Система уравнений должна быть дополнена граничными условиями на $\partial\Omega$. Примем условие периодичности решения по переменной η, отражающее условия моделирования замкнутых оболочек, вида

$$U^{(n)}\Big|_{\partial\Omega_1} = U^{(n)}\Big|_{\partial\Omega_2}. \tag{3.7}$$

На участках $\partial\Omega_3$ и $\partial\Omega_4$ необходимо задать граничные условия вида

$$\partial\Omega_k : G_j^k\left(\eta, U^{(n-1)}\Big|_{\partial\Omega_k}\right) = 0,\ j = \overline{1,n_k},\ k = 3,4, \tag{3.8}$$

где G_j^k – некоторые ограниченные кусочно-непрерывные алгебраические функции всех своих аргументов.

Необходимо найти решение системы (3.6) на множестве U, удовлетворяющее граничным условиям (3.7) – (3.8) на $\partial\Omega$. Количество независимых граничных условий (3.8) определяется конкретным видом системы уравнений. Они должны давать возможность определения всех произвольных функций, входящих в общее решение системы.

Задача (3.6) – (3.8) является, в общем случае, нелинейной краевой задачей со сложными граничными условиями для неканонической (возмущенной) области. Вид (3.6) – (3.8) имеют некоторые задачи механики, в частности, краевые задачи теории пластин и торсовых оболочек [19].

4 Схема вычислительного метода для общего случая

Данный метод представляет собой последовательность формализованных действий для получения эффективного решения поставленной граничной задачи (3.6) – (3.8) в виде двумерных дробно-рациональных моделей. Эта последовательность такова: краевая задача для возмущенной области приводится к нормальной форме, а затем приближается рекуррентной последовательностью предельных краевых задач с полиномиальной нелинейностью для прямоугольника. Решение предельных краевых задач в виде тригонометрического ряда по циклической переменной сводится к граничным

задачам для систем ОДУ. Граничные задачи для систем ОДУ решаются путем специального вида возмущения при помощи искусственного параметра так, что последовательные приближения получаются в виде полиномов. Для обеспечения сходимости решения одновременно по параметру возмущения и переменной интегрирования используется обобщенное суммирование его двумерными дробно-рациональными функциями. Суммирование производится на основании специальной схемы выбора членов ряда, используемых для построения приближения, что обеспечивает его существование, единственность и сходимость к точному решению.

Приведение системы уравнений к нормальной форме позволяет провести в дальнейшем ее регуляризацию при помощи введения искусственного параметра. Метод возмущения формы границы позволяет перейти к рассмотрению последовательности рекуррентных задач в канонической области – на прямоугольнике. Замена нелинейности системы и решений полиномиальными приближениями позволяет ограничить вид предельных систем и обеспечить общую форму решения задачи в виде полиномов. Разложение искомого решения и входящих в систему функций в тригонометрические ряды является стандартным методом разделения переменных на прямоугольнике и сводит решение задачи к интегрированию граничных задач для систем ОДУ. Особенностью метода является специальная форма возмущения этих задач, которая приводит к получению приближений в виде двумерных полиномов по параметру возмущения и переменной интегрирования. Такая форма приближенного решения позволяет выявить некоторое условие сходимости к точному решению и в этом предположении провести его обобщенное суммирование дробно-рациональными функциями, обеспечивающее расширение радиуса сходимости регулярного приближения по обеим переменным.

5.1. Сведение к краевой задаче для уравнений первого порядка

Хорошо известно, что система дифференциальных уравнений в частных производных произвольного порядка может быть представлена эквивалентной системой 1-го порядка. Для этого достаточно ввести в качестве дополнительных искомых функций все частные производные от каждой функции U_i до порядка $n_i - 1$ включительно (эта замена корректна, если хотя бы одна производная порядка n_i входит в какое-либо уравнение рассматриваемой системы). В наших обозначениях это означает переход от системы искомых функций U к системе $U^{(n)}$ с разверткой мультииндекса компонент в линейную последовательность номеров функций. Обозначим $U^{(1)}$ упорядоченное таким образом множество вида

$$\left\{U^{(1)} : U_j^{(1)} = U_{ikp},\ j = \overline{1,r},\ 0 \le k + p \le n_i,\ k,p = \overline{1,n_i},\ i = \overline{1,N}\right\}. \qquad (5.1)$$

Без потери общности можно считать, что $U_i^{(1)} = U_i, U_{r-N+i}^{(1)} = U_i^{(\max)}, i = \overline{1,N}$. Ясно, что при этом систему следует пополнить уравнениями, отражающими равенство различных смешанных производных. В итоге граничная задача (3.6) – (3.8)

записывается относительно r компонент $U^{(1)}$ взамен $U^{(n)}$, но не меняет своего общего вида. Запишем систему вида (4.6) в нормальной форме, в которой исходная часть уравнений разрешена относительно производных по переменной ξ, вида

$$LU_j = F_j\left(\eta,\xi,\left\{U_i^{(1)}, i=\overline{1,r-N}\right\}\right),\ L=\frac{\partial(\ )}{\partial\xi}, \tag{5.2}$$

а остальные являются уравнениями вида

$$\frac{\partial U_j}{\partial\xi}=U_k,\quad \frac{\partial U_p}{\partial\eta}=U_q. \tag{5.3}$$

Здесь F_j – алгебраические аналитические функции, представимые своими рядами Маклорена относительно η,ξ,U_i.

5.2. Асимптотический подход к построению решения на прямоугольнике

При решении задач методом возмущения формы границы требуется представление искомых и известных функций в виде рядов по переменной, в направлении оси которой происходит возмущение [19, 20]. Представим аналитические функции F_j и G_j^k в области своего определения сходящимися кратными рядами Маклорена относительно ξ и всех входящих в них компонент $U^{(1)}$ вида

$$F_j\left(\eta,\xi,U^{(1)}\right)=\sum_{q=0}^{\infty}F_{j0q}(\eta)\xi^q+\sum_{k=1}^{r-N}\left[\sum_{q=0}^{\infty}F_{jkq}(\eta)\xi^q\right]U_k^{(1)}+$$
$$+\sum_{k=1}^{r-N}\sum_{i=1}^{r-N-k+1}\left[\sum_{q=0}^{\infty}F_{jikq}(\eta)\xi^q\right]U_i^{(1)}U_k^{(1)}+\dots, \tag{5.4}$$
$$G_j^k\left(\eta,U^{(1)}\right)=G_{j0}^k+\sum_{l=1}^{r-N}G_{jl0}^kU_l^{(1)}\Big|_{\partial\Omega_k}+\sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1}G_{jil}^kU_i^{(1)}\Big|_{\partial\Omega_k}U_l^{(1)}\Big|_{\partial\Omega_k}+\dots$$

при этом η рассматривается как параметр, а ξ и $U_i^{(1)}$ – как независимые переменные.

Переходя к общим условиям, которыми ограничим рассмотрение метода, отметим, что в большинстве прикладных задач механики и теории оболочек функции, определяющие нелинейность уравнений, изначально имеют вид полинома, причем степень его редко превышает куб [17, 18]. Нами, предполагая по постановке задачи поиск «ограниченного» или «медленного» решения, имеющего ограниченные производные до n_i порядка, ряды в (12) заменяются своими отрезками – многомерными полиномами от $U^{(1)}$ конечной степени. Это получит обоснование после установления соответствующей (условной) сходимости данных рядов. Иначе – при высокой степени нелинейности и т. н. «быстром» или быстро осциллирующем решении, – следовало бы использовать упрощенные предельные системы, позволяющие выявить главные члены

уравнений для конкретных параметров граничной задачи и характера искомого решения [19].

Нелинейная граничная задача (3.7), (3.8), (5.2) – (5.3) на криволинейной трапеции Ω в предлагаемом методе приводится к последовательности граничных задач на прямоугольнике Ω_0 в ходе процедуры асимптотического возмущения формы границы, что является стандартным шагом [20–22]. Для этого рационально использовать искусственный параметр возмущения ε, который введем специальным образом:

$$\left\{\Omega_\varepsilon : -\pi < \eta < \pi, -\varepsilon f_1(\eta) < \xi < 1 + \varepsilon f_2(\eta)\right\},$$

$$\partial\Omega_{\varepsilon 1} = \partial\Omega\big|_{\eta=-\pi},\ \partial\Omega_{\varepsilon 2} = \partial\Omega\big|_{\eta=\pi},\ \partial\Omega_{\varepsilon 3} = \partial\Omega_\varepsilon\big|_{\xi=-\varepsilon f_1(\eta)},\ \partial\Omega_{\varepsilon 4} = \partial\Omega_\varepsilon\big|_{\xi=1+\varepsilon f_2(\eta)}.$$

$$LU_j = \sum_{q=0}^{\infty} F_{j0q}(\eta)\xi^q + \sum_{k=1}^{r-N}\left[\sum_{q=0}^{\infty} F_{jkq}(\eta)\xi^q\right]U_k^{(1)} + \sum_{k=1}^{r-N}\sum_{i=1}^{r-N-k+1}\left[\sum_{q=0}^{\infty} F_{jikq}(\eta)\xi^q\right]U_i^{(1)}U_k^{(1)} + \ldots,$$

$$\frac{\partial U_j}{\partial \xi} = U_k,\quad \frac{\partial U_j}{\partial \eta} = U_q,\ U^{(1)}\big|_{\partial\Omega_1} = U^{(1)}\big|_{\partial\Omega_2}, \tag{5.5}$$

$$\partial\Omega_{\varepsilon k} : \sum_{q=0}^{\infty} G_{j0q}^k(\eta)\varepsilon^q + \sum_{l=1}^{r-N}\left[\sum_{q=0}^{\infty} G_{jl0q}^k(\eta)\varepsilon^q\right]U_l^{(1)}\big|_{\partial\Omega_k} +$$

$$+\sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1}\left[\sum_{q=0}^{\infty} G_{jilq}^k(\eta)\varepsilon^q\right]U_i^{(1)}\big|_{\partial\Omega_k}\,U_l^{(1)}\big|_{\partial\Omega_k} + \ldots = 0,\ k = 3, 4.$$

При этом, для $\varepsilon = 0$ получаем упрощенную краевую задачу на прямоугольнике, а при $\varepsilon = 1$ – исходную краевую задачу. Представляя все функции в виде рядов по степеням параметра вида

$$U_i = \sum_{q=0}^{\infty} U_{iq}(\eta,\xi)\varepsilon^q,\ U_i^{(1)} = \sum_{q=0}^{\infty} U_{iq}^{(1)}(\eta,\xi)\varepsilon^q,$$

$$U_{iq}^{(1)}\big|_{\partial\Omega_3} = U_{iq}^{(1)}(\eta,\xi)\big|_{\xi=-\varepsilon f_1(\eta)} = U_{iq}^{(1)}(\eta,0) -$$

$$-\frac{\partial U_{iq}^{(1)}(\eta,0)}{\partial\xi} f_1(\eta)\varepsilon + \frac{1}{2}\frac{\partial^2 U_{iq}^{(1)}(\eta,0)}{\partial\xi^2}\left(f_1(\eta)\right)^2\varepsilon^2 + \ldots \tag{5.6}$$

$$U_{iq}^{(1)}\big|_{\partial\Omega_4} = U_{iq}^{(1)}(\eta,\xi)\big|_{\xi=1+\varepsilon f_2(\eta)} = U_{iq}^{(1)}(\eta,1) +$$

$$+\frac{\partial U_{iq}^{(1)}(\eta,1)}{\partial\xi} f_2(\eta)\varepsilon + \frac{1}{2}\frac{\partial^2 U_{iq}^{(1)}(\eta,1)}{\partial\xi^2}\left(f_2(\eta)\right)^2\varepsilon^2 + \ldots,$$

получим системы для определения $U_{iq}^{(1)}$ вида

$$\sum_{q=0}^{\infty} LU_{jq}\varepsilon^q = \sum_{q=0}^{\infty} F_{j0q}(\eta)\xi^q + \sum_{k=1}^{r-N}\sum_{i=1}^{r-N-k+1}\left[\sum_{q=0}^{\infty} F_{jikq}(\eta)\xi^q\right]\left[\sum_{q=0}^{\infty} U_{iq}^{(1)}\varepsilon^q\right]\left[\sum_{q=0}^{\infty} U_{kq}^{(1)}\varepsilon^q\right] + \ldots,$$

$$\sum_{q=0}^{\infty}\frac{\partial U_{jq}}{\partial\xi}\varepsilon^q = \sum_{q=0}^{\infty} U_{kq}\varepsilon^q,\quad \sum_{l=0}^{\infty}\frac{\partial U_{jl}}{\partial\eta}\varepsilon^l = \sum_{l=0}^{\infty} U_{ql}\varepsilon^l,\ U_{iq}^{(1)}\big|_{\partial\Omega_1} = U_{iq}^{(1)}\big|_{\partial\Omega_2}, \tag{5.6}$$

$$\partial\Omega_{\varepsilon k}: \sum_{q=0}^{\infty} G^k_{j0q}(\eta)\varepsilon^q + \sum_{l=1}^{r-N}\left[\sum_{q=0}^{\infty} G^k_{jl0q}(\eta)\varepsilon^q\right]\left[\sum_{q=0}^{\infty} U^{(1)}_{lq}\Big|_{\partial\Omega_k}\varepsilon^q\right] +$$

$$+\sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1}\left[\sum_{q=0}^{\infty} G^k_{jilq}(\eta)\varepsilon^q\right]\left[\sum_{q=0}^{\infty} U^{(1)}_{iq}\Big|_{\partial\Omega_k}\varepsilon^q\right]\left[\sum_{q=0}^{\infty} U^{(1)}_{lq}\Big|_{\partial\Omega_k}\varepsilon^q\right]+\ldots=0,\ k=3,4.$$

Группируя коэффициенты при одинаковых степенях параметра, приравнивая их затем нулю, получим последовательность бесконечных, в общем случае, краевых задач для коэффициентов разложения искомых функций и вида

$$\varepsilon^0: LU_{j0} = F_{j00}(\eta) + \sum_{k=1}^{r-N}\sum_{i=1}^{r-N-k+1}\left[\sum_{q=0}^{\infty} F_{jikq}(\eta)\xi^q\right]U^{(1)}_{i0}U^{(1)}_{k0}+\ldots,$$

$$LU_{j0} = U_{k0},\ \frac{\partial U_{j0}}{\partial\eta} = U_{q0},\ U^{(1)}_{i0}(-\pi,\xi) = U^{(1)}_{i0}(\pi,\xi), \tag{5.7}$$

$$\partial\Omega_{03}: G^3_{j01}(\eta) + \sum_{l=1}^{r-N} G^3_{jl00}U^{(1)}_{l0}(\eta,0) + \sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1} G^3_{jil0}U^{(1)}_{i0}(\eta,0)U^{(1)}_{l0}(\eta,0)+\ldots=0,$$

$$\partial\Omega_{04}: G^4_{j01}(\eta) + \sum_{l=1}^{r-N} G^4_{jl00}U^{(1)}_{l0}(\eta,1) + \sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1} G^4_{jil0}U^{(1)}_{i0}(\eta,1)U^{(1)}_{l0}(\eta,1)+\ldots=0,$$

$$\varepsilon^1: LU_{j1} = F_{j01}(\eta) + \sum_{k=1}^{r-N}\sum_{i=1}^{r-N-k+1}\left[\sum_{q=0}^{\infty} F_{jikq}(\eta)\xi^q\right]\left[U^{(1)}_{i1}U^{(1)}_{k0} + U^{(1)}_{i0}U^{(1)}_{k1}\right]+\ldots,$$

$$LU_{j1} = U_{k1},\ \frac{\partial U_{j1}}{\partial\eta} = U_{q1},\ U^{(1)}_{i1}(-\pi,\xi) = U^{(1)}_{i1}(\pi,\xi), \tag{5.8}$$

$$\partial\Omega_{03}:\ G^3_{j00}(\eta) + \sum_{l=1}^{r-N}\left[G^3_{jl01}U^{(1)}_{l0}(\eta,0) + G^3_{jl00}f_1(\eta)LU^{(1)}_{l0}(\eta,0)\right]+$$

$$+\sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1}\left[G^3_{jil1}U^{(1)}_{i0}(\eta,0)U^{(1)}_{l0}(\eta,0)+\right.$$

$$\left.+G^3_{jil0}U^{(1)}_{l0}(\eta,0)f_1(\eta)LU^{(1)}_{i0}(\eta,0) + G^3_{jil0}U^{(1)}_{i0}(\eta,0)f_1(\eta)LU^{(1)}_{l0}(\eta,0)\right]+\ldots=0,$$

$$\partial\Omega_{04}:\ G^4_{j00}(\eta) + \sum_{l=1}^{r-N}\left[G^4_{jl01}U^{(1)}_{l0}(\eta,1) + G^4_{jl00}f_2(\eta)LU^{(1)}_{l0}(\eta,1)\right]+$$

$$+\sum_{l=1}^{r-N}\sum_{i=1}^{r-N-l+1}\left[G^4_{jil1}U^{(1)}_{i0}(\eta,1)U^{(1)}_{l0}(\eta,1)+\right.$$

$$\left.+G^4_{jil0}U^{(1)}_{l0}(\eta,1)f_2(\eta)LU^{(1)}_{i0}(\eta,1) + G^4_{jil0}U^{(1)}_{i0}(\eta,1)f_2(\eta)LU^{(1)}_{l0}(\eta,1)\right]+\ldots=0,$$

$$\ldots.$$

При этом граничные условия заданы на прямоугольнике $\partial\Omega_0$.

5.3. Разделение переменных для приведения последовательности краевых задач к граничным задачам для систем ОДУ.

Проведем разделение переменных интегрирования на прямоугольнике при помощи представления всех функций в виде тригонометрических рядов по

периодической переменной [17, 18]. В силу периодичности граничных условий по переменной η, представим функции в предельных краевых задачах в виде рядов Фурье по этой переменной на отрезке $[-\pi,\pi]$ вида

$$U_{jq}=\sum_{k=0}^{\infty}\left[u_{jqk}^{\cos}\cos(k\eta)+u_{jqk}^{\sin}\sin(k\eta)\right],\ f_i=\sum_{k=0}^{\infty}\left[f_{ik}^{\cos}\cos(k\eta)+f_{ik}^{\sin}\sin(k\eta)\right],$$

$$G_{j0l}^{q}=\sum_{k=0}^{\infty}\left[\left(G_{j0l}^{q}\right)_k^{\cos}\cos(k\eta)+\left(G_{j0l}^{q}\right)_k^{\sin}\sin(k\eta)\right], \tag{5.9}$$

$$G_{jilm}^{q}=\sum_{k=0}^{\infty}\left[\left(G_{jilm}^{q}\right)_k^{\cos}\cos(k\eta)+\left(G_{jilm}^{q}\right)_k^{\sin}\sin(k\eta)\right],$$

$$\cdots.$$

Произведем подстановку (5.9) в разложение (5.7), (5.8), преобразование произведения тригонометрических функций к функциям кратного аргумента и сгруппируем коэффициенты при одинаковых тригонометрических функциях. В силу независимости базиса разложения коэффициенты при каждом компоненте базиса должны быть равны нулю, что и дает необходимые ОДУ для расчета коэффициентов. Отметим, что при разделении переменных с использованием отрезков тригонометрических рядов допустимо использование также проекционных методов [18], в частности, метода Бубнова–Галеркина.

В результате последовательного применения преобразований для произведений, получаем граничную задачу для системы обыкновенных дифференциальных уравнений для линеаризованных функций u и $u^{(1)}$ вида

$$\left\{u: u_j\in u, u_j=u_{iqk}^{m}(\xi), i=\overline{1,r},\ k,q=\overline{0,\infty}, j=\overline{1,\infty}, m=\sin\cup\cos\right\},$$

$$\left\{u^{(1)}: u_j^{(1)}\in u^{(1)}, u_j^{(1)}=du_{iqk}^{(1)}\big/d\xi, i=\overline{1,r}, k=\overline{0,\infty}, j=\overline{1,\infty}, q=1\cup 2\right\}.$$

Граничная задача для системы ОДУ относительно переменной ξ на интервале $\xi\in\Xi=]0,1[$ с границей $\partial\Xi=0\cup 1$ имеет следующий общий вид

$$Lu_i+R_i(\xi,u)+N_i(\xi,u)=g_i(\xi),\ G_j(u)\big|_{\partial\Xi}=0, \tag{5.10}$$

где R_i – линейные, а N_i – нелинейные функции относительно u.

5.4. Приближенное интегрирование последовательности краевых задач для системы ОДУ.

Приближенное интегрирование граничных задач (5.10) проведем способом возмущения вида уравнений при помощи искусственного параметра. В отличие от известных асимптотических подходов [20–22], когда в качестве невозмущенной части оставляется наиболее близкая к исходным уравнениям линейная форма, произведем возмущение по искусственному параметру λ новым специальным образом: в качестве невозмущенной части оставим только старшую производную:

$$u_i=\sum_{j=0}^{\infty}u_{ij}^{M}\lambda^j,\ Lu_i=\lambda(g_i-R_i-N_i),\ G_j(u)\big|_{\partial\Xi}=0. \tag{5.11}$$

Представим R_i, N_i, и G_j в виде многомерных рядов Тейлора вида

$$R_i + N_i = \sum_{j=1}^{n}\left(\left(\sum_{r=0}^{\infty} N_{ij}^{r}\xi^{r}\right)u_j + \frac{1}{2!}\sum_{p=1}^{n}\left(\sum_{r=0}^{\infty} N_{ijp}^{r}\xi^{r}\right)u_j u_p + \ldots\right),\ g_i = \sum_{j=0}^{\infty} g_{ij}\xi^{i},$$

$$G_j = \sum_{q=1}^{n}\left(G_{jq}\left(u_q - u_q\big|_{\partial\Xi}\right) + \frac{1}{2!}\sum_{p=1}^{n} G_{jqp}\left(u_q - u_q\big|_{\partial\Xi}\right)\left(u_p - u_p\big|_{\partial\Xi}\right) + \ldots\right). \qquad (5.12)$$

Подставим теперь разложения (21) в (20), сгруппируем и приравняем нулю коэффициенты при одинаковых степенях λ. Проведем непосредственное интегрирование полученных предельных граничных задач. В случае решения задачи Коши несложно проверить, что результат имеет вид:

$$\begin{aligned} u_i = {} & \xi^{0}\lambda^{0}\,u_i\big|_{\partial\Xi} + \xi^{1}\lambda^{1}\left(g_{i0} - \sum_{r=1}^{n}\left(N_{ir}^{0}u_r\big|_{\partial\Xi} + \frac{1}{2}\sum_{p=1}^{n} N_{irp}^{0}\,u_r\big|_{\partial\Xi}\,u_p\big|_{\partial\Xi} + \ldots\right)\right) + \\ & + \xi^{2}\left(\left[\frac{g_{i1}}{2} - \frac{1}{2}\sum_{r=1}^{n}\left(N_{ir}^{1}\,u_r\big|_{\partial\Xi} + \frac{1}{2}\sum_{p=1}^{n} N_{irp}^{1}\,u_r\big|_{\partial\Xi}\,u_p\big|_{\partial\Xi} + \ldots\right)\right]\lambda^{1} + \right. \\ & + \left[-\sum_{r=1}^{n}\left(N_{ir}^{0}\left(\frac{g_{r0}}{2} - \frac{1}{2}\sum_{l=1}^{n}\left(N_{rl}^{0}\,u_l\big|_{\partial\Xi} + \frac{1}{2}\sum_{q=1}^{n} N_{rlq}^{0}\,u_l\big|_{\partial\Xi}\,u_q\big|_{\partial\Xi} + \ldots\right)\right) + \right.\right. \\ & + \frac{1}{2}\sum_{p=1}^{n} N_{irp}^{0}\left(u_p\big|_{\partial\Xi}\left(\frac{g_{r0}}{2} - \frac{1}{2}\sum_{l=1}^{n}\left(N_{rl}^{0}\,u_l\big|_{\partial\Xi} + \frac{1}{2}\sum_{q=1}^{n} N_{rlq}^{0}\,u_l\big|_{\partial\Xi}\,u_q\big|_{\partial\Xi} + \ldots\right)\right)\right) + \\ & \left.\left.\left. + u_r\big|_{\partial\Xi}\left(\frac{g_{p0}}{2} - \frac{1}{2}\sum_{l=1}^{n}\left(N_{pl}^{0}\,u_l\big|_{\partial\Xi} + \frac{1}{2}\sum_{q=1}^{n} N_{plq}^{0}\,u_l\big|_{\partial\Xi}\,u_q\big|_{\partial\Xi} + \ldots\right)\right)\right)\right)\right)\right]\lambda^{2}\right) + \ldots . \end{aligned} \qquad (5.13)$$

Таким образом, результатом специального введения искусственного параметра является решение в виде двумерного асимптотического степенного ряда по параметру и переменной интегрирования. Кроме того, из вида полученного решения непосредственно вытекает его устойчивость для задачи Коши – первые члены ряда по степеням переменной определяются окончательно решениями первых предельных граничных задач.

5.5. Выбор схемы усечения рядов и суммирование их двумерными дробно-рациональными аппроксимантами.

Полученное приближенное асимптотическое решение граничной задачи (5.13) является формальным. Для непосредственного применения приближения в технических задачах необходимо определить диапазон его сходимости по параметру возмущения и переменной интегрирования, или произвести суммирование. При этом, в отличие от естественного малого параметра, для искусственного параметра должна быть обеспечена сходимость вплоть до единицы, что не всегда достижимо. Эта проблема решается, в основном, методами продолжения по параметру или обобщенного суммирования [21], наиболее перспективными и естественными из которых являются аппроксимации Паде [10]. Использование указанных методов производилось ранее только для продолжения по параметру возмущения, т. е. для одномерного случая (1D), и на основе преимущественно эмпирических предположений.

Переменная интегрирования при этом рассматривалась как параметр, интервал сходимости по ней рассчитывался апостериорно. Применение двумерного (2D) суммирования практически не производилось, в частности, из-за множественности выбора коэффициентов ряда и вытекающей из этого проблемы существования и единственности обобщенной суммы. Хорошо известно, что это фундаментальная проблема для кратных рядов и ее решение требует использования результатов специальных исследований в области многомерных функций [23]. Только в последние годы математиками школы А.Н. Колмогорова под руководством В.В. Вавилова была обоснована специальная схема преобразования отрезков кратного ряда Тейлора в двумерные дробно-рациональные функции Паде–типа, обеспечивающая существование и единственность аппроксимант с заданной структурой [14].

Пусть для комплексных переменных z_1, z_2 в окрестности начала координат задана голоморфная функция $F(z_1,z_2)=\sum_{i=0}^{\infty} f_{ij} z_1^i z_2^j$. Пусть для любых целых множеств $n=(n_1,n_2)$ и $m=(m_1,m_2)$, т. е. для любых $n,m\in\mathbb{Z}_+^2$, задан класс рациональных функций – отношение двумерных полиномов степени не выше $n=(n_1,n_2)$ и $m=(m_1,m_2)$ соответственно по каждой переменной – вида

$$R(n,m)=\left\{ r=\frac{p}{q}, p=\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} p_{ij} z_1^i z_2^j, q=\sum_{i=0}^{m_1}\sum_{j=0}^{m_2} q_{ij} z_1^i z_2^j, q_{00}\equiv 1, \deg(p)\le n, \deg(q)\le m \right\}.$$

Каждая рациональная функция $r\in R(n,m)$ может быть задана своим сходящимся в некоторой окрестности начала координат рядом Маклорена. Функция r зависит от $\tau_{nm}=(n_1+1)(n_2+1)+(m_1+1)(m_2+1)-1$ параметров (коэффициентов p и q). Множество целых точек $I=I(n,m)\subset\mathbb{Z}_+^2$ называется определяющим множеством, если для фиксированных n и m

$$\dim I=\tau_{nm},$$

$$\{(n_1+m_1,0),(0,n_2+m_2)\}\subset I,$$

$$[0,k]=\{k\in I,(s_1,s_2)\in\mathbb{Z}_+^2: 0\le s_j\le k_j, j=1,2\}\subset I,$$

$$\{(n_1+m_1,m_2),(m_1,n_2+m_2)\}\cap I\ne\varnothing,\ n\in I.$$

Существует два и только два варианта определяющих множеств I_1 и I_2, удовлетворяющих всем требованиям, с компонентами (i,j) вида

$$I_1:$$

$$\{[0\le i\le n_1, 0\le j\le n_2]\cup[n_1+1\le i\le n_1+m_1, 0\le j\le m_2]\cup[i=0, n_2+1\le j\le n_2+m_2]\},$$

$$I_2:$$

$$\{[0\le i\le n_1, 0\le j\le n_2]\cup[0\le i\le m_1, n_2+1\le j\le n_2+m_2]\cup[n_1+1\le i\le n_1+m_1, j=0]\}.$$

Обобщенная аппроксиманта Паде для заданных значений n и m определяется как рациональная функция F_{nm}, для которой коэффициенты ряда Маклорена разности F_{nm} с F равны нулю на I:

$$T_{ij}\left(F-F_{nm}\right)=0,\ (i,j)\in I.$$

Пусть для фиксированного значения $m\in\mathbb{Z}_+^2$ определен класс функций

$$M_m=M_m\left(\mathbb{C}^2\right)=\left\{F:F\left(z_1,z_2\right)=P\left(z_1,z_2\right)/Q_m\left(z_1,z_2\right)\right\}$$

со следующими свойствами: $P\left(z_1,z_2\right)$ – полная функция; $\deg Q_m=m$, $\deg Q_m\left(z_1,0\right)=m_1, \deg Q_m\left(0,z_2\right)=m_2$; $Q_m\left(0,0\right)=1$; функции $P\left(z_1,0\right), P\left(0,z_2\right)$ и полиномы $Q_m\left(z_1,0\right), Q_m\left(0,z_2\right)$ не равны нулю одновременно.

Сформулируем условия корректного применения двумерного дробно-рационального преобразования.

Пусть $F\left(z_1,z_2\right)\in M_m$ представлена рядом Маклорена, $m\in\mathbb{Z}_+^2$ фиксировано и $n\in\mathbb{Z}_+^2$. Тогда для всех достаточно больших $n'=\min\left(n_1,n_2\right)$ существует единственная обобщенная аппроксиманта Паде $F_{nm}=P_n/q_n$ на каждом из определяющих множеств $I_j, j=1,2$, и последовательность F_{nm} при $n'=\min\left(n_1,n_2\right)\to\infty$ равномерно сходится к F по всем компактным подмножествам $G=\mathbb{C}^2\setminus\left\{Q_m=0\right\}$. Для любого компакта $E\subset\mathbb{C}^2$ верны оценки сходимости по мере $\left\|f\left(z\right)\right\|_E=\sup\limits_{z\in E}\left|f\left(z\right)\right|$ вида

$$\lim_{n'\to\infty}\left\|Q_m-q_n\right\|_E^{1/n'}=0,\ \lim_{n'\to\infty}\left\|F-F_{nm}\right\|_E^{1/n'}=0.$$

В предлагаемом методе за счет специальной схемы введения искусственного параметра и, как следствие, полиномиальной структуры приближенного решения асимптотического решения граничной задачи (5.13) впервые появляется возможность использовать приведенный результат для определения порядка усечения двумерного ряда, его обобщенного суммирования и оценки скорости сходимости. Приближенное решение (5.13) преобразуется по приведенной схеме к дробно-рациональной аппроксиманте вида

$$PA_{nm}\left(u_i\right)=\sum_{j=0}^{n_1}\sum_{k=0}^{n_2}p_{ijk}\xi^j\lambda^k\Bigg/\sum_{j=0}^{m_1}\sum_{k=0}^{m_2}q_{ijk}\xi^j\lambda^k,\ q_{i00}\equiv 1. \qquad (5.14)$$

Согласно условиям корректного применения двумерного дробно-рационального преобразования, оно осуществляет мероморфное продолжение двумерного асимптотического ряда. Границей изменения параметров ряда является расстояние до ближайшей существенно особой точки (а не ближайшего полюса как у традиционных приближений) аппроксимируемой функции, что значительно расширяет диапазон пригодности модели.

5.6. Сходимости построенного приближенного решения

В работах ряда авторов, в частности Я. Ф. Каюка [21], А. Н. Гузя и Ю. Н. Немиша [22], приведен общий порядок нахождения области сходимости разложения по параметру и обосновано успешное применение метода возмущения формы границы в некоторых отдельных задачах механики сплошных сред. При этом отмечено [21], что доказать сходимость приближенного решения таких задач при произвольном выборе области

интегрирования, вида системы уравнений и граничных условий не представляется возможным.

В нашем подходе возможна иная постановка задачи обоснования построенного метода приближенного решения задачи (3.6) – (3.8). Будем опирается на гипотезу о пригодности асимптотического разложения в методе возмущения формы границы. Эта гипотеза основана на традиции, восходящей к Пуанкаре [21], она заключается в предположении существования точного аналитического решения возмущенной системы для некоторых малых отличных от нуля значений параметра возмущения и сходимости к нему разложения в методе возмущения формы границы. Поскольку аналитическое решение на некоторой части области определения параметра возмущения позволяет построить его аналитическое продолжение на всю область определения, принятие гипотезы является логичным основанием искать решение в виде обобщенной суммы полученных рядов.

Лемма 5.1. Для того, чтобы (5.13) при $\lambda = 1$ было рядом Маклорена точного решения граничной задачи (5.9), голоморфного в некоторой окрестности $\xi = 0$, необходимо и достаточно, чтобы ряды (5.10) сходились равномерно на замыкании $\Xi \bigcup \partial\Xi$.

Доказательство необходимости условий леммы вытекает из способа получения (5.13). Действительно, легко проверить, что, если представить искомые функции в виде рядов Маклорена, подставить их в (5.9), перегруппировать и приравнять коэффициенты при одинаковых степенях переменной, то получатся выражения для коэффициентов (5.13).

Достаточность приведенных условий непосредственно следует из единственности степенного разложения функций в области голоморфности. Требование равномерности для оценки сходимости разложения необходимо для изменения порядка суммирования при бесконечном числе членов.

Лемма 5.1 заимствована из нашей работы [17], в которой изложение более детально и рассмотрены применения к уравнениям выше первого порядка.

Теорема 5.1. Пусть решение (точное) краевой задачи (5.10) существует и единственно. Тогда в условиях леммы 5.1 применение к асимптотическому ряду (5.13) усечения и дробно-рационального преобразования в двумерные Паде–аппроксиманты описанным в разделе 5.5 методом определяет корректное обобщенное суммирование этого ряда, которое сохраняет смысл во всей области мероморфности точного решения.

Доказательство непосредственно вытекает из леммы 5.1 и приведенных условий корректного применения двумерного дробно-рационального преобразования.

Следствие 5.1 (достаточное условие сходимости метода) Для того, чтобы дробно-рациональные функции (5.14) определяли асимптотическую мероморфную модель задачи (3.6) – (3.8) в области, близкой к прямоугольнику, необходимо, чтобы точное аналитическое решение этой задачи существовало и допускало сходящееся разложение (5.6) с коэффициентами, удовлетворяющими методу возмущения формы границы (5.7).

Действительно, если сходятся разложения аналитических функций (5.14) и соответствующая им последовательность предельных краевых задач метода

возмущения формы границы для криволинейной трапеции, то уравнения (5.9) для амплитуд их гармонического разложения имеет, согласно утверждениям 1 и 2, мероморфное приближение в виде (5.14). Ограничение на аналитичность решения связано с характером исследуемых задач и, как правило, вытекает из физической природы исследуемых систем. В то же время, наличие полюсов аппроксимации, моделирующих положение особенностей точного решения, позволяет рассчитывать критические значения параметров модели – положение точек бифуркации, предельных точек, нарушение условий устойчивости процесса и т. д.

Несмотря на то, условие следствия 5.1 не вполне конструктивно, в задачах механики замкнутых торсовых оболочек, которые исследованы в [12, 13, 15], оно выполняется. Больше того, там оно может быть заменено непосредственно доказательством существования и аналитичности решения возмущенной задачи. В тех задачах, где условие следствия не выполняется, всё еще остаётся возможность установить практическую сходимость модели в смысле Пуанкаре путем оценки нормы последовательных приближений [21].

6. Выводы по результатам и направления дальнейших исследований

В статье автором разработан комплексный вычислительный метод для расчета возмущённых периодических краевых задач для систем уравнений в частных производных, включающего известные задачи теории упругих гибких торсовых оболочек, который обеспечивает сходимость решений и получение приемлемого для приложений результата при наиболее широких условиях, допустимых в рамках ассимптотического подхода к использованию рядов с параметрами возмущения. Наряду с разработкой общей схемы, в которую укладываются частные случаи [12, 13, 1], впервые дано условие успешного применения нашего варианта асимптотического подхода в моделировании деформированных состояний замкнутых торсовых оболочек, выделяющее определённый класс задач с наличием, вообще говоря, полиномиальных нелинейностей в уравнениях для областей, которые можно получить деформацией из прямоугольника. Даётся схема вычислительного метода, применимая при выполнении данного условия, и доказывается для этого случая условная сходимость асимптотических рядов. Это позволяет осуществить в данной схеме обоснованное усечений двумерных асимптотических рядов с оценкой точности, причем, не с апелляцией к физическим соображениям, а на основе найденного адекватного порядка их обобщенного суммирования. Накопленные к настоящему времени результаты численных экспериментов наглядно подтверждают [15, 16], что в исследованных задачах использование обобщенного суммирования двумерного асимптотического ряда по предложенному методу действительно приводят существенному повышению точности численного моделирования.

Прикладная значимость этих результатов заключена в их существенном отличии от ранее использовавшихся вариантов асимптотического подхода для построения решений краевых задач механики оболочечных и других строительных конструкций в двумерных возмущённых областях. А именно, в предложенном методе возможна оценка погрешности усечения ряда и

увеличение интервала изменения параметра возмущения, на котором получаются достоверные численные результаты, вплоть до ближайшей существенно особой точки точного решения.

Общность схемы применения метода позволяет получать новые математические модели и позитивные результаты их испытаний для возмущенных краевых задач в различных прикладных областях. Это – естественное направление развития результатов статьи на дальнейшее.

ЛИТЕРАТУРА

1. Олевский, В. И. Асимптотический метод моделирования технических систем с технологическими отклонениями [Текст] / В. И. Олевский // Восточно-Европейский журнал передовых технологий. – 4/11(70). – 2014. – С. 25–31.
2. Teng, J. G. Buckling of thin shells: Recent advances and trends [Text] / J. G. Teng // Appl. Mech. Rev. – 1996. – Vol. 49, Issue 4. – pp. 263–274.
3. Teng, J. G. Buckling of Thin Metal Shells [Text] / J. G. Teng and , J. M. Rotter. – London and New York: Spon Press, 2006. – 520 p.
4. Григолюк, Э. И. Проблемы нелинейного деформирования: Метод продолжения решения по параметру в нелинейных задачах механики деформируемого твердого тела [Текст] / Э. И. Григолюк, В. И. Шалашилин. – М.: Наука, Гл. ред. физ.-мат. лит., 1988. – 232 с.
5. Численные методы в механике [Текст] / В.А. Баженов, А.Ф Дащенко, Л.В. Коломиец и др. – Одесса, Стандартъ. – 2005. – 563 с.
6. Дородницын, А. А. Применение метода малого параметра к численному решению дифференциальных уравнений [Текст] / А. А. Дородницын. – Соврем. проблемы матем. физики и вычисл. матем. – М.: Наука, 1982. – С. 145–155.
7. Adomian, G. A review of the decomposition method and some recent results for nonlinear equations [Text] / G. Adomian // Comp. Math. Appl. – 1989. – Vol. 21. – pp. 101–127.
8. He, J. H. Recent developments of the homotopy perturbation method [Text] / J. H. He // Top. Meth. Nonlin. Anal. – 2008. – Vol.31. – pp. 205-209.
9. Андрианов, И.В. Расчет нелинейного деформирования оболочек с развертывающейся срединной поверхностью приближенными аналитическими методами [Текст] / И.В. Андрианов, А.М. Мильцын, В.И. Олевский, В.В. Плетин // Східно-Європейський журнал передових технологій. – Харьков, 2010. - № 3/9 (45). – С. 27-34.
10. Андрианов, И. В. Применение метода Паде-аппроксимант для устранения неоднородностей асимптотических разложений [Текст] / И. В. Андрианов // Изв. АН СССР. Механика жидкости и газа. – 1984. – № 3. – С. 166–167.
11. Basto, M. Numerical study of modified Adomian's method applied to Burgers equation / M. Basto, V. Semiao, F. L. Calheiros // Journal of Computational and Applied Mathematics. – 2007. – Vol. 206(2). – pp. 927-949.
12. Andrianov, I. V. Analytical perturbation method for calculation of shells based on 2-D Padé approximants [Text] / I. V. Andrianov, V. I. Olevs'kyy, J. Avrejcevwicz // Int. J. Str. Stab. Dyn. – 13 (7). – 2013. – pp. 1340003-1–1340003-7.

13. Andrianov, I. V. Application of 2-D Padé approximants in nonlinear shell theory: Stability calculation and experimental justification [Text] / I. V. Andrianov, V. I. Olevs'kyy, J. Avrejcevwicz // Nunliarity, bifurcation and chaos – Theory and applications. – Rijeca (Croatia): InTech. – 2012. – pp. 1–26.
14. Vavilov, V. V. Design of multidimensional Recursive Systems through PadéType Rational Approximation [Text] / V. V. Vavilov, M. K. Tchobanou, P. M. Tchobanou // Nonlinear Analysis: Modelling and Control. – 2002. – Vol. 7, Issue 1. – P. 105–125.
15. Олевский, В. И. Математическое моделирование оболочечных конструкций с отклонениями [Текст] / В. И. Олевский. – Днепропетровск, изд-во Маковецкий. – 2014. – 382 с.
16. Андрианов, И. В. Модифицированный метод декомпозиции Адомяна [Текст] / И. В. Андрианов, В. И. Олевский, С. Токажевский // ПММ. – 1998. – Т. 62, №.2. – С. 334-339.
17. Полянин, А. Д. Методы решения нелинейных уравнений математической физики и механики [Текст] / А. Д. Полянин, В. Ф. Зайцев, А. И. Журов. – М.: Физматлит. – 2005. – 256 с.
18. Власова, Б. А. Приближенные методы математической физики [Текст]: Учеб. для вузов / Б. А. Власова, В. С. Зарубин, Г. Н. Кувыркин // Под. ред. В. С. Зарубина, А. П. Крищенко. – М.: Изд-во МГТУ им. Н.Э. Баумана. – 2001. – 700 с.
19. Andrianov, I. V. Approximate non-linear boundary value problems of reinforced shell dynamics [Text] / I. V. Andrianov, E. G. Kholod, V. I. Olevsky // J. Sound Vibr. – 1996. – Vo l. 194, Issue3. – P. 369 – 387.
20. Andrianov, I. V. Asymptotic Approaches in Mechanics: New Parameters and Procedures [Text] / I. V. Andrianov, J. Awrejcewicz, R. G Barantsev // Appl. Mech. Rev. – 2003. – Vol. 56, Issue 1. – P. 87–110.
21. Каюк, Я. Ф. некоторые вопросы методов разложения по параметру [Текст] // Я. Ф. Каюк. – Киев: Наук. думка. – 1980. – 168 с.
22. Гузь, А. Н. Метод возмущения формы границы в механике сплошных сред: Учебное пособие для студ. ун-тов и втузов [Текст] // А. Н. Гузь, Ю. Н. Немиш. – Киев : Выща школа, 1989 . – 352 с.
23. Олевская, Ю. Б. О соотношении некоторых систем суммирования кратных рядов Фурье [Текст] / Ю. Б. Олевская, В. И. Олевский // Шляхи сучасної математики:освіта, наука, індустрія: матеріали конф. (18 квіт., м. Дніпропетровськ). –Д.: Національний гірничий університет. – 2013. – с. 26–32.

Надійшла у першій редакції 23.10.2014, в останній - 03.11.2014.

УДК 517.922+519.642.2

Комбинированный численный метод решения вырожденного нелинейного интегро-дифференциального уравнения с запаздываниями

А. Л. Пивень

Харьковский национальный университет имени В.Н. Каразина, Украина

Разработан численный метод решения нелинейного вырожденного интегро-дифференциального уравнения запаздывающего типа, которое возникает при описании переходных процессов в радиотехнических системах. Матричный коэффициент при производной может быть необратимым. Для дискретизации этого уравнения используются метод спектральных проекторов типа Рисса, явная схема Эйлера и квадратурная формула левых прямоугольников. Полученное при дискретизации разностное уравнение решается с помощью метода простых итераций. Доказана теорема сходимости разработанного численного метода.

Ключевые слова: *вырожденное интегро-дифференциальное уравнение с запаздываниями, численное решение,комбинированный метод, сходимость*

Розроблено чисельний метод розв'язання нелінійного виродженого інтегро-диференціального рівняння запізнюючого типу, що виникає при опису перехідних процесів у радіотехнічних системах. Матричний коефіцієнт при похідній може бути необерненим. Для дискретизації цього рівняння використовуються метод спектральних проекторів типу Риса, явна схема Ейлера та квадратурна формула лівих прямокутників. Отримане при дискретизації різницеве рівняння розв'язується за допомоги методу простих ітерацій. Доведено теорему про збіжність розробленого чисельного методу.

Ключові слова: *вироджене інтегро-диференціальне рівняння із запізненнями, комбінований метод, чисельний розв'язок, збіжність.*

We have developed a numerical method for solving nonlinear degenerated integro-differential equation with delay, which appears when transients in radio technical systems are described. The matrix coefficient of derivative may be non-invertible. To discretize this equation, the method of Rietz type spectral projectors, explicit Euler scheme and left rectangles quadrature are used. The difference equation obtained during discretization is solved using simple iteration method. Convergence of developed numerical method is proved as a theorem.

Key words: *degenerate delay integro-differential equation, numeric solution, combined method, convergence.*

1. Введение

В данной работе предложен численный метод решения начальной задачи

$$\frac{d}{dt}(A_0u(t)) + \sum_{j=0}^{N} B_j u(t-\omega_j) + \sum_{j=0}^{N} \int_{t_0-\omega_j}^{t-\omega_j} \Phi_j(t,\tau)u(\tau)d\tau = f(t,u(t)), \quad t_0 \le t \le T \quad (1.1)$$

$$u(t) = g(t), \qquad t_0 - \omega_N \le t \le t_0 \quad (1.2)$$

Здесь A_0, B_j $(j = 0,...,N)$ – постоянные квадратные матрицы порядка n с вещественными элементами, элементы $n \times n$ матриц $\Phi_j(t,\tau)$ непрерывны по

совокупности переменных на множествах $\{(t,\tau)\in[t_0,T]\times[t_0-\omega_j,T]: t_0-\omega_j \le \le \tau \le t-\omega_j\}$ соответственно, $f(t,x)\in C([t_0,T]\times R^n,R^n)$, $g(t)\in C([t_0-\omega_N,t_0],R^n)$. Запаздывания упорядочены : $0=\omega_0<\omega_1<...<\omega_N$. Следуя [1,2], под решением задачи (1.1),(1.2) на отрезке $[t_0-\omega_N,T]$ будем понимать вектор-функцию $u(t)\in C([t_0-\omega_N,T],R^n)$ такую, что $A_0u(t)\in C^1([t_0,T],R^n)$, $u(t)$ удовлетворяет уравнению (1.1) при $t\in[t_0,T]$ и начальному условию (1.2).

Уравнение (1.1) называется *неявным*, а в случае необратимости матрицы A_0 – *вырожденным* [1,3]. Нелинейное уравнение (1.1) с вырожденной матрицей A_0 получено в [4,5] при описании переходных процессов в радиотехнических системах.

Различные численные методы решения частного случая (1.1) – вырожденных нелинейных интегро-дифференциальных уравнений Вольтерра

$$\frac{d}{dt}(A_0u(t))+B_0u(t)+\int_{t_0}^{t}\Phi_0(t,s,u(s))ds=f(t,u(t)),\quad t_0\le t\le T. \tag{1.3}$$

строились в [10–12] с путем замены производной в (1.3) конечной разностью и интегрального слагаемого (1.3) – квадратурной суммой. В [6] строилось численное решение дифференциально-алгебраического уравнения с запаздываниями индекса 1 вида $\frac{dx}{dt}=f(x(t),x(t-1),y(t),y(t-1));$ $y(t)=g(x(t),x(t-1),y(t))$ методом Рунге-Кутта и методом BDF. Построенные таким образом численные методы часто приводят к нелинейным разностным уравнениям, для которых не всегда удается найти точное решение. В [8,9] для системы полулинейных дифференциально-алгебраических уравнений индекса 1 вида

$$\frac{dx}{dt}=f(t,x(t),y(t)); \tag{1.4}$$

$$y(t)=g(t,x(t),y(t)) \tag{1.5}$$

предлагались так называемые *комбинированные* численные методы, в которых замены производной конечной разностью в (1.4) приводила к системе разностных нелинейных уравнений, для численного решения которой, в свою очередь, применялись методы итераций или метод Ньютона. Именно, в [8] численное решение системы уравнений (1.4),(1.5) строилось на основе комбинаций разностных схем Эйлера и простых итераций и комбинаций схем Адамса и Ньютона, а в [9] для этой же системы рассмотрены комбинации методов Рунге-Кутта с методами простых итераций и Ньютона.

Как и в [8] , в настоящей работе для построения численного метода решения задачи (1.1),(1.2) с характеристическим пучком λA_0+B_0 индекса 1 используются комбинации явной схемы Эйлера и простых итераций, а замена интегральные слагаемые в (1.1) заменяются по формуле левых прямоугольников, как в [12]. Предварительно применяется метод спектральных проекторов типа Рисса [13] для расщепления уравнения на дифференциальное и

алгебраическое. Расщепленные уравнения превращаются в систему вида (1.4),(1.5) в случае отсутствия в (1.1) интегральных слагаемых и слагаемых, отвечающих запаздываниям. В отличие от [11,12] для доказательства сходимости вычислительного процесса не требуется знание точного решения ни в одной точке $t > t_0$.

2. Теорема существования и единственности решения

Линейной части уравнения (1.1) отвечает характеристический матричный пучок $\lambda A_0 + B_0$. Всюду в дальнейшем он предполагается регулярным $(\det(\lambda A_0 + B_0) \not\equiv 0)$ [14] индекса 0 или 1. Введем пару взаимно дополнительных спектральных проекторов [13]

$$P_1 = \frac{1}{2\pi i} \oint_{|\lambda|=C_0} (\lambda A_0 + B_0)^{-1} A_0 d\lambda, \qquad P_2 = E - P_1;$$

$$Q_1 = \frac{1}{2\pi i} \oint_{|\lambda|=C_0} A_0 (\lambda A_0 + B_0)^{-1} d\lambda, \quad Q_2 = E - Q_1, \tag{2.1}$$

где контур $\{\lambda : |\lambda| = C_0\}$ охватывает конечный спектр пучка $\lambda A_0 + B_0$, а E – единичная матрица порядка n. По аналогии с оператором G из [3] введем матрицу

$$G = A_0 + B_0 P_2 = A_0 + Q_2 B_0 .$$

Как и оператор G из [3], матрица G обратима. Для проекторов вида (2.1) свойства матрицы G установлены в [1].

Рассмотрим множество D, содержащее точку T и все точки отрезка $[t_0, T]$, представимые в виде $t_0 + \sum_{j=0}^{N} l_j \omega_j$, где $l_j (j = 0,...,N)$ –целые неотрицательные числа. В следующей теореме устанавливаются условия существования и единственности решения $u(t) \in \in C^1([t_0 - \omega_N, T] \setminus D, R^n)$ задачи (1.1),(1.2).

Теорема 1. *Пусть пучок матриц* $\lambda A_0 + B_0$ *регулярный индекса не выше 1,* $f(t,x) \in C^1([t_0,T] \times R^n, R^n)$, , $g(t) \in C^1([t_0 - \omega_N, t_0], R^n)$, *элементы* $n \times n$ *матриц* $\Phi_j(t,\tau)$ *вещественны и непрерывно дифференцируемы по совокупности переменных на множествах* $\{(t,\tau) \in [t_0,T] \times [t_0 - \omega_j, T]: \quad t_0 - \omega_j \le \tau \le t - \omega_j\}$ $(j = 0,...,N)$ *соответственно и при некоторой постоянной* $M > 0$ *выполнено условие Липшица*

$$\| G^{-1} Q_1 f(t,x) - G^{-1} Q_1 f(t,y) \| \le M \| x - y \|, \qquad x, y \in R^n, t \in [t_0, T]. \tag{2.2}$$

Предположим, что при некоторой постоянной $q \in (0,1)$ *матрица Якоби* $J(t,x) = \{\partial (G^{-1} Q_2 f)_j / \partial x_k\}$ *удовлетворяет ограничению*

$$\| J(t,x) \| \le q, \qquad (t,x) \in [t_0,T] \times R^n \tag{2.3}$$

и выполнено условие согласования

$$Q_2 \sum_{j=0}^{N} B_j g(t_0 - \omega_j) = Q_2 f(t_0, g(t_0)) \tag{2.4}$$

на начальный вектор в (1.2) и правую часть в (1.1). Тогда существует единственное решение начальной задачи (1.1),(1.2) на $[t_0 - \omega_N, T]$, *которое непрерывно-дифференцируемо, возможно, за исключением точек множества* D, *в которых производная может иметь скачки. При этом* $P_1 u(t) \in C^1([t_0, T], R^n) \cap C^2([t_0, T] \setminus D, R^n)$

Доказательство теоремы проводится по схеме доказательства теоремы 2.8 работы [16], выполненной для нелинейного уравнения $\frac{d}{dt}(A_0 u(t)) + B_0 u(t) + B_1 u(t - \omega_1) = f(t, u(t))$.

Замечание. В [1,15,16] получены локальные и глобальные теоремы существования и единственности решения задачи (1.1),(1.2) в бесконечномерных пространствах в случае $\Phi_j(t, \tau) \equiv 0 \, (j = 0, ..., N)$.

3. Построение численного решения задачи (1.1),(1.2)

Всюду далее предполагаем выполнение условий теоремы 1. Как и в [1,3] применим к уравнению (1.1) проекторы Q_1, Q_2. Получим эквивалентную систему дифференциально-алгебраических уравнений

$$\frac{d}{dt}(x(t)) + Sx(t) + G^{-1} Q_1 \sum_{j=1}^{N} B_j u(t - \omega_j) +$$

$$+ G^{-1} Q_1 \sum_{j=0}^{N} \int_{t_0 - \omega_j}^{t - \omega_j} \Phi_j(t, \tau) u(\tau) d\tau = G^{-1} Q_1 f(t, u(t)); \tag{3.1}$$

$$y(t) = G^{-1} Q_2 f(t, u(t)) - G^{-1} Q_2 \sum_{j=1}^{N} B_j u(t - \omega_j) - G^{-1} Q_2 \sum_{j=0}^{N} \int_{t_0 - \omega_j}^{t - \omega_j} \Phi_j(t, \tau) u(\tau) d\tau, \tag{3.2}$$

где $x(t) = P_1 u(t), \; y(t) = P_2 u(t)$, $S = G^{-1} Q_1 B_0$.

Рассмотрим сеточную область $\{t_i = t_0 + ih : i = -m_N, ..., K\}$, где шаг сетки $h = \frac{T - t_0}{K}$ выбирается сколь угодно малым и таким, что величины $m_k = \frac{\omega_k}{h} \, (k = 0, ..., N)$ являются целыми. Обозначим через u_i значение приближенного решения задачи (1.1),(1.2) в узле сетки $t_i \, (i = -m_N, .., K)$ и через $x_i = P_1 u_i, y_i = P_2 u_i$ его проекции: $u_i = x_i + y_i$. Заменим производную в уравнении (3.1) конечной разностью согласно явной схеме Эйлера, а интегральные слагаемые в (3.1),(3.2) – квадратурной формулой левых прямоугольников, как в [12]. Получаем следующую разностную схему:

$$x_{i+1} = (E - Sh)x_i - h\sum_{j=1}^{N} G^{-1}Q_1 B_j u_{i-m_j} -$$

$$- h^2 \sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j-1} G^{-1}Q_1 \Phi_j(t_i, t_k)u_k + hG^{-1}Q_1 f(t_i, x_i + y_i); \qquad (3.3)$$

$$y_{i+1} = G^{-1}Q_2 f(t_{i+1}, x_{i+1} + y_{i+1}) - \sum_{j=1}^{N} G^{-1}Q_2 B_j u_{i+1-m_j} -$$

$$- h\sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j} G^{-1}Q_2 \Phi_j(t_{i+1}, t_k)u_k ; \qquad (3.4)$$

$$u_{i+1} = x_{i+1} + y_{i+1}, \qquad i = 0,..., K-1, \qquad (3.5)$$

В силу принципа сжимающих отображений [20, с.44] уравнение (3.4) относительно неизвестного y_{i+1} однозначно разрешимо, если выполнено условие (2.3). Разностные уравнения (3.3)–(3.5) решаются при известных начальных условиях:

$$x_i = P_1 g(t_0 + ih),\ y_i = P_2 g(t_0 + ih),\ u_i = x_i + y_i,\ i = -m_N, -m_N + 1,...,0. \qquad (3.6)$$

Однако процесс построения численного решения задачи (1.1),(1.2) по разностной схеме (3.3) – (3.6) вызывает трудности, связанные с необходимостью решения уравнения (3.4) относительно y_{i+1}. Уравнение (3.4) возникает только в случае необратимости матрицы A_0. В явном виде y_{i+1} легко находится из (3.4), если проекция $Q_2 f(t,x)$ не зависит от x. Такое ограничение было использовано в [2,15] при доказательстве теорем существования и единственности решения, а также в [17] при построении и доказательстве сходимости численного метода (3.3)–(3.6) решения задачи (1.1),(1.2). В следующей теореме указываются достаточные условия сходимости вычислительного метода (3.3)–(3.6) в условиях теоремы 1 без предположения о независимости проекции $Q_2 f(t,x)$ от x.

Теорема 2. *Пусть выполнены условия теоремы 1. Тогда разностная схема (3.3)–(3.6) имеет первый порядок сходимости :*

$$\max_{i=0,...,K} \| u(t_i) - u_i \| = O(h).$$

Доказательство. Пусть $u(t) \in C^1([t_0 - \omega_N, T] \setminus D, R^n)$ – точное решение задачи (1.1),(1.2), которое существует в силу теоремы 1. Так как $x(t) = P_1 u(t) \in C^1([t_0, T], R^n) \cap C^2([t_0, T] \setminus D, R^n)$, то в силу формулы Тейлора и квадратурной формулы левых прямоугольников [18, с. 164]

$$x'(t_i) = \frac{x(t_{i+1}) - x(t_i)}{h} + O(h), \quad i = 0,..., K-1; \qquad (3.7)$$

$$\int_{t_0-\omega_j}^{t_i-\omega_j} \Phi_j(t_i, \tau)u(\tau)d\tau = h \sum_{k=-m_j}^{i-m_j-1} \Phi_j(t_i, t_k)u(t_k) + O(h), \quad i = 0,..., K,\ j = 0,..., N. \qquad (3.8)$$

При $i=0$ правая часть уравнения (3.8) равна $O(h)$. Положим в уравнении (3.1) $t=t_i$, в уравнении (3.2) $t=t_{i+1}$ и воспользуемся формулами (3.7),(3.8). Получим

$$x(t_{i+1})=(E-hS)x(t_i)-h\sum_{j=1}^{N}G^{-1}Q_1B_ju(t_{i-m_j})-$$

$$-h^2\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j-1}G^{-1}Q_1\Phi_j(t_i,t_k)u(t_k)+hG^{-1}Q_1f(t_i,u(t_i))+O(h^2),\ i=0,...,K-1; \tag{3.9}$$

$$y(t_{i+1})=G^{-1}Q_2f(t_{i+1},u(t_{i+1}))-\sum_{j=1}^{N}G^{-1}Q_2B_ju(t_{i+1-m_j})-$$

$$-h\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j}G^{-1}Q_2\Phi_j(t_{i+1},t_k)u(t_k)+O(h),\ i=0,...,K-1. \tag{3.10}$$

Обозначим $\varepsilon_i^x=\|x_i-x(t_i)\|, \varepsilon_i^y=\|y_i-y(t_i)\|$, $i=-m_N,...,K$. Очевидно, что

$$\varepsilon_i^x=\varepsilon_i^y=0,\qquad i=-m_N,...,0. \tag{3.11}$$

В силу условия Липшица (2.2) и ограничения (2.3) имеем

$$\|G^{-1}Q_1(f(t_i,u_i)-f(t_i,u(t_i)))\|\le M(\varepsilon_i^x+\varepsilon_i^y),\quad i=0,...,K. \tag{3.12}$$

$$\|G^{-1}Q_2(f(t_i,u_i)-f(t_i,u(t_i)))\|\le q(\varepsilon_i^x+\varepsilon_i^y),\quad i=0,...,K. \tag{3.13}$$

Вычитая из уравнений (3.3),(3.4) уравнения (3.9),(3.10) соответственно, с учетом (3.12),(3.13) получим оценки

$$\varepsilon_{i+1}^x\le(1+Mh)\varepsilon_i^x+O(h)\varepsilon_i^y+O(h)\sum_{j=1}^{N}\left(\varepsilon_{i-m_j}^x+\varepsilon_{i-m_j}^y\right)+$$

$$+O(h^2)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j-1}(\varepsilon_k^x+\varepsilon_k^y)+O(h^2);\quad i=0,...,K-1, \tag{3.14}$$

$$\varepsilon_{i+1}^y\le q_0\varepsilon_{i+1}^x+\sum_{j=1}^{N}L_j\left(\varepsilon_{i+1-m_j}^x+\varepsilon_{i+1-m_j}^y\right)+$$

$$+O(h)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j}(\varepsilon_k^x+\varepsilon_k^y)+O(h);\qquad i=0,...,K-1, \tag{3.15}$$

где $L_j=\dfrac{\|G^{-1}Q_2B_j\|}{1-q}(j=1,...,N), q_0=\dfrac{q}{1-q}$. Оценим ε_{i+1}^x в правой части (3.15) с помощью неравенства (3.14). Тогда (3.15) примет вид

$$\varepsilon_{i+1}^y\le q_0(1+Mh)\varepsilon_i^x+\sum_{j=1}^{N}L_j\left(\varepsilon_{i+1-m_j}^x+\varepsilon_{i+1-m_j}^y\right)+$$

$$+O(h)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j}(\varepsilon_k^x+\varepsilon_k^y)+O(h);\qquad i=0,...,K-1, \tag{3.16}$$

Применяя рекуррентно (3.14),(3.16) и учитывая начальные условия (3.11), получим оценки

$$\varepsilon_i^x\le\sum_{l=1}^{i}(1+Mh)^{i-l}F_{l-1}^1+F_{i-1}^1,\qquad i=1,...,K \tag{3.17}$$

$$\varepsilon_i^y\le q_0\sum_{l=1}^{i}(1+Mh)^{i-l}F_{l-1}^1+F_{i-1}^2,\quad i=1,...,K\,, \tag{3.18}$$

где

$$F_i^1=F_i^1(\varepsilon_1^x,\varepsilon_1^y,...,\varepsilon_i^x,\varepsilon_i^y)=O(h)\varepsilon_i^y+O(h)\sum_{j=1}^{N}\left(\varepsilon_{i-m_j}^x+\varepsilon_{i-m_j}^y\right)+$$

$$+O(h^2)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j-1}(\varepsilon_k^x+\varepsilon_k^y)+O(h^2),$$

$$F_i^2=F_i^2(\varepsilon_1^x,\varepsilon_1^y,...,\varepsilon_i^x,\varepsilon_i^y)=\sum_{j=1}^{N}L_j\left(\varepsilon_{i+1-m_j}^x+\varepsilon_{i+1-m_j}^y\right)+$$

$$+O(h)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j}(\varepsilon_k^x+\varepsilon_k^y)+O(h).$$

Тогда с учетом нулевых начальных условий (3.11), вида вектор - функций F_i^1,F_i^2, оценок (3.17),(3.18) и неравенств

$$(1+Mh)^i\le e^{M(T-t_0)},\qquad i=1,...,K\,, \tag{3.19}$$

получаем линейное дискретное разностное неравенство для норм векторов $\varepsilon_i=\begin{pmatrix}\varepsilon_i^x\\ \varepsilon_i^y\end{pmatrix}$

$$\|\,\varepsilon_i\,\|\le O(h)\sum_{j=0}^{i-1}\|\,\varepsilon_j\,\|+O(h),\qquad i=1,...,k_1=\min\{m_1,K\}\,. \tag{3.20}$$

В силу следствия 4.1.2 [19, с. 186] из (3.20) получаем

$$\max_{i=1,...,k_1}\|\,\varepsilon_i\,\|\le O(h)(1+O(h))^{k_1}=O(h)\,, \tag{3.21}$$

что доказывает сходимость вычислительного процесса на отрезке $[t_0,t_0+\omega_1]\bigcap[t_0,T]$. Если $t_0+\omega_1<T$, то рассмотрим систему разностных

неравенств (3.17),(3.18) для $i = m_1,...,k_2 - 1$, где $k_2 = \min\{K, 2m_1\}$. Эта система исследуется при начальном условии

$$\| \varepsilon_i \| = O(h), \qquad i = -m_N,...,m_1, \tag{3.22}$$

справедливом в силу соотношений (3.11) и (3.22). Повторяя проведенные выше рассуждения относительно неравенств (3.14), (3.16) с учетом начального условия (3.22) и оценки (3.19), получим $\max\limits_{i=1,...,k_2} \| \varepsilon_i \| = O(h)$. За конечное число шагов получим оценку

$$\max_{i=1,...,K} \| \varepsilon_i \| = O(h),$$

которая доказывает теорему.

Как уже отмечалось выше, процесс построения численного решения задачи (1.1),(1.2) по разностной схеме (3.3)–(3.6) связан с трудностями разрешимости уравнения (3.4) относительно y_{i+1}. Условие сжимаемости (2.3) позволяет построить приближенное решение этого уравнения с использованием метода простых итераций подобно тому, как это делалось в [9] для дифференциально-алгебраических уравнений вида (1.4). Поэтому мы изменим вычислительную схему (3.3)–(3.6). Будем искать приближенное решение $\hat{u}_i = \hat{x}_i + \hat{y}_i = P_1\hat{u}_i + P_2\hat{u}_i$ в узлах сетки $t_i (i = -m_N,...,K)$. Разностное уравнение (3.4) заменим следующим

$$\hat{x}_{i+1} = (E - Sh)\hat{x}_i - h\sum_{j=1}^{N} G^{-1}Q_1 B_j \hat{u}_{i-m_j} - h^2 \sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j-1} G^{-1}Q_1 \Phi_j(t_i, t_k)\hat{u}_k +$$

$$+ hG^{-1}Q_1 f(t_i, \hat{x}_i + \hat{y}_i), \ \ i = 0,...,K-1 \tag{3.23}$$

Численное решение $\hat{u}_i$ также удовлетворяет начальным условиям (3.6):

$$\hat{x}_i = P_1 g(t_0 + ih), \hat{y}_i = P_2 g(t_0 + ih), \hat{u}_i = \hat{x}_i + \hat{y}_i, \ i = -m_N, -m_N + 1,...,0. \tag{3.24}$$

Для нахождения $\hat{y}_{i+1} (i = 0,...,K-1)$ с учетом найденных уже векторов $\hat{x}_{i+1}, \hat{u}_k (k = 0,...,i)$ определяем последовательность итераций

$$z_{i+1,0} = \hat{y}_i, \ z_{i+1,s} = G^{-1}Q_2 f(t_{i+1}, \hat{x}_{i+1} + z_{i+1,s-1}) - \sum_{j=1}^{N} G^{-1}Q_2 B_j \hat{u}_{i+1-m_j} -$$

$$-h\sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j} G^{-1}Q_2 \Phi_j(t_{i+1}, t_k)\hat{u}_k, \qquad s = 1,...,m; \ i = 0,...,K-1, \tag{3.25}$$

где $m = m(h) = \left[\dfrac{2\ln h}{\ln q} \right] + 1$, [x]–целая часть числа x. В силу принципа сжимающих отображений [20, с.44] существуют предельные векторы $z_{i+1} = \lim\limits_{s \to \infty} z_{i+1,s} (i = 0,...,K-1)$, удовлетворяющие соответствующим уравнениям

$$z_{i+1} = G^{-1}Q_2 f(t_{i+1}, \hat{x}_{i+1} + z_{i+1}) - \sum_{j=1}^{N} G^{-1}Q_2 B_j \hat{u}_{i+1-m_j} -$$

$$-h\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j} G^{-1}Q_2\Phi_j(t_{i+1},t_k)\hat{u}_k,\ i=0,...,K-1 \qquad (3.26)$$

и оценкам [20,с. 45]

$$\| z_{i+1} - \hat{y}_{i+1} \| \le \frac{q^m}{1-q} \| \hat{y}_i - G^{-1}Q_2 f(t_{i+1}, \hat{x}_{i+1} + \hat{y}_i) + \sum_{j=1}^{N} G^{-1}Q_2 B_j \hat{u}_{i+1-m_j} +$$

$$+h\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j} G^{-1}Q_2\Phi_j(t_{i+1},t_k)\hat{u}_k \|, \quad i=0,...,K-1 \quad (3.27)$$

Таким образом, соотношения (3.23)–(3.25),(3.6) приводят к построению численного решения $\hat{u}_i$ задачи (1.1),(1.2) в узле сетки t_i. Докажем теорему сходимости вычислительного метода (3.23)–(3.25),(3.6). Предварительно докажем вспомогательную лемму, в которой по существу устанавливается сходимость вычислительной схемы (3.23)–(3.25),(3.6) к решению задачи (3.3)–(3.6).

Лемма 1. *Пусть выполнены условия теоремы 1. Тогда*

$$\max_{i=0,...,K} \| \hat{u}_i - u_i \| = O(h).$$

Доказательство. Обозначим $\delta_i^x = \| x_i - \hat{x}_i \|, \delta_i^y = \| y_i - \hat{y}_i \|, i = -m_N,...,K$. Вычитая из уравнения (3.3) уравнение (3.23), с учетом условия Липшица (2.2), получаем неравенство (ср. с (3.14))

$$\delta_{i+1}^x \le (1+Mh)\delta_i^x + O(h)\delta_i^y + O(h)\sum_{j=1}^{N}\left(\delta_{i-m_j}^x + \delta_{i-m_j}^y\right) +$$

$$+O(h^2)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j-1}(\delta_k^x + \delta_k^y) + O(h^2); \quad i=0,...,K-1 \qquad (3.28)$$

Вычитая из уравнения (3.4) уравнение (3.26) с учетом условия сжимаемости (2.3) , получаем неравенство (ср. с (3.15))

$$\| y_{i+1} - z_{i+1} \| \le q_0\delta_{i+1}^x + \sum_{j=1}^{N} L_j\left(\delta_{i+1-m_j}^x + \delta_{i+1-m_j}^y\right) +$$

$$+O(h)\sum_{j=0}^{N}\sum_{k=-m_j}^{i-m_j}(\delta_k^x + \delta_k^y), \quad i=0,...,K-1. \qquad (3.29)$$

Теперь с учетом (3.27) и (3.29) получаем оценку для δ_{i+1}^y:

$$\delta_{i+1}^y \le \| y_{i+1} - z_{i+1} \| + \| \hat{y}_{i+1} - z_{i+1} \| \le q_0 \delta_{i+1}^x + \sum_{j=1}^{N} L_j \left(\delta_{i+1-m_j}^x + \delta_{i+1-m_j}^y \right) +$$

$$+ O(h) \sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j} (\delta_k^x + \delta_k^y) + \frac{q^m}{1-q} \| \hat{y}_i - G^{-1} Q_2 f(t_{i+1}, \hat{x}_{i+1} + \hat{y}_i) +$$

$$+ \sum_{j=1}^{N} G^{-1} Q_2 B_j \hat{u}_{i+1-m_j} + h \sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j} G^{-1} Q_2 \Phi_j(t_{i+1}, t_k) \hat{u}_k \|, \quad i = 0,..., K-1. \quad (3.30)$$

В силу теоремы 2 и непрерывности функции $f(t,x)$ существует постоянная $C_2 > 0$ такая, что при достаточно малых $h > 0$ справедливо неравенство

$$\| y_i - G^{-1} Q_2 f(t_{i+1}, x_{i+1} + y_i) +$$

$$+ \sum_{j=1}^{N} G^{-1} Q_2 B_j u_{i+1-m_j} + h \sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j} G^{-1} Q_2 \Phi_j(t_{i+1}, t_k) u_k \| \le C_2, i = 0,..., K-1. (3.31)$$

Оценим правую часть неравенства (3.30) с помощью (3.28),(3.31) и неравенства $q^m \le h^2 = O(h^2)$, вытекающего из определения $m = m(h)$. Получим оценку

$$\delta_{i+1}^y \le q_0 (1 + Mh) \delta_i^x + \sum_{j=1}^{N} L_j \left(\delta_{i+1-m_j}^x + \delta_{i+1-m_j}^y \right) +$$

$$+ O(h) \sum_{j=0}^{N} \sum_{k=-m_j}^{i-m_j} (\delta_k^x + \delta_k^y) + O(h^2); \quad i = 0,..., K-1, \quad (3.32)$$

Система неравенств (3.28),(3.32) исследуется аналогично системе неравенств (3.14),(3.16), полученной при доказательстве теоремы 2, с учетом начальных условий $\delta_i^x = \delta_i^y = 0, i = -m_N,...,0$. Повторяя рассуждения, проведенные при доказательстве теоремы 2 для (3.14),(3.16), получим $\max\limits_{i=0,...,K} (\delta_i^x + \delta_i^y) = O(h)$. Лемма доказана.

Теперь из утверждений теоремы 2 и леммы 1 непосредственно вытекает теорема сходимости вычислительного метода (3.23)–(3.25).

Теорема 3. *В условиях теоремы 1 численный метод (3.23)–(3.25),(3.6) имеет первый порядок сходимости :* $\max\limits_{i=0,...,K} \| u(t_i) - u_i \| = O(h)$

4. Пример

На рис. 1–2 приведены в графической форме результаты численного решения задачи (1.1),(1.2). Эта задача рассмотрена в пространстве R^2 на отрезке $[t_0, T] = [0,100]$, причем $N = 2, \omega_j = j (j = 1,2)$, $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B_0 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$,

$B_1 = \begin{pmatrix} 0 & 0.2 \\ 0.3 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 0.2 & 0.4 \\ -0.3 & 0.1 \end{pmatrix}, f(t,x) = \begin{pmatrix} \cos t + 0.01\sin^2 x_1 \\ \sin t + 0.01\sin(x_1 + x_2) \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$g(t) \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Phi_0(t,\tau) = \frac{1}{(t-\tau)^2+1} \begin{pmatrix} (t-\tau)^2 & t-\tau \\ (t-\tau)+4 & t-\tau+1 \end{pmatrix},$

$\Phi_1(t,\tau) = \frac{1}{(t-\tau)^2+1} \begin{pmatrix} t-\tau+2 & t-\tau \\ (t-\tau)^2 & t-\tau \end{pmatrix}, \Phi_2(t,\tau) = \frac{1}{(t-\tau)^2+1} \begin{pmatrix} 2(t-\tau) & 1 \\ (t-\tau) & 0 \end{pmatrix}.$ Пучок $\lambda A_0 + B_0$ регулярный и имеет индекс 1. При этом $P_1 = Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$ $P_2 = Q_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$ $G = E$ и выполняются соотношения (2.3),(2.4). Функция $f(t,x)$ удовлетворяет ограничениям (2.2),(2.3) с постоянными $M = q = 0.02$. Поэтому условия теорем 1–3 выполнены. В силу теоремы 3 для нахождения численного решения этой задачи на отрезке $[0,T]$ можно использовать вычислительную схему (3.23)–(3.25), (3.6), которая имеет первый порядок сходимости.

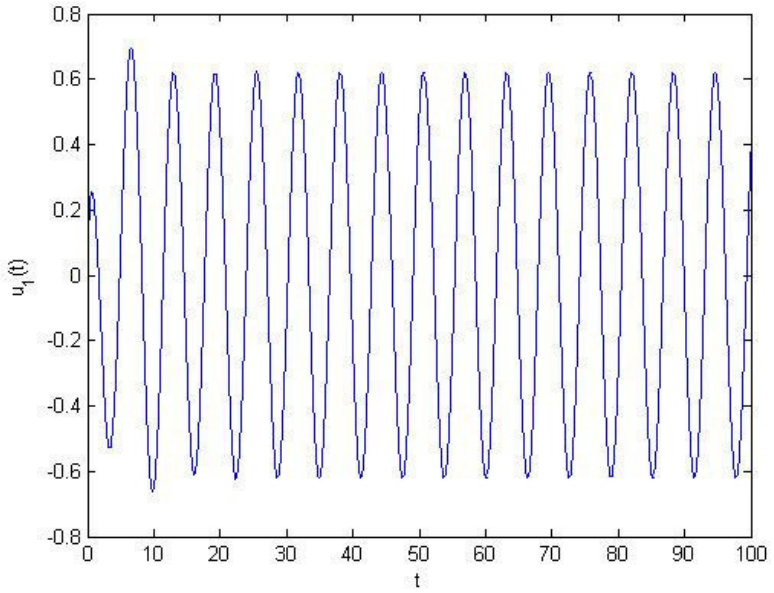


Рис. 1

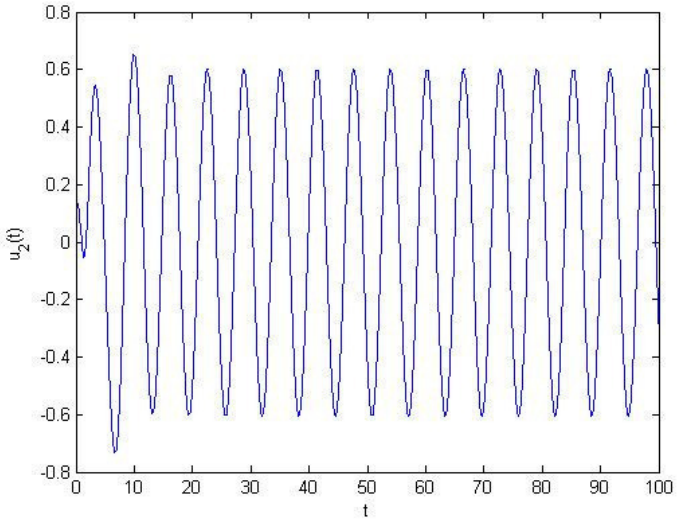


Рис.2

5. Выводы

Вычислительная схема (3.23)–(3.25),(3.6) позволяет находить численное решение определенных классов вырожденных интегро-дифференциальных уравнений вида (1.1). Условия сходимости этой схемы установлены в теореме 3 . Разработанный в статье численный метод может быть использован для построения приближенного решения уравнений вида (1.1), удовлетворяющих условиям теоремы 1.

ЛИТЕРАТУРА

1. Rutkas A.G., Vlasenko L.A. Existence, uniqueness and continuous dependence for implicit semilinear functional differential equations // Nonlinear Analysis. TMA.–2003–V. 55, № 1-2.–P.125–139.
2. Пивень А.Л. Существование и единственность решения одного вырожденного интегро-дифференциального уравнения с запаздываниями // Вісник Харк. Нац. Університету імені В.Н. Каразіна. Серія Математика, прикладна математика і механіка.–2011.–№ 967.– С. 17–31.
3. Власенко Л.А., Мышкис А.Д., Руткас А.Г. Об одном классе дифференциальных уравнений параболического типа с импульсными воздействиями // Дифференциальные уравнения.–2008.–Т.44, № 2.– С. 222–231.
4. Власенко Л.А. Руткас А.Г. Переходные процессы в цепях с диспергирующими многопроводными линиями передачи // Радиотехника.–2010.– № 161.– С. 105–114.
5. Rutkas A.G., Vlasenko L.A. Time-domain descriptor models for circuits with multiconductor transmission lines and lumped elements // Proceedings of IEEE

5-th International Conference on Ultrawideband and Ultrashort Impulse Signals.–Sevastopol, Ukraine, September 6–10.–2010.–P. 102–104.
6. Ascher M., Petzold L. The numerical solution of delay-differential-algebraic equations of retarded and neutral type // SIAM J.Numer.Anal.–1995.–Vol.32, №5.–P.1635–1657.
7. Brennan K.E., Campbell S.L., Petzold L.R. Numerical solution of initial-value problems in differential algebraic equations.–SIAM: Classic in Applied Mathematics, 1995.–256 p.
8. Куликов Г.Ю. О численном решении автономной задачи Коши с алгебраической связью на фазовые переменные // Журнал выч. математики и мат. физики. – 1993.– Т.33, № 4.–С. 522–540.
9. Куликов Г.Ю. Теоремы сходимости для итеративных методов Рунге-Кутты с постоянным шагом интегрирования // Журнал выч. математики и мат. физики. – 1996.– Т.36, № 8.–С. 73–89.
10. Булатов М.В., Чистякова Е.В. Численное решение интегро-дифференциальных систем с вырожденной матрицей перед производной многошаговыми методами // Дифференциальные уравнения.–2006.–Т. 42, № 9.– С. 1248–1255.
11. Чистякова Е.В. Дифференциально-алгебраические уравнения с малым нелинейным членом // Дифференциальные уравнения.–2009.–Т. 45, № 11.–С. 1365–1368.
12. Чистякова Е.В., Чистяков В.Ф. О разрешимости вырожденных систем квазилинейных интегро-дифференциальных уравнений общего вида // Вычислительные технологии.–2011.–Т. 16, № 5.– С. 100–113.
13. Руткас А.Г. Задача Коши для уравнения $Ax'(t)+Bx(t)=f(t)$ // Дифференциальные уравнения.–1975.–Т.11, № 11.–С. 1996–2010.
14. Гантмахер Ф.Р. Теория матриц.– М.: Наука, 1988.– 548 с.
15. Власенко Л.А. Эволюционные модели с неявными и вырожденными дифференциальными уравнениями. – Днепропетровск: Системные технологии, 2006.–273 с.
16. Власенко Л.А. Математические модели с уравнениями типа Соболева: Учебное пособие .– Харьков: ХНУ им.В.Н.Каразина, 2011.– 112 с.
17. Пивень А.Л. Численное решение вырожденного интегро-дифференциального уранения с запаздываниями// Вісник Харк. Нац. Університету імені В.Н. Каразіна. Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління».–2013.–№ 1058.–С. 132–141.
18. Agarval R. Difference equations and inequalities. Theory, methods and applications. – Marcel: Chapman and Hall/CRC Pure and Applied Mathematics, 2000.–971 p.
19. Самарский А.А., Гулин А.В. Численные методы. – М.:Наука, 1989.– 432 с.
20. Люстерник Л.А., Соболев В.И. Элементы функционального анализа.–М.: Физматлит, 1965.–520 с.

Надійшла у першій редакції 15.05.2014, в останній – 16.05.2014.

УДК 519.6

Competitive diffusion of benzene-hexane mixtures in microporous medium: mathematical modeling and parameters identification

J. Fraissard[1], S. Leclerc[2], D. Mykhalyk[3], M. Petryk[3]

1 – LPEM-ESPCI ParisTech and Université Pierre et Marie Curie, France

2 – UMR7563 Laboratoire d'energetique et de mecanique theorique et appliquee, University of Lorraine, France

3 – Ternopil National Ivan Pul'uj Technical University, Ukraine

В статье представлена эффективная процедура идентификации коэффициентов компетитивной диффузии двух газов (бензола и гексана) в пространствах между частицами и в частицах среды, реализована с использованием градиентных методов, разработанной математической модели для процессов компетитивнои и моно диффузии и результатов ЯМР-анализа распределений адсорбированной массы каждой из компонент в цеолитной среде. Распределения коэффициентов диффузии получены как функции от времени для различных положений вдоль исследуемой цеолитной среды. Построены модельные профили изменения концентрации бензола и гексана в среде во времени на основе полученных коэффициентов диффузии и разработанной математической модели.

Ключевые слова: *компетитивная диффузия газов, адсорбция, математическое моделирование, коэффициенты диффузии, микропористая среда, градиентные методы идентификации*

В статі представлена ефективна процедура ідентифікації коефіцієнтів компетитивної дифузії двох газів (бензолу та гексану) у міжчастинковому та внутрішньочастинковому просторах, реалізована з використанням градієнтних методів, розробленої математичної моделі для процесів компетитивної та моно дифузії та результатів ЯМР-аналізу розподілів адсорбованої маси кожної з компонент в цеолітному середовищі. Розподіли коефіцієнтів дифузії отримано як функції від часу для різних положень уздовж досліджуваного цеолітного середовища. Побудовано модельні профілі зміни концентрації бензолу та гексану в середовища в часі на основі отриманих коефіцієнтів дифузії та розробленої математичної моделі.

Ключові слова: *компетитивна дифузія газів, адсорбція, математичне моделювання, коефіцієнти дифузії, мікропористе середовище, градієнтні методи ідентифікації*

In the paper we propose new effective identification procedures for competitive diffusion coefficients of two gases (benzene and hexane) in intra- and inter- crystallite space. This procedure is developed using high-speed gradient methods, mathematical models of competitive as well as monodiffusion, and NMR imaging of adsorbed masses distribution for each component in zeolite crystallite bed. Distributions of diffusion coefficients as functions of time were obtained for different positions along the microspores. The benzene and hexane concentration curves as functions of time were built at every level of crystallite bed for inter crystallite space, based on identified coefficients and developed mathematical models.

Key words: *Competitive diffusion of gases, adsorption, modeling, diffusion coefficient, microporous zeolite bed, intra- and intracrystallites space, gradient methods of identification.*

1. General problem

Knowledge of the co-diffusion and co-adsorption coefficients of reactants and products is essential when a heterogeneous catalysis reaction is performed by means of gases flowing through a porous catalyst bed. It can be even more important in a

fluidized bed where the contact times between the catalyst pellets and the reactants are very short, which means that one hardly ever knows whether the pellet is used completely or only superficially. Under these experimental conditions the system is never at adsorption equilibrium. The distribution of the various reactants adsorbed on the catalyst is very heterogeneous and, moreover, it varies greatly from one reactant to another. However, calculations on the kinetics are generally performed by assuming the ideal case of a homogeneous assembly, both for the totality of the catalyst and for the distribution of all the reactants, which makes these calculations, even when they are mathematically rigorous, rather approximate as reality regards.

Mass transfer in catalysis is a rather complex process involving several parameters which are seldom defined by a single technique. Nuclear magnetic resonance (NMR) has provided much information about these problems. It can be divided into two parts: conventional NMR and Magnetic Resonance Imaging (MRI) [1]. With the first one, measurements characterize the entire sample without spatial resolution. Its application to the study of the diffusion of an adsorbat most often concerns self-diffusion measurements on species in adsorption equilibrium.

Unlike the known research on diffusion of separate gases for the adsorptions catalysts, this project includes research of kinetics and development of methods of kinetic parameters identification for competitive (compatible) diffusive transfer and adsorption of two hydrocarbon components in macro- and nanoporous of heterogeneous catalysts.

Many researches in this area of problem mainly concerned molecular transport of certain substances in the porous medium and considered mass transfer at the macro level without significantly influence the effects and characteristics of micro- and nano-porous of particles itself. But this kind of masstransfer are limiting and determining factor in overall diffusion kinetics of such type of mediums [2]. Another unexplored aspect is co-diffusion of two or more components in the same time or so-called competitive diffusion [3]. It is therefore essential to clarify mechanisms and factors of competitive diffusion kinetics of microporous mediums and to identify each of interacting components.

2. Mathematical model

The presented mathematical model is similar to biporous model [4, 5], with a system of complex competitive masstransfer between two diffused components (benzene and hexane) in a heterogeneous media (crystallite bed) of porous crystallites. We suppose that diffusion process causes two types of mass transfer: diffusion in the macropore (intracrystallite space) by the space between particles and diffusion in micro- and nanopores of crystallites (intraparticle space).

A cylindrical bed of microporous zeolite crystallites, assumed to be spherical (radius R), is exposed to a constant concentration of adsorbate in the gas phase. One face of this bed is permeable to the two gases (benzene at hehane). In this case one can consider that two gases diffusion is axial in the macropores (z direction along the height, l, of the bed) and radial in the micropores. We assume that the crystallites zeolite bed consists of a large number N, of very thin solid layers of thickness $\Delta l_k = l_k - l_{k-1}$, perpendicular to the propagation of the gas in the z direction. In the proposed mathematical model we assume that: (i) during the evolution of the system

towards equilibrium there has to be a concentration gradient in the macropores and/or in the micropores; (ii) the effect of heat is negligible; (iii) diffusion occurs in Henry's law region of the adsorption isotherm; (iv) all solid particles are spherical; (v) all solid particles are of the same size; uniform packing through the crystallite bed is assumed.

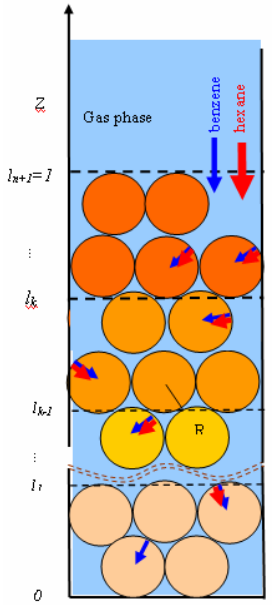


Fig. 1 The scheme of diffusion completive in crystallite bed

Diffusion in macropores
Length of the bed: l
Characteristic position of the layer: l_k, $k=\overline{1,n+1}$
Thickness of layer k: $\Delta l_k = l_{k+1} - l_k$, $k=\overline{1,n+1}$
Intercrystallite diffusion coefficient in k[th] layer: $D_{inter,k}$;
Corresponding characteristic time: $\tau_{inter,k}$

Diffusion in micropores
Crystallite radius: R
Intracrystallite diffusion coefficient in k[th] layer: $D_{intra,k}$
Corresponding characteristic time: $\tau_{intra,\,k}$
$l >> R$; $D_{inter,k} >> D_{intra,k}$; $k=\overline{1,n+1}$

The mathematical model of gas diffusion kinetics in the crystallite zeolite bed (which we consider to be a heterogeneous and multilayer porous medium) is defined in domains Ω_{m_T} by the solutions of the system of differential equations [6]

$$\frac{\partial C_{s_m}(t,Z)}{\partial t} = \frac{D_{inter_{s_m}}}{l^2}\frac{\partial^2 C_{s_m}}{\partial Z^2} - e_{\text{inter}_m} K_{s_m} \frac{D_{intra_{s_m}}}{R^2}\left(\frac{\partial Q_{s_m}}{\partial X}\right)_{X=1}, \quad (1)$$

$$\frac{\partial Q_{s_m}(t,X,Z)}{\partial t} = \frac{D_{intra_{s_m}}}{R^2}\frac{\partial^2 Q_{s_m}}{\partial X^2}. \quad (2)$$

Domain: $\Omega_{m_T} = (0,T)\times\Omega_m$, $\left(\Omega_m = (L_{m-1}, L_m), m=\overline{1,N+1}, L_0 = 0 < L_1 < \ldots < L_{N+1} = 1\right)$.

With initial conditions

$$C_{s_m}(t=0,Z)=0; \quad Q_{s_m}(t=0,X,Z)=0;\ X\in(0,R),\ Z\in\Omega_m,\ m=\overline{1,N+1}, \quad (3)$$

boundary and interface conditions for Z coordinate:

$$C_{s_1}(t,L_1)=1,\ \frac{\partial C_{s_1}}{\partial Z}(t,Z=0)=0,\ t\in(0,T); \quad (4)$$

$$\left[C_{s_m}(t,Z) - C_{s_m}(t,Z)\right]_{Z=L_m} = 0\ ,\ m=\overline{1,N}$$

$$\frac{\partial}{\partial Z}\left[D_{inter_{s_{m-1}}} C_{s_{m-1}}(t,Z) - D_{inter_{s_m}} C_{s_m}(t,Z) \right]_{Z=L_m} = 0,\ t \in (0,T) \qquad (5)$$

and boundary conditions for X coordinate:

$$\frac{\partial}{\partial X} Q_{s_m}(t, X=0, Z) = 0;\ Q_{s_m}(t, X=1, Z) = C_{s_m}(t,Z),\ Z \in \Omega_m,\ m = \overline{1, N+1}. \qquad (6)$$

Here

$$Z = \frac{z}{\ell},\ X = \frac{r}{\mathrm{R}},\ C = \frac{c}{\mathrm{c}_\infty},\ Q = \frac{q}{\mathrm{q}_\infty};$$

$$\mathrm{e}_{\mathrm{inter}_m} = \frac{\mathrm{e}_{\mathrm{inter}_m} c_{sm}}{\mathrm{e}_{\mathrm{inter}_m} c_{sm} + (1-\mathrm{e}_{\mathrm{inter}_m}) q_{sm}} \approx \frac{\mathrm{e}_{\mathrm{inter}_m}}{(1-\mathrm{e}_{\mathrm{inter}_m})\mathrm{K}_{sm}};\quad \mathrm{e}_{\mathrm{intra}_m} = 1 - \mathrm{e}_{\mathrm{inter}_m},\ \mathrm{K}_{s_m} = \frac{\mathrm{q}_{s_m\infty}}{\mathrm{c}_{s_m\infty}}.$$

$\bar{Q}_s(t,Z) = \int_0^1 Q_s(t,X,Z)\,dX$ - average concentration for s^{th} adsorbed component ($s = \overline{1,2}$) in microporous of crystallite; c_s, q_s – adsorbent concentration in macro and micro porous (molecule/cm^3); D_{intra_s}, D_{inter_s} – diffusion coefficient in micro- and macro porous (m^2/s); K_s – adsorption equilibrium constant, $\varepsilon_{\mathrm{inter}}$ – porosity; e_{inter} – porosity coeficient; R – partical radius (m); ℓ – thickness of working area of catalytic media (m).

3. Analytical solution

To find the analytical solutions of direct problems (1)-(6) we used the Heaviside method according to procedure originally used by Leniuk & Petryk [9]. As result we obtained functions C_{sm} and N_{sm} in theform:

$$C_{sm}(t,Z) = 1 + \frac{2\pi}{\Delta L}\frac{R^2}{\mathrm{D}_{\mathrm{intra}_{sm}}}\frac{D_{inter_{sm}}}{\Delta L^2}\times$$

$$\times \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} exp\left(-\frac{D_{intra_{sm}}\,\varepsilon_{kn_m}^2}{\mathrm{R}^2}t\right)\frac{n\left(\theta_{sm}\sin\left[\frac{n\pi}{\Delta L}(Z-L_{m-1})\right] + \theta_{s_{m-1}}\sin\left[\frac{n\pi}{\Delta L}(L_m - Z)\right]\right)}{(-1)^n\,\varepsilon_{kn_m}^2\left(\frac{3}{\mathrm{e}_{\mathrm{inter}_m}}\left(\frac{1}{\sin^2(\varepsilon_{kn_m})} - \frac{ctg(\varepsilon_{kn_m})}{\varepsilon_{kn_m}}\right) + 2\right)} \qquad (7)$$

$$N_{sm}(t,X,Z) = 1 + \frac{2\pi}{\Delta L}\frac{R^2}{\mathrm{D}_{\mathrm{intra}_{sm}}}\frac{D_{inter_{sm}}}{\Delta L^2}\times$$

$$\times \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} exp\left(-\frac{D_{intra_{sm}}\,\varepsilon_{kn_m}^2}{\mathrm{R}^2}t\right)\frac{n\cdot\sin(\varepsilon_{kn_m}X)\left(\theta_{sm}\sin\left[\frac{n\pi}{\Delta L}(Z-L_{m-1})\right] + \theta_{s_{m-1}}\sin\left[\frac{n\pi}{\Delta L}(L_m - Z)\right]\right)}{(-1)^n\,\varepsilon_{kn_m}^2\sin(\varepsilon_{kn_m})\left(\frac{3}{\mathrm{e}_{\mathrm{inter}_m}}\left(\frac{1}{\sin^2(\varepsilon_{kn_m})} - \frac{cot(\varepsilon_{kn_m})}{\varepsilon_{kn_m}}\right) + 2\right)} \qquad (8)$$

where $в_{kn_m}$ are roots of transcendent equations:

$$\varphi_{s1}^2(\beta) \equiv \frac{3}{e_{\text{inter}_1}} \frac{\Delta L^2}{R^2} \frac{D_{\text{intra}_{s1}}}{D_{\text{inter}_{s1}}} \left(\frac{e_{\text{inter}_1}}{3} \frac{R^2}{D_{\text{intra}_{s1}}} \beta^2 - \beta \cot \beta + 1 \right) = \frac{2n-1}{2\Delta L} \pi, \; n,k = \overline{1,\infty}. \quad (9)$$

4. Identification of diffusion coefficients.

Assuming, that coefficients of competitive diffusion in intracrystallite space D_{inter_s} and inter crystallite space D_{intra_s}, $s = \overline{1,2}$ (s=1 – corresponded benzene and s=2 corresponded hexane) are unknown.

Taking into account mathematical model (1)–(6), the identification problem can be formulated as: to find unknown functions $D_{\text{intra}_s} \in \Omega_T$, $D_{\text{inter}_s} \in \Omega_T$ ($D_{\text{intra}_s} > 0$, $D_{\text{inter}_s} > 0$, $s = \overline{1,2}$), when absolute absorbed masses for every point $\gamma_m \subset \Omega_m$ of m-th crystallite bed segment satisfy the condition:

$$\left[C_{s_m}(t,Z) + \bar{Q}_{s_m}(t,Z) \right]\Big|_{\gamma_m} = M_{s_m}(t,Z)\Big|_{\gamma_m}, \quad s = \overline{1,2}; \; \gamma_m \in \Omega_m, \gamma_m = L_m - L_{m-1} \quad (10)$$

According to paper [8] and using errors minimization gradient method for identification of competitive diffusion coefficients as function of time for intracrystallite space $D_{\text{intra}_{s_m}}$ (mycro level) and intercrystallite space $D_{\text{inter}_{s_m}}$ (macro level) of s-th diffused components (s=1 -benzene and s=2 - hexane), we obtain regulation expression for $(n+1)^{\text{th}}$ identification step:

$$D_{\text{intra}_{s_m}}^{n+1}(t) = D_{\text{intra}_{s_m}}^{n}(t) - \nabla J_{D_{\text{intra}_{s_m}}}^{n}(t) \times$$

$$\times \frac{\left[C_{s_m}\left(D_{\text{inter}_{s_m}}^{n}, D_{\text{intra}_{s_m}}^{n}; t, \gamma_m\right) + \left(\frac{1}{X}\right)_{X=\frac{1}{2}} N_{s_m}\left(D_{\text{inter}_{s_m}}^{n}, D_{\text{intra}_{s_m}}^{n}; t, \frac{1}{2}, \gamma_m\right) - M_{s_m}(t) \right]^2}{\left\| \nabla J_{D_{\text{intra}_{s_m}}}^{n}(t) \right\|^2 + \left\| \nabla J_{D_{\text{inter}_{s_m}}}^{n}(t) \right\|^2}, \; t \in (0,T) \quad ,(11)$$

$$D_{\text{inter}_{s_m}}^{n+1}(t) = D_{\text{inter}_{s_m}}^{n}(t) - \nabla J_{D_{\text{inter}_{s_m}}}^{n}(t) \times$$

$$\times \frac{\left[C_{s_m}\left(D_{\text{inter}_{s_m}}^{n}, D_{\text{intra}_{s_m}}^{n}; t, \gamma_m\right) + \left(\frac{1}{X}\right)_{X=\frac{1}{2}} N_{s_m}\left(D_{\text{inter}_{s_m}}^{n}, D_{\text{intra}_{s_m}}^{n}; t, \frac{1}{2}, \gamma_m\right) - M_{s_m}(t) \right]^2}{\left\| \nabla J_{D_{\text{intra}_{s_m}}}^{n}(t) \right\|^2 + \left\| \nabla J_{D_{\text{inter}_{sm}}}^{n}(t) \right\|^2}, \; t \in (0,T) \quad .(12)$$

Where:

$J\left(D_{\text{inter}_{s_m}}, D_{\text{intra}_{s_m}}\right)$- error's functional of model solution from experimental data (observation traces) on $\gamma_m \in \Omega_m$:

$$J\left(D_{\text{inter}_{s_m}}, D_{\text{intra}_{s_m}}\right) = \frac{1}{2}\int_0^T \left[C_{s_m}\left(\tau, Z, D_{\text{inter}_{s_m}}, D_{\text{intra}_{s_m}}\right) + \left(\frac{1}{X}\right)_{X=\frac{1}{2}} N_{s_m}\left(t, Z, D_{\text{inter}_{s_m}}, \frac{1}{2}, D_{\text{intra}_{s_m}}\right) - M_{s_m}(t) \right]^2_{\gamma_m} d\tau,$$

$$\gamma_m \in \Omega_m,\ m = \overline{1, N+1}. \qquad (13)$$

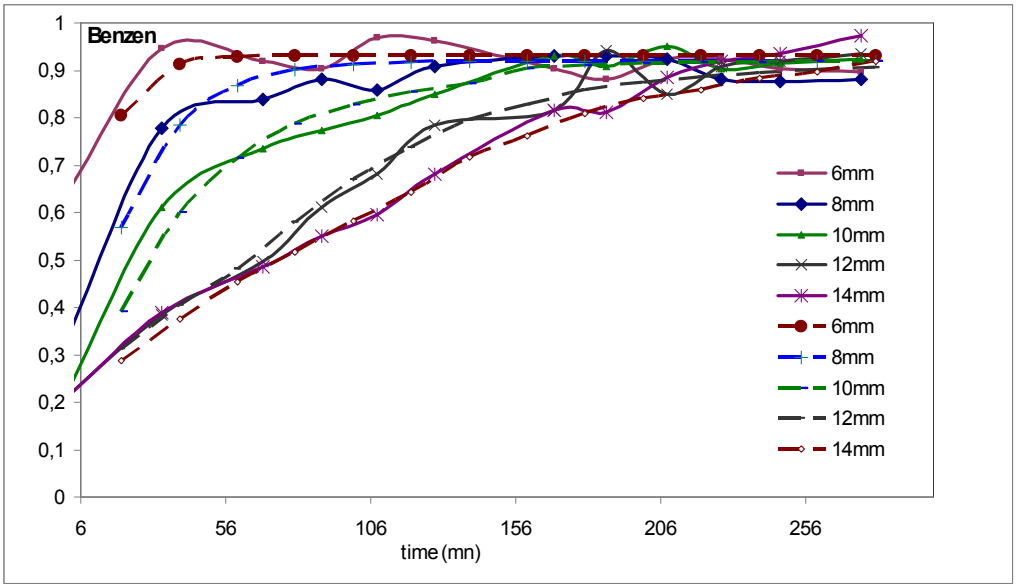
Here:

$\nabla J^n_{D_{\text{intra}_{s_m}}}(t)$, $\nabla J^n_{D_{\text{intra}_{s_m}}}(t)$ are errors's functional gradients $J\left(D_{\text{inter}_{s_m}}, D_{\text{intra}_{s_m}}\right)$;

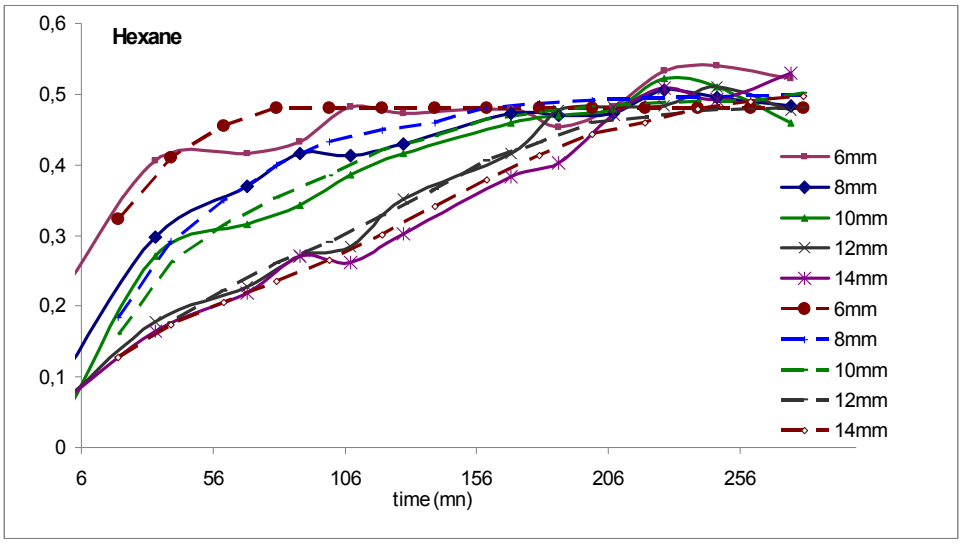
$\left\|\nabla J^n_{D_{\text{intra}_{s,m}}}(t)\right\|^2 = \int_0^T \left[\nabla J^n_{D_{\text{intra}_{s,m}}}(t)\right]^2 dt$, $\left\|\nabla J^n_{D_{\text{inter}_{s,m}}}(t)\right\|^2 = \int_0^T \left[\nabla J^n_{D_{\text{inter}_{s,m}}}(t)\right]^2 dt$ - norms of functional gradient;

$M_s(t,Z)\big|_{\gamma_m}$ - data vector of experimental distribution (observational trace) of absorbed mass in macro- and micro porous on surface area $\gamma_m \subset \Omega_m$ as results of NMR-analysis (Fig. 2a – experimental data for benzene, Fig. 2b – experimental data for hexane).

The technology of calculation of each component in case of monodiffusion based on the works optimal control theory [7] and is described in paper [8] in details.



a)



b)

Fig. 2. Distribution of the benzene and hexane concentrations in time for different sample layers (experimental curves and approximations).

5. Numerical simulation and analysis.

According to proposed methodology of parameter identification based on optimal control theory [7] and using experimental data approximations (fig.2) the distributions of diffusion coefficients were obtained. The identified benzene diffusion coefficients $D_{\text{int}ra_{1,k}}$ and $D_{\text{int}er_{1,k}}$ are presented as functions of time for the five coordinates thickness positions: 6, 8, 10, 12, 14 mm in fig. 3, 4.

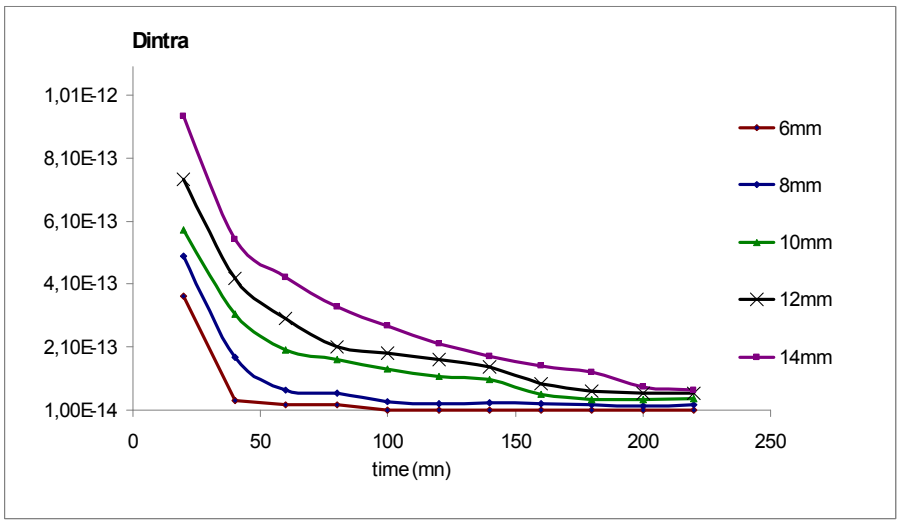


Fig. 3. Diffusion coefficients profiles $D_{\text{int}ra_{1,k}}$ in intraparticle space for benzene in time for different positions in catalytic bed

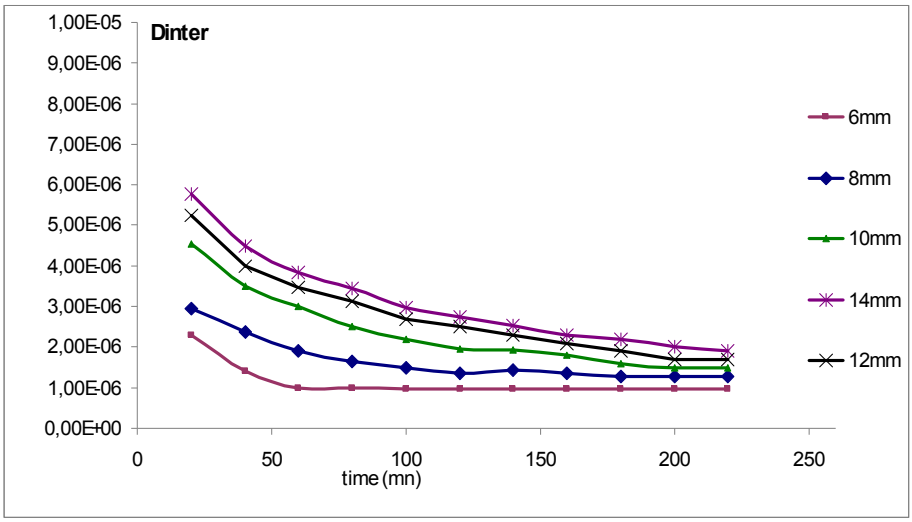


Fig. 4. Diffusion coefficients profiles $D_{\text{int}\,er_{1,k}}$ *in interparticles space for benzene in time for different positions in catalytic bed*

Competitive diffusion coefficient curves for the intraparrticle space are of the pseudo exponential nature and in a case of benzene varying in the range from 1.0 e-12 till 5.0 e-13. Since the diffusion time t = 80-100mn relatively gentle picture of changes in their values is observed testifying the diffusion approximation toward the equilibrium level.

The diffusion coefficient distributions (fig. 4) are of more gentle nature and changed in the range of 6.0 e-6 till 1.0 e-6. Since the diffusion time t = 130-140 mn also relatively gentle picture of changes in their values is observed testifying the process of convergence to equilibrium.

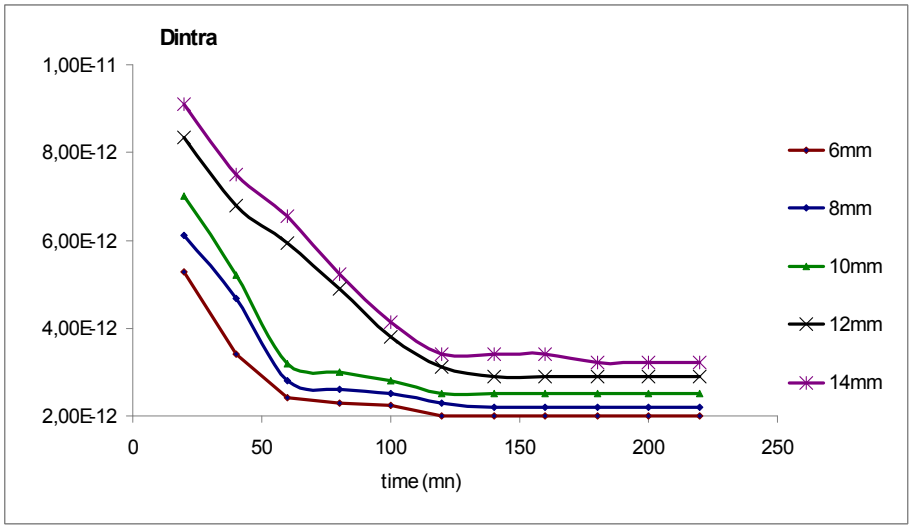


Fig. 5. Diffusion coefficients profiles $D_{\text{int}\,ra_{2,k}}$ *in intraparticles space for hexane in time for different positions in catalytic bed*

Fig. 5 and 6 show similar results but for hexane. Identified competitive diffusion coefficients distributions as functions of time for the same positions of coordinates thickness (6, 8, 10, 12, 14 mm) change in time in the range of 1.0 e-11 till 1.0 e-12. Since the diffusion time t = 75-90 mn a slight decrease in their values of time is observed.

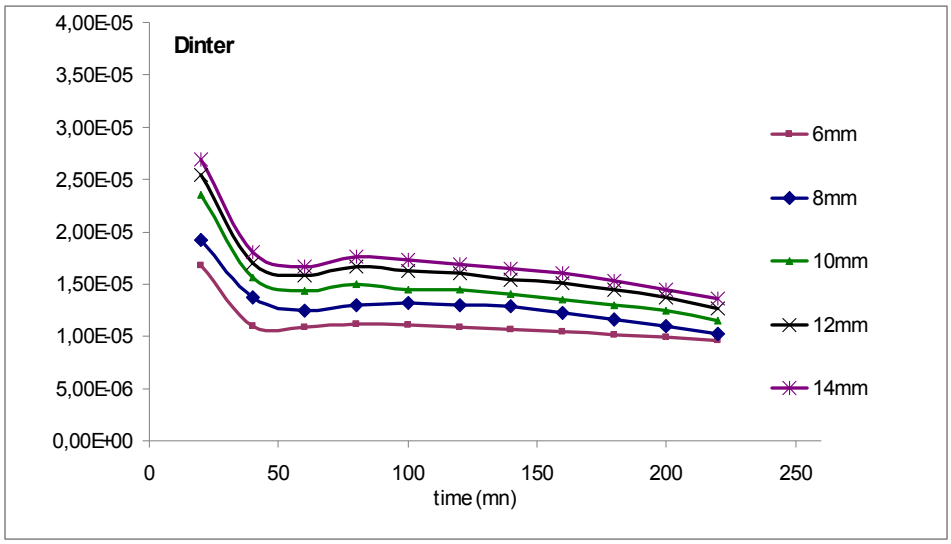


Fig. 6 Diffusion coefficients profiles $D_{\text{int}er_{2,k}}$ *in interparticles space for hexane in time for different positions in catalytic bed*

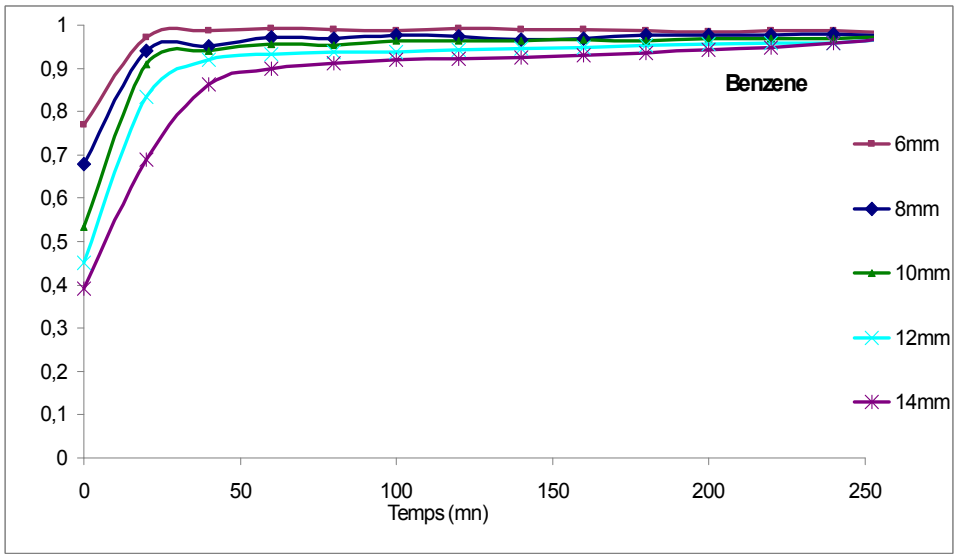


Fig. 7. Diffusion concentration distributions in the intercrystalites space in time at different positions of the catalytic bed for benzene

The hexane diffusion coefficient distributions in the macropore (fig. 6) vary in the range from 2.7 E-6 till 1.0 E5. Since the diffusion time t = 70-80 mn also relatively gentle pattern changing of their values and slightly higher at the end of diffusion is observed.

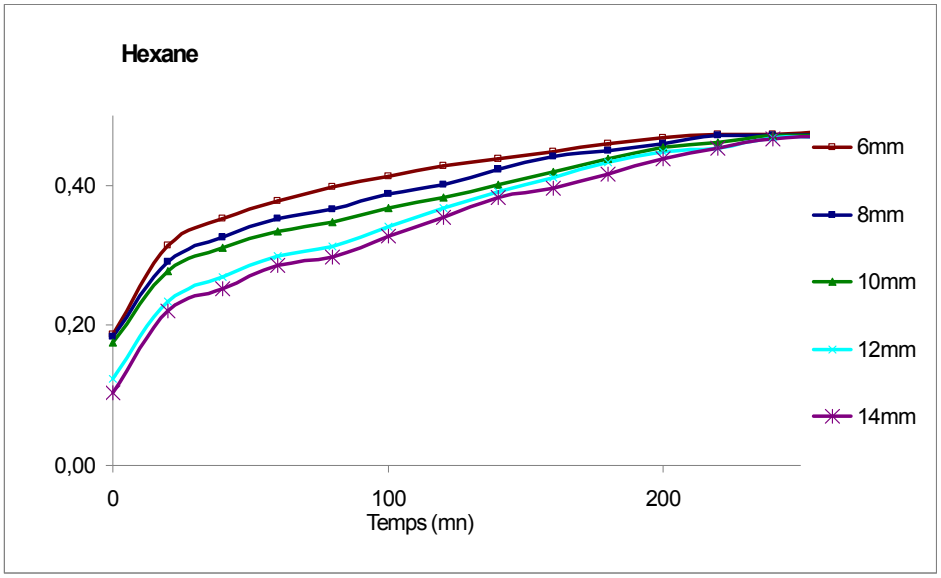


Fig. 8. Diffusion concentration distributions in the intercrystalites space in time at different positions of the catalytic bed for hexane

Fig.7, 8 shows concentration curves calculated for benzene and hexane in intercrystallites space with taking into account identified diffusion coefficients both for intra- and inter particles spaces. It is easily noticed that model distributions are very similar to experimental data approximations.

7. Conclusions

In this paper we propose the new effective identification procedures of diffusion coefficients for both diffused components (benzene and hexane) in intra- and inter crystallites spaces. Procedure implementation are achieved with using of gradient methods of the complex multi-component systems state control, mathematical models of competitive diffusion and mono diffusion in porous zeolite mediums and NMR imaging of the adsorbed masses distribution for each component in the zeolite crystallite bed. As result we obtain distributions of diffusion coefficients as a function of time for different positions along the microspores medium.

References

1. M. Fernandez, J. Kärger, D. Freude, A. Pampel, J.M. van Baten, R. Krishna Mixture diffusion in zeolites studied by MAS PFG NMR and molecular simulation. Microporous and Mesoporous Materials, 105, 124 (2007).

2. S. Leclerc, G. Trausch, B. Cordier, D. Grandclaude, A. Retournard, J. Fraissard, D. Canet Chemical shift imaging (CSI) by precise object displacement . Magn. Reson. Chem. 44, 311 (2006).
3. P. N'Gokoli-Kekele, M.-A. Springuel, J.-J. Bonardet, J.-M. Dereppe and J. Fraissard Use of 1H NMR imaging to study the diffusion and co-diffusion of gaseous hydrocarbons in HZSM-5 catalysts. Studies in Surface Science and Catalysis, 135, 93 (2001)
4. M. Petryk, S. Leclerc, D. Canet, J. Fraissard Modeling of gas transport in a microporous solid using a sclice selection procedure: Application to the diffusion of benzene in ZSM5. Catalysis Today, Elsevier B.V., – Volume 139, Issue 3. – P. 234-240. (2008)
5. M. Petryk, S. Leclerc, D. Canet, J. Fraissard Mathematical modeling and visualization of gas transport in a zeolite bed using a slice selection procedure. Diffusion Fundamentals. – Volume 4. – P. 11.1-11.23 (2007)
6. S.Lecler, M.Petryk, D.Canet, J.Fraissard Competitive Diffusion of Gases in a Zeolite Using Proton NMR and Sclice Selection Procedure. Catalysis Today, Elsevier B.V.. – Volume 187, Issue 1. – P. 104-107 (2012).
7. Sergienko I.V., Deineka V.S. Optimal Control of Distributed Systems with Conjugation Conditions. New York: Kluwer Academic Publishers. 400 p. (2005)
8. V. S. Deineka, M. R. Petryk, J. Fraissard Identifying kinetic parameters of mass transfer in components of multicomponent heterogeneous nanoporous media of a competitive diffusion system. Cybernetics and Systems Analysis; 47(5) (2011)
9. Ленюк М.П. Інтегральні перетворення Фур'є, Бесселя із спектральним параметром в задачах математичного моделювання масопереносу в неоднорідних середовищах / М.П. Ленюк, М.Р. Петрик. − К: Наукова думка, 2000. − 371с.

Надійшла у першій редакції 03.10.2014, в останній - 28.10.2014.

УДК 519.6:37.015.6

Применение теории сплайнов, построенных на неравномерной сетке узлов, в моделировании образовательных процессов

Е. В. Ярмош
Украинская инженерно-педагогическая академия, Украина

В статье построены и исследованы модели процесса формирования контингента студентов высшего учебного заведения с помощью сплайнов второго порядка. Рассмотрены два способа построения базисных сплайнов на неравномерной сетке узлов при разном количестве узлов сплайна (основных и вспомогательных). Приведены результаты вычислительного эксперимента.

Ключевые слова: *математическая модель, сплайн второго порядка, неравномерная сетка узлов, контингент студентов, цена обучения, рейтинг вуза.*

В статті побудовано та досліджено математичні моделі процесу формування контингенту студентів вищого навчального закладу за допомогою сплайнів другого степеня. Розглянуто два способи побудови базисних сплайнів на нерівномірній сітці вузлів за різної кількості вузлів сплайну (основних та опоміжних). Наведено результати обчислювального експерименту.

Ключові слова: *математична модель, сплайн другого степеня, нерівномірна сітка вузлів, контингент студентів, ціна навчання, рейтинг вищого навчального закладу.*

In the article, models of student contingent formation in tertiary institutions are developed and investigated with second-order splines. Two methods of constructing basis splines on irregular grid with different number of nodes (main and auxiliary) in spline are considered. The computational experiment results are provided.

Keywords: *mathematical model, second-order spline, irregular grid nodes, students' contingent, educational price, university ranking.*

1. Введение

Широкое применение в последние десятилетия сплайн-методов в инженерных расчетах свидетельствует о преимуществах сплайнов перед классическими методами. При этом широко используются билинейные, биквадратичные и бикубические сплайны.

На практике применяют сплайны, построенные на равномерной и неравномерной сетках узлов. Сегодня исследуется значительное количество процессов и явлений, данные о характеристиках которых представлены нерегулярно, что свидетельствует о перспективах исследования и использования в моделировании именно сплайнов, построенных на неравномерной сетке узлов. Построению сплайнов разных порядков на регулярной и нерегулярной сетках узлов посвящены работы Ю.Н. Субботина [1], В.Т. Шевалдина [2], Ю.С. Завьялова, Б.И. Квасова [3], В.А. Василенко [4], Н.П. Корнейчука [5], В.Л. Макарова, В.В. Хлобистова [6], О.Н. Литвина, А.В. Ткаченко [7].

В связи с переходом системы образования к рыночным методам функционирования становится актуальным, по мнению автора, исследование различных социально-экономических аспектов деятельности высших учебных заведений как особых субъектов рынка образовательных услуг с помощью сплайнов, построенных на неравномерной сетке узлов.

Целью данной работы является построение модели процесса формирования контингента студентов вуза с помощью указанных сплайнов второго порядка.

2. Изложение основного материала

Информация, получаемая от различных субъектов хозяйствования, как правило, носит нерегулярный характер. Не являются исключением данные рынка образовательных услуг Украины, что связано с влиянием большого количества факторов на деятельность вузов. В условиях демографического кризиса, который определил количество поступивших в вузы всех уровней аккредитации последних лет, ведения конкурентной ценовой борьбы за контингент студентов, законодательных изменений образовательного процесса рассмотрим процесс формирования контингента студентов вузов III-IV уровней аккредитации как специфическую задачу математического моделирования.

В работе [8] были рассмотрены математические модели процесса формирования контингента студентов, построенные с помощью кусочно-линейных сплайнов двух переменных и рациональных функций, где контингент, выраженный в относительной доле студентов контрактной формы обучения к их общему количеству, определяется как функция двух переменных: цены обучения на определенной специальности (p) и рейтинга вуза по соответствующему направлению подготовки (r), размещенных на нерегулярной сетке узлов (R_k, P_k), $k = \overline{1, M}$ (рис. 1). Рейтинг вузов принят по результатам рейтингования вузов Украины по методике ЮНЕСКО, согласно которой вузы сгруппированы по 10-ти уровням рейтинга, где 1 – наивысший рейтинг, 10 - наименьший. Для удобства вычислений значение рейтинга представлено в интервале $[0{,}1;1]$, то есть значение 0,1 соответствует десятому рейтингу, 1 - первому.

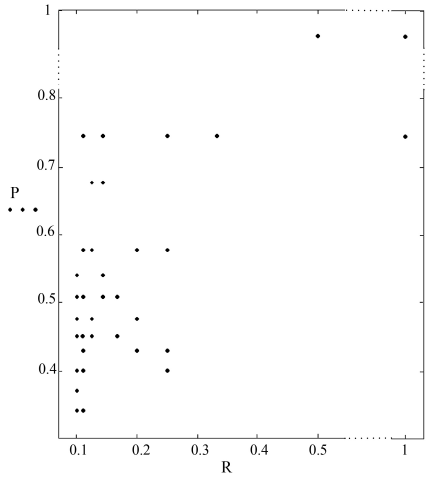


Рис. 1. Размещение узлов (R_k, P_k)

При построении модели использованы следующие обозначения:

1. $R_i = 1/RR_i$ ($R_i = 1$ при $RR_i = 1$, $i = \overline{1,M}$), RR_i - рейтинг вуза по направлению подготовки, M - количество исследуемых вузов.

2. $P_i = PP_i / PP_{\max}$, $PP_{\max} = \max\limits_{1 \le i \le M} PP_i$, $i = \overline{1,M}$, PP_i - стоимость обучения на конкретной специальности в исследуемом учебном году.

3. $F_i^{к} = \dfrac{FF_i^{к}}{FF_i^{б} + FF_i^{к}}$ или $F_i^{б} = \dfrac{FF_i^{б}}{FF_i^{б} + FF_i^{к}}$, соответственно $F_i^{к} + F_i^{б} = 1$,

$FF_i^{б}$, $FF_i^{к}$ - количество зачисленных студентов на бюджетную или за средства физических лиц (контрактная) формы обучения.

Рассмотрим случай, когда модель процесса формирования контингента студентов вуза от его рейтинга и цены обучения имеет вид

$$E(r,p) = \sum_{q=1}^{M} B_q(r) Q(p,q), \tag{1}$$

где $B_q(r) = \sum\limits_{k=1}^{N} S_2(r, R^{<k>}) C_{k,q}$, $C_{k,q}$ находим из условия $B_q(r_\mu) = \delta_{q,\mu}$, $1 \le q, \mu \le N$, $S_2(r, R^{<k>})$ - сплайн второго порядка на неравномерной сетке узлов вида [7].

Приведем явный аналитический вид для базисных сплайнов второго порядка на неравномерной сетке узлов, построенных в [7].

$$S_2(x,X) = \begin{cases} 0, x \le X_0 \\ \dfrac{(x-X_0)^2}{X_1 - X_0}, X_0 < x \le X_1 \\ X_2 - X_0 + \dfrac{(x-X_2)^2}{X_1 - X_2} - \dfrac{(x-X_1)^2}{X_2 - X_1} \dfrac{X_2 - X_0}{X_3 - X_1}, X_1 < x \le X_2 \\ \dfrac{(X_2 - X_0)(X_3 - X_2)}{X_3 - X_1} - \dfrac{(x-X_3)^2 - (X_3 - X_2)^2}{X_2 - X_3} \dfrac{X_2 - X_0}{X_3 - X_1}, X_2 < x < X_3 \\ 0, x \ge X_3 \end{cases} \tag{2}$$

Отметим, при $X_k = k, k = 0,1,2,3,4$ формула для $SS_2(x,X)$ совпадает с формулами для *B*-сплайнов второй степени на равномерной сетке узлов [7].

Графики сплайнов $B_q(r)$ на интервале [0,1] при $q = \overline{5,10}$ приведем на рис. 2.

Рассмотрим также случай, когда $S_2(r, R^{<k>})$ - сплайн 2-го порядка, построенный в [2] с использованием большего количества узлов сплайна (основных и вспомогательных) по сравнению с [6]. В работе [2] указано, что локальные полиномиальные сплайны порядка r минимального дефекта обычно строятся как линейные комбинации соответствующих B-сплайнов $B_{r,k}(x)$. А именно для функции f класса непрерывных функций локальный полиномиальный сплайн $S(x) = S(x, f)$ определяется так

$$S(x) = \sum_k c_k(f) B_{r,k}(x),$$

где при фиксированном $x = x^*$ коэффициенты $c_k(f)$ определяются значениями $f(\tau_k)$ в некоторой окрестности точки x^* и определяется носителем тех B-сплайнов $B_{r,k}(x)$, которые в точке x^* не равны нулю. Следует отметить, что самый простой и удобный с вычислительной точки зрения вариант $c_k(f) = f(\tau_k)$

$$S(x) = \sum_k f(\tau_k) B_{r,k}(x). \quad (3)$$

имеет невысокую точность.

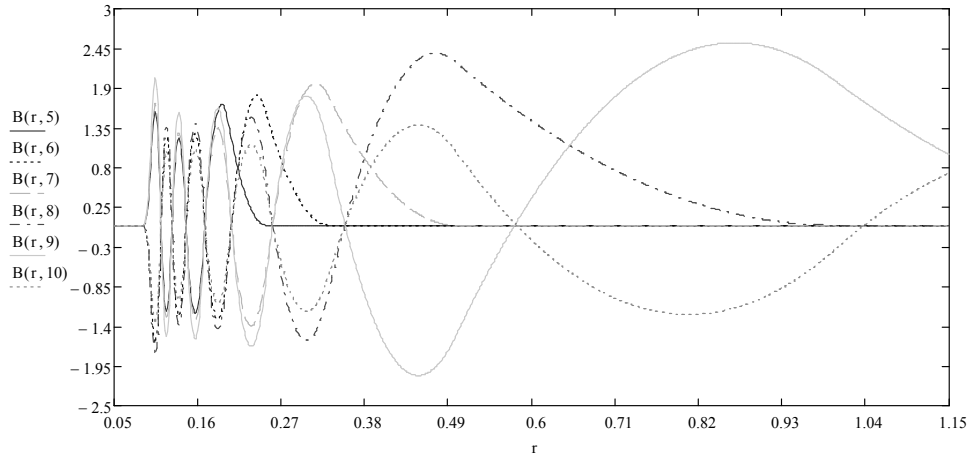


Рис. 2. График сплайна $B_q(r_\mu) = \delta_{q,\mu}$, $1 \le q, \mu \le N$, *простроенного с помощью базисного сплайна второго порядка вида (2), при* $q = \overline{5,10}$

В случае функций f, заданных на отрезке $[0,1]$, этим сплайном определяется линейный положительный оператор, который каждой непрерывной функции $f \in C[0,1]$ ставит в соответствие полиномиальный сплайн $S(x) \in S(\Delta_n)$ порядка r на сетке $\Delta_n : 0 = x_0 < x_1 < ... < x_{n-1} < x_n = 1$. Положительность оператора S означает, что если $f(x) \ge 0$, то $S(x, f) \ge 0$ для всех $x \in [0,1]$.

Также в работе [2] построен параболический сплайн $S(x) = S(x, f)$ для неравномерного размещения узлов в виде линейной комбинации функций типа B-сплайнов. Для построения автором рассмотрена на оси $\mathbf{R}$ бесконечная в обе стороны сетка узлов $... < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < ...$; $h_j = x_{j+1} - x_j (j \in \mathbf{Z})$, для функций $f \in W_\infty^2(\mathbf{R})$ и введено обозначение

$$[y_j, y_{j+1}, y_{j+2}] = f[x_j, x_{j+1}, x_{j+2}] = \frac{y_{j+2}}{h_{j+1}(h_{j+1} + h_j)} - \frac{y_{j+1}}{h_{j+1} h_j} + \frac{y_j}{h_j(h_{j+1} + h_j)}, \quad (4)$$

$j \in \mathbf{Z}$, для разделенной разности 2-го порядка по значениям функции $y_j = f(x_j)$ в точках x_j, x_{j+1} и x_{j+2}. Функции $f \in W_\infty^2(\mathbf{R})$ ставится в соответствие параболический сплайн

$$\begin{aligned} S(x) = S(x,f) = f(x_j) + \frac{h_{j-1}h_j}{4} f[x_{j-1}, x_j, x_{j+1}] + \\ + \frac{f(x_{j+1}) - f(x_{j-1})}{h_j + h_{j+1}}(x - x_j) + \frac{h_{j-1}}{h_j}(x - x_j)^2 f[x_{j-1}, x_j, x_{j+1}] + \\ + \left(\frac{h_{j+1}}{h_j} f[x_j, x_{j+1}, x_{j+2}] - \frac{h_{j-1}}{h_j} f[x_{j-1}, x_j, x_{j+1}] \right) \left(x - \frac{x_j + x_{j+1}}{2} \right)_+^2, \\ x \in [x_j, x_{j+1}] \, (j \in \mathbf{Z}) \end{aligned} \tag{5}$$

где $(x-\alpha)_+^2 = \max\{0, (x-\alpha)\}^2$. Для формулировки теоремы о свойствах функции $S(x)$ введено несколько вспомогательных функций (далее j - произвольное целое число). Пусть

$t_1 = \frac{x_{j-2} + x_{j-1}}{2}$, $t_2 = x_{j-1}$, $t_3 = \frac{x_{j-1} + x_j}{2}$, $t_4 = x_j$, $t_5 = \frac{x_j + x_{j+1}}{2}$, $t_6 = x_{j+1}$, $t_7 = \frac{x_{j+1} + x_{j+2}}{2}$

и сплайн на каждом интервале $[t_k, t_{k+1}]$, $k = \overline{1,6}$ задается такими функциями

$$\varphi_1(x) = \frac{(x - t_1)^2}{h_{j-1}(h_{j-1} + h_{j-2})}, \; x \in [t_1, t_2];$$

$$\varphi_2(x) = \frac{1}{h_{j-1} + h_{j-2}} \left(\frac{h_{j-2}}{4} + x - t_2 + \frac{h_{j-2}}{h_{j-1}^2}(x - t_2)^2 \right), \; x \in [t_2, t_3];$$

$$\varphi_3(x) = \varphi_2(x) - \frac{2h_{j-2} + h_{j-1}}{h_{j-1}^2(h_{j-1} + h_{j-2})}(x - t_1)^2, \; x \in [t_3, t_4];$$

$$\varphi_4(x) = \frac{3}{4} - \frac{1}{h_j^2}(x - t_4)^2, \; x \in [t_4, t_5];$$

$$\varphi_5(x) = \varphi_4(x) + \frac{2h_{j+1} + h_j}{h_j^2(h_j + h_{j+2})}(x - t_5)^2, \; x \in [t_5, t_6];$$

$$\varphi_6(x) = \frac{1}{h_{j+1}(h_{j+1} + h_j)}(x - t_7)^2, \; x \in [t_6, t_7].$$

Целесообразность использования сплайнов такого вида связана со свойствами процесса аппроксимации с использованием сплайнов, сформулированными в работе [2] в виде теоремы.

Теорема 1. Локальный процесс аппроксимации (5) имеет следующие свойства:

1. Наследует локально свойство монотонности исходных данных $\{y_i\}$ в том смысле, что

а) если $y_{j-1} \le y_j \le y_{j+1}$ (или $y_{j-1} \ge y_j \ge y_{j+1}$), то функция $S(x)$ не убывает (не возрастает) на промежутке $\left(x_j, \frac{x_j + x_{j+1}}{2}\right)$;

б) если $y_j \le y_{j+1} \le y_{j+2}$ (или $y_j \ge y_{j+1} \ge y_{j+2}$), то функция $S(x)$ не убывает (не возрастает) на промежутке $\left(\frac{x_j + x_{j+1}}{2}, x_{j+1}\right)$;

2. Наследует локально свойство выпуклости исходных данных, а именно:

а) если $[y_{j-1}, y_j, y_{j+1}] \ge 0$ ($[y_{j-1}, y_j, y_{j+1}] \le 0$), то функция $S(x)$ выпуклая вниз (вверх) на промежутке $\left(x_j, \frac{x_j + x_{j+1}}{2}\right)$;

б) если $[y_j, y_{j+1}, y_{j+2}] \ge 0$ ($[y_j, y_{j+1}, y_{j+2}] \le 0$), то функция $S(x)$ выпуклая вниз (вверх) на промежутке $\left(\frac{x_j + x_{j+1}}{2}, x_{j+1}\right)$;

3. Функция $S(x)$ непрерывна на всей оси $\mathbf{R}$, причем

$$S(x_j) = y_j + \frac{h_{j-1}h_j}{4}[y_{j-1}, y_j, y_{j+1}] \quad (j \in \mathbf{Z});$$

4. Функция $S(x)$ имеет на всей оси $\mathbf{R}$ непрерывную первую производную $S'(x)$, причем $S'(x_j) = \frac{1}{h_{j-1} + h_j}(y_{j+1} - y_{j-1})$ $(j \in \mathbf{Z})$;

5. а) для любой функции $f \in W_\infty^2[x_{j-1}, x_{j+1}]$ $(j \in \mathbf{Z})$ имеет место точное неравенство $\left|S''(x)\right| \le \frac{h_{j-1}}{h_j}$, $x \in \left(x_j, \frac{x_j + x_{j+1}}{2}\right)$;

б) для любой функции $f \in W_\infty^2[x_j, x_{j+2}]$ $(j \in \mathbf{Z})$ имеет место точное неравенство $\left|S''(x)\right| \le \frac{h_{j+1}}{h_j}$, $x \in \left(\frac{x_j + x_{j+1}}{2}, x_{j+1}\right)$;

6. При любом $x \in \mathbf{Z}$ справедлива формула $S(x) = \sum_j y_j B_j(x)$, причем при $x \in \left(\frac{x_{l-1} + x_l}{2}, \frac{x_l + x_{l+1}}{2}\right)$ $(l \in \mathbf{Z})$ эта сумма состоит из трех слагаемых, а именно

$$S(x) = y_{l-1}B_{l-1}(x) + y_l B_l(x) + y_{l+1}B_{l+1}(x).$$

Теорема 1 доказана в работе [2], а для случая равномерной сетки узлов Ю.Н. Субботиным [1].

Отметим, что можно построить локальные параболические сплайны с произвольными фиксированными узлами для отрезка [0,1] и в непериодическом случае. Пусть $\Delta_n : 0 = x_0 < x_1 < ... < x_{n-1} < x_n = 1$ - сетка узлов на отрезке [0,1]. Полагаем

$$S(x) = y_0 + \frac{y_1 - y_0}{h_0} x + \frac{h_1}{h_0}\left(x - \frac{x_1}{2}\right)_+^2 [y_0, y_1, y_2],\ x \in [0, x_1],$$

$$S(x) = y_{n-1} + \frac{y_{n-2} y_{n-1}}{4}[y_{n-2}, y_{n-1}, y_n] + \frac{y_n - y_{n-2}}{h_{n-1} + h_{n-2}}(x - x_{n-1}) +$$

$$+ \frac{h_{n-2}}{h_{n-1}}(x - x_{n-1})^2 [y_{n-2}, y_{n-1}, y_n] - \frac{h_{n-2}}{h_{n-1}}[y_{n-2}, y_{n-1}, y_n]\left(x - \frac{1 + x_{n-1}}{2}\right)_+^2,$$

$x \in [x_{n-1}, 1]$.

На отрезках $[x_j, x_{j+1}]$ ($j = 1,2,...,n-2$) функция $S(x)$ строится по формулам (5). Построенная таким образом на отрезке [0,1] функция $S(x)$ такая, что $S'(x) \in C[0,1]$. При этом сплайн $S(x)$ локально наследует свойства монотонности и выпуклости исходных данных $y_j = f(x_j)$, $j = 1,2,...,n$.

Важным является также утверждение приведенной ниже теоремы 2 о погрешности аппроксимации.

Теорема 2. Для любой функции $f \in W_\infty^2[x_{j-1}, x_{j+2}]$ ($j \in \mathbf{Z}$) для сплайна $S(x) = S(x, f)$, определенного формулой (5), имеет место точное неравенство

$$|f(x) - S(x)| \le \psi_j(x),\ x \in [x_j, x_{j+1}]\ \ (j \in \mathbf{Z}),$$

где $\psi_j(x) = \begin{cases} \psi_{j,1}(x), x \in \left[x_j, \dfrac{x_j + x_{j+1}}{2}\right] \\ \psi_{j,2}(x), x \in \left[\dfrac{x_j + x_{j+1}}{2}, x_{j+1}\right] \end{cases}$,

$$\psi_{j,1}(x) = \frac{1}{2}\left[(x - x_j)^2 \frac{h_{j-1} - h_j}{h_j} + (x - x_j)(h_j - h_{j-1}) + \frac{h_j h_{j-1}}{4}\right],$$

$$\psi_{j,2}(x) = \frac{1}{2}\left[(x - x_{j+1})^2 \frac{h_{j+1} - h_j}{h_j} - (x - x_{j+1})(h_j - h_{j+1}) + \frac{h_j h_{j+1}}{4}\right].$$

Обе функции $\psi_{j,1}(x)$ и $\psi_{j,2}(x)$ являются неотрицательными в областях своего определения.

Рассмотрим случай построения математической модели процесса формирования контингента студентов с помощью формулы вида (1), где $B_q(r)$ получаем с помощью сплайна вида (5).

Графики сплайна $B_q(r)$ на интервале [0,1] при $q = \overline{5,10}$ приведем на рис. 3.

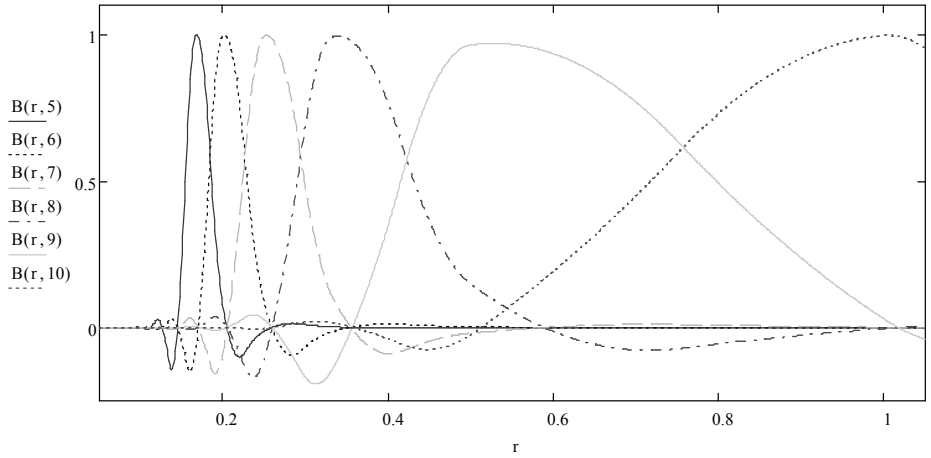


Рис. 3. Графики базисных сплайнов $B_q(r_\mu) = \delta_{q,\mu}$, $1 \le q, \mu \le N$ *вида (5) при* $q = \overline{5,10}$

При построении модели процесса формирования контингента студентов вуза при помощи рассмотренных сплайнов, проведен расчет для двух случаев задания функции цены:

1) полиномом вида $Q1(p,k) = \sum_{l=1}^{N_k} F(r_k, p_l) \prod_{\substack{q=1 \\ q \ne l}}^{N_k} \frac{p - P_q}{P_l - P_q}$;

2) в виде $Q2_k(p) = b_{0,k} + b_{1,k} p + b_{2,k} p^2$ для заданного значения рейтинга r, где $b_{i,k}, i = \overline{0,2}$ находятся из условия наилучшего приближения методом наименьших квадратов экспериментальных данных $F(r_k, p_l)$.

Отметим, что результаты вычислительного эксперимента модели (1) позволяют проводить анализ критических значений цены, при которой спрос на исследуемую специальность, при известном значении рейтинга, падает. Также полученные зависимости подтверждают гипотезу, что при фиксированном значении цены и при росте рейтинга увеличивается количество зачисленных студентов, которое по условиям модели выражается в доле студентов контрактной формы обучения к их общему количеству.

3. Выводы

Построенные модели формирования контингента студентов с помощью рассмотренных сплайнов второго порядка позволяют проводить эконометрический анализ зависимости процесса формирования контингента студентов вуза от изменений его рейтинга и цены обучения. Сложности построения прогнозов с использованием полученных зависимостей связаны с проблемой исходных данных, шаг между которыми мал. Кроме того, в рейтинге вузов Украины всего 5 вузов, которые входят в группу высоких рейтингов (от 1 до 3), а высокий уровень цены характерен для вузов рейтинга от 1 до 9.

Таким образом, анализ результатов вычислительного эксперимента, проведенного на основании указанных выше экспериментальных данных по Украине, для одной из специальностей показывает, что использование квадратичных сплайнов по одной переменной на неравномерной сетке узлов для приближения зависимости принятых студентов $E(r, p)$ позволяет найти области монотонности, а также локальные экстремумы (максимумы и минимумы). Это приводит к возможности найти зависимость между переменными r и p, при которых функция $E(r, p)$ будет иметь наибольшее значение.

ЛИТЕРАТУРА

1. Субботин Ю.Н. Исследование свойств монотоности и выпуклости при локальной аппроксимации // ЖВМиМФ. – 1993. – 33, № 7. – С. 996–1003.
2. Шевалдин В.Т. Аппроксимация локальними параболическими сплайнами с произвольным расположением узлов // Сиб. журн. вычисл. математики/ РАН. Сиб. отд-ние.-Новосибирск. – 2005. – 8, № 1. – С. 77–88.
3. Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л. Методы сплайн-функций. – М.: Наука, 1980. – 220 с.
4. Василенко В.А. Сплайн-функции: теория, алгоритмы, программы. – Новосибирск: Наука, 1983. – 215 с.
5. Корнейчук Н.П. Сплайны в теории приближения. – М.: Наука, 1984. – 352 с.
6. Макаров В.Л., Хлобыстов В.В. Интерполирование операторов. – К.: Наукова думка, 2000. – 406 с.
7. Литвин О.М., Ткаченко О.В. Математичне моделювання процесів інтерполяційними сплайнами на нерегулярній сітці вузлів // Доповіді НАН України. – 2010. - №1. – С.34-39.
8. Литвин О., Ярмош О. Математичне моделювання процесу формування контингенту студентів ВНЗ // Вісник Львівського університету. Серія прикладна математика та інформатика. – 2012. – Випуск 18. –С.191-200.

Надійшла у першій редакції 22.01.2014, в останній - 19.06.2014.

UDC 519.24

The Sequences with Stationary differences

Sh. Assadi, Gh. Jouja, F. Farhood
Dept of Mathematics, Faculty of Science, University of Aleppo, Syria

This paper studies nonstationary random sequences with stationary increments. General representations are obtained for their correlation function and correlation differences. The general case is studied for non-stationary sequence, which is the solution of difference equation with stationary right-hand side. Derived spectral representations prove that such sequences are harmonizable. The general representation of solution correlation function is obtained for equation, the right-hand side of which is a non-stationary sequence of finite non-stationarity rank.

Key words: *difference equation, stationary increment, the correlation function, the correlation difference, spectral expansion, harmonizable, non-stationarity rank.*

У статті вивчаються нестаціонарні випадкові послідовності зі стаціонарними приростами. Для них отримані загальні зображення кореляційної функції і кореляційних різниць. Досліджено загальний випадок нестаціонарної послідовності, що є рішенням різницевого рівняння зі стаціонарною правою частиною. Отримано спектральні зображення, що доводять гармонізуємість таких послідовностей. Для рівняння, права частина якого є нестаціонарна послідовність кінцевого рангу нестаціонарності, отримано загальне зображення кореляційної функції розв'язку.

Ключові слова: *різницеве рівняння, стаціонарний приріст, кореляційна функція, кореляційний різниця, спектральне зображення, гармонізуємість, ранг нестаціонарності.*

В статье изучаются нестационарные случайные последовательности со стационарными приращениями. Для них получены общие представления корреляционной функции и корреляционных разностей. Исследован общий случай нестационарной последовательности, являющейся решением разностного уравнения со стационарной правой частью. Получены спектральные представления, доказывающие гармонизируемость таких последовательностей. Для уравнения, правая часть которого является нестационарной последовательностью конечного ранга нестационарности, получено общее представление корреляционной функции решения.

Ключевые слова: *разностное уравнение, стационарное приращение, корреляционная функция, спектральное разложение, гармонизируемость, ранг нестационарности.*

Introduction

Linear random difference equations with discrete variables, the right side of which is a random function having effect of discrete white noise [1], are widely used in modeling of random processes with discrete times. The type of non-stationarity of equation, the right side of which is a random sequence that belongs to one of random sequence classes, have not been studied yet.

To study sequences in *Hilbert* space, an operator concept was suggested and adequate non-stationarity characteristics were introduced in [2,3].

The related close subjects see also [4-6].

1. Cauchy function

For further investigation, we will use the representation of linear difference equations of order k with constant coefficients in the form:

$$\begin{cases} L_k\, x(n) = \sum_{j=0}^{k} \alpha_j\, \Delta_j\, x(n) = f(n) \\ \alpha_k = 1 \quad , \quad \Delta_0\, x(n) = x(n) \end{cases} \tag{1.1}$$

where n is a natural number, x is the unknown function, and $\alpha_j \left(j = \overline{0\,,\,k} \right)$ are known real numbers;
or in equivalent form:

$$\begin{cases} L_k\, x(n) = \sum_{j=0}^{k} a_j\, x(n+j) = f(n) \\ a_k = 1 \end{cases} \tag{1.2}$$

Definition:

For each given natural number m, the Cauchy function $\Phi(n,m)$ is defined as solution of homogenous linear difference equation with constant coefficients

$$\begin{cases} L_k\, \Phi(n,m) = \sum_{j=0}^{k} \alpha_j\, \Delta_j\, \Phi(n,m) = 0 \\ \alpha_k = 1 \quad , \quad \Delta_0\, \Phi(n,m) = \Phi(n,m) \end{cases} \tag{1.3}$$

under such initial conditions

$$\Phi(n,m)\big|_{n=m} = 0 \ , \ \Delta_1\, \Phi\big|_{n=m} = 0 \ , \ \ldots \ , \Delta_{k-2}\, \Phi\big|_{n=m} = 0 \ , \ , \Delta_{k-1}\, \Phi\big|_{n=m} = 1 \tag{1.4}$$

Let us consider some examples of Cauchy function for linear difference equations with constant coefficients [7].

Example 1

Let us consider the difference equation

$$\Delta_1\, x(n) + \alpha_0 x(n) = f(n) \tag{1.5}$$

The expression for its Cauchy is following: $\Phi(n,m) = (1-\alpha_0)^{n-m}$.

Example 2

In the case of linear difference equation of second order

$$\Delta_2\, x(n) + \alpha_1\, \Delta_1\, x(n) + \alpha_0\, x(n) = f(n) \tag{1.7}$$

using the equality $\Delta_k\, x(n) = \sum_{j=0}^{k} (-1)^{k-j} C_k^j\, x(n+j)$ we can transform (1.7) to the form

$$x(n+2) + (\alpha_1 - 2)\, x(n+1) + (1 - \alpha_1 + \alpha_0)\, x(n) = f(n) \tag{1.8}$$

which is a linear difference equation of second order with constant coefficients.
Cauchy function $\Phi(n,m)$ is that solution of difference equation

$$\Phi(n+2,m) + (\alpha_1 - 2)\Phi(n+1,m) + (1 - \alpha_1 + \alpha_0) = 0 \tag{1.9}$$

which satisfies this two initial conditions

$$\Phi(n,m)\big|_{n=m}=0 \qquad , \qquad \Delta_1\Phi(n,m)\big|_{n=m}=1$$

$$\begin{cases} c_1(m)\lambda_1{}^n + c_2(m)\lambda_2{}^n = 0 \\ c_1(m)\left(\lambda_1{}^{m+1} - \lambda_1{}^m\right) + c_2(m)\left(\lambda_2{}^{m+1} - \lambda_2{}^m\right) = 1 \end{cases}$$

Thus, for $\Phi(n,m)$ we have $$\Phi(n,m)=\frac{1}{\lambda_1-\lambda_2}\lambda_1{}^{n-m}+\frac{1}{\lambda_2-\lambda_1}\lambda_2{}^{n-m}$$

Remark:

If the roots λ_1 and λ_2 are complex, they can be written in the form $\lambda_1=\rho e^{i\theta}$ and $\lambda_2=\rho e^{-i\theta}$, and making this substitution, we bring the Cauchy function to the form

$$\Phi(n,m)=\rho^{n-m-1}\frac{\sin\left((n-m)\theta\right)}{\sin\theta}=c\sin\left((n-m)\theta\right)$$

Further we prove several general theorems.

Theorem 1

Cauchy function $\Phi(n,m)$ for linear difference equation (1.3) of order k with homogenous constant coefficients, which satisfies the initial conditions (1.4), is a function of the difference $(n-m)$.

Theorem 2

The difference equation (1.1) has such a solution:

$$x(n)=\sum_{j=1}^{n-1}\Phi(n-1,\,j)\,f(j) \qquad (1.11)$$

where Cauchy function $\Phi(n,m)$ is a solution of homogenous difference equation (1.3) with initial conditions (1.4) [7].

Application of theorem (2) to both examples (1) and (2)

Due to mentioned above, we can transform solution of equation (10) to the form

$$x(n)=\sum_{j=1}^{n-1}(1-\alpha_0)^{n-1-j}f(j)$$

Solution of the equation (12) gets the form:

$$x(n)=\sum_{j=1}^{n-1}\left(\frac{1}{\lambda_1-\lambda_2}\lambda_1{}^{n-1-j}+\frac{1}{\lambda_2-\lambda_1}\lambda_2{}^{n-1-j}\right)f(j)$$

If n_0 is any integer then solution of non-homogenous equation (1.1) has the form

$$x(n)=\sum_{j=n_0}^{n-1}\Phi(n-1,\,j)\,f(j) \qquad (1.12)$$

where $\Phi(n,m)$ is Cauchy function.

So we have found that if $\lambda_k, \ldots, \lambda_2, \lambda_1$ are the roots of characteristic equation and they all are different then the general form of solution for (1.1) is

$$\begin{cases} x(n) = \sum_{j=1}^{k} \gamma_j \ \lambda_j^{n-n_0} \ + \sum_{j=n_0}^{n-1} \Phi(n-1, j) \ f(j) \\ \gamma_j = \dfrac{1}{\prod\limits_{\substack{\ell<i \\ \ell=j}} (\lambda_i - \lambda_\ell)} \end{cases} \qquad (1.15)$$

It is obvious that if $\left(j = \overline{1,k}\right) \quad |\lambda_j| < 1$, then for $n_0 \to -\infty$ we obtain:

$$x(n)_{stat} = \sum_{j=-\infty}^{n-1} \Phi(n-1, j) \ f(j) \qquad (1.16)$$

This solution of difference equations (1.1) is called *steady state*.

We can consider each of non-homogenous stationary equations (1.1) and (1.2) as stochastic linear difference equation, the right side of which is a stochastic function $f(n)$ that behaves as discrete white noise $[1]$.

In this paper, we consider $f(n)$ as random sequence belonging to one of the random sequence classes.

2. Analytic approach to difference equations with random right side of the form $L_k \ \eta(n) = \xi(n)$

Let L_k be a linear difference operator having real constant coefficients, and let us consider the difference equation

$$L_k \ \eta(n) = \xi(n) \qquad (2.1)$$

We have shown above that if all roots of characteristic equation for difference equation (2.1) are different and their absolute value is less than one, then the stationary solution will be given by:

$$\eta(n)_{stat} = \sum_{j=-\infty}^{n-1} \Phi(n-1, j) \ \xi(j) \qquad (2.2)$$

where $\Phi(n, m)$ is Cauchy function that satisfies the difference equation (1.3) and initial conditions (1.4).

When the correlation function for equation right side is given the correlation function for solution has the following appearance

$$K_{\eta\eta}(n, m) \ = \sum_{j,\ell=-\infty}^{n-1, m-1} \Phi(n-1, j) \ \overline{\Phi(m-1, \ell)} \ K_{\xi\xi}(j, \ell) \qquad (2.3)$$

and according to the theorem (1):

$$\Phi(n-1, j) \ = \ \Phi(n-1-j) \qquad (2.4)$$

Thus the relationship (2.3) takes form:

$$K_{\eta\eta}(n,m) = \sum_{j,\ell=-\infty}^{n-1,m-1} \Phi(n-1-j)\ \overline{\Phi(m-1-\ell)}\ K_{\xi\xi}(j,\ell) \qquad (2.5)$$

In addition, since the sequence $\xi(n)$ is stationary, then:

$$K_{\xi\xi}(j,\ell) = K_{\xi\xi}(j-\ell)$$

Consequently, the correlation function of the sequence $\eta(n)$ takes the form:

$$K_{\eta\eta}(n,m) = \sum_{j,\ell=-\infty}^{n-1,m-1} \Phi(n-1-j)\ \Phi(m-1-\ell)\ K_{\xi\xi}(j-\ell) \qquad (2.6)$$

Using the following transformation: $n-1-j=n_1 \ \Rightarrow\ j=n-1-n_1$

$$m-1-\ell=m_1 \ \Rightarrow\ \ell=m-1-m_1$$

we obtain: $K_{\eta\eta}(n,m) = \sum_{n_1,m_1=+\infty}^{0,0} \Phi(n_1)\,\overline{\Phi(m_1)}\,K_{\xi\xi}(n-m-(n_1-m_1)) = f(n-m)$

which means that in this case the sequence $\eta(n)$ is stationary too.

Therefore, we have proved the following theorem.

Theorem 3:

If L_k is linear difference operator with real constant coefficients, and the sequence $\xi(n)$ is stationary, then the steady-state solution of difference equation, is also stationary.

According to the theorem (1), the equation (2.1) has solution given by the following expression:

$$\eta(n) = \sum_{j=1}^{n-1} \Phi(n-1-j)\,\xi(j)$$

Consequently, for initial conditions, we have:

$$K_{\eta\eta}(n,m) = \sum_{j,\ell=1}^{n-1,m-1} \Phi(n-1-j)\ \overline{\Phi(m-1-\ell)}\ K_{\xi\xi}(j,\ell)$$

For the case when coefficients of operator L_k are constant and roots of characteristic equation are different, we obtain the following expression for Cauchy function:

$$\Phi(n,m) = \Phi(n-m) = \sum_{p=1}^{k} \gamma_p\ \lambda_p^{n-m} \quad ; \quad \gamma_p = \frac{1}{\prod\limits_{\substack{j<i \\ j=p}} (\lambda_i - \lambda_j)}$$

Now we will shall consider in general the case when $\xi(n)$ is a stationary sequence. In this case, the correlation function of sequence $\eta(n)$ takes the form

$$K_{\eta\eta}(n,m) = \sum_{j,\ell=1}^{n-1,m-1} \sum_{p,q=1}^{k} \gamma_p \overline{\gamma_q}\ \lambda_p^{n-1-j}\ \overline{\lambda_q^{m-1-\ell}} \int_0^{2\pi} e^{i\lambda(j-\ell)}\ dF(\lambda)$$

or $$K_{\eta\eta}(n,m) = \int_0^{2\pi} \sum_{p,q=1}^{k} \gamma_p \overline{\gamma_q} \sum_{j=1}^{n-1} \lambda_p^{n-1-j} e^{i\lambda j} \sum_{\ell=1}^{m-1} \overline{\lambda_q^{m-1-\ell}} e^{-i\lambda\ell} dF(\lambda)$$

Supposing that: $\chi_p(n,\lambda) = \sum_{j=1}^{n-1} \lambda_p^{n-1-j} e^{i\lambda j}$

we get: $k_{\eta\eta}(n,m) = \int_0^{2\pi} \theta(\lambda,n)\overline{\theta(\lambda,m)}\, dF(\lambda)$

where: $\theta(\lambda,n) = \sum_{p=1}^{\kappa} \gamma_p \chi_p(n,\lambda),\ \Delta F(\lambda) = \|\Delta E_\lambda\|^2$

E_λ is the solution of equation

$$\chi_p(n,\lambda) = \sum_{j=1}^{n-1} \lambda_p^{n-1-j} e^{i\lambda j} = \lambda_p^{n-1} \sum_{j=1}^{n-1} \left(\frac{e^{i\lambda}}{\lambda_p}\right)^j =$$

$$= \lambda_p^{n-1} \frac{\left(\frac{e^{i\lambda}}{\lambda_p}\right)^n - \frac{e^{i\lambda}}{\lambda_p}}{\frac{e^{i\lambda}}{\lambda_p} - 1} = \frac{e^{i\lambda n} - e^{i\lambda}\lambda_p^{n-1}}{e^{i\lambda} - \lambda_p}$$

Supposing that $F'(\lambda)$ exists and equals to $f(\lambda)$, we can conclude that the correlation function can be written in the form:

$$K_{\eta\eta}(n,m) = \int_0^{2\pi} \tilde{\theta}(\lambda,n)\,\overline{\tilde{\theta}(\lambda,m)}\, d\lambda$$

where $\tilde{\theta}(\lambda,n) = \theta(\lambda,n)\sqrt{f(\lambda)}$

(i. e. the sequence $\eta(n)$ is harmonizable [10]).

So we have proved the following theorem:

Theorem 4:

Let L_k be a linear difference operator with constant coefficients, all roots of characteristic equation $L_k\eta(n) = 0$ are different, and sequence $\xi(n)$ where $L_k\eta(m) = \xi(n)$ is stationary. Then the correlation function of process $\eta(n)$ takes form:

$$K_{\eta\eta}(n,m) = \int_0^{2\pi} \tilde{\theta}(\lambda,n)\overline{\tilde{\theta}(\lambda,m)}\, d\lambda$$

where: $$\tilde{\theta}(\lambda,n) = \theta(\lambda,n)\sqrt{f(\lambda)}$$

Application:

Let $\tilde{\xi}(n)=\sqrt{2\pi}\,\tilde{\theta}\left(n,\xi_0(w)\right)$ be a random sequence, where $\xi_0(w)$ is uniform over the interval $[0,2\pi]$. Then the correlation function of this sequence is:

$$\mathrm{M}\,\tilde{\xi}(n)\;\overline{\tilde{\xi}(m)} = 2\pi\int_0^{2\pi}\tilde{\theta}(n,\lambda)\;\overline{\tilde{\theta}(m,\lambda)}\,P(\lambda)\,d\lambda$$

where $P(\lambda)=\dfrac{1}{2\pi}$, thus it will be: $\mathrm{K}_{\tilde{\xi}\,\tilde{\xi}}(n,m) = \mathrm{K}_{\eta\eta}(n,m)$

This means that the two sequences $\xi(n)$ and $\tilde{\xi}(n)$ are unitary equivalent $[2,3]$.

3. The Relation between correlation differences of two sequences $\xi(n)$ and $\eta(n)$ in case of stationary solution

Let us rearrange the stationary equation $L_k\,\eta(n) = \xi(n)$ in the form

$$L_k^{\eta(n)}\,\eta(n) = \xi(n)$$

and suppose that $\Phi(n-m)$ is Cauchy function of L_k difference operator having constant real coefficients. Then we find that the correlation function of sequence $\xi(n)$ can be written in the form:

$$L_k^{\eta(n)}\;L_k^{\eta(m)}\left\langle\eta(n)\,,\eta(m)\right\rangle = \left\langle\xi(n),\xi(m)\right\rangle$$

and thus

$$L_k^{\eta(n)}\;L_k^{\eta(m)}\quad \mathrm{K}_{\eta\eta}(n,m) = \mathrm{K}_{\xi\zeta}(n,m)$$

The steady state solution of this equation is

$$\mathrm{K}_{\eta\eta}(n,m) = \sum_{p=-\infty}^{n-1}\sum_{q=-\infty}^{m-1}\Phi(n-1-p)\;\overline{\Phi(m-1-q)}\,\mathrm{K}_{\xi\zeta}(p,q)$$

we can obtain the same representation if we write stationary solution of difference equation of the form:

$$\eta(n) \;=\; \sum_{p=-\infty}^{n-1}\Phi(n-1,p)\;\xi(p)$$

Let $\mathrm{K}_{\eta\eta}(n,m)=\left\langle\eta(n)\;,\;\eta(m)\right\rangle_{H_\eta} = \mathrm{M}\,\eta(n)\,\overline{\eta(m)}$

then $\mathrm{K}_{\eta\eta}(n,m)=\sum_{p=-\infty}^{n-1}\sum_{q=-\infty}^{m-1}\Phi(n-1-p)\,\overline{\Phi(m-1-q)}\;\mathrm{K}_{\xi\xi}(p,q)$

Replacing n with $n+1$ in equation: $L_k^{\eta(n)}\quad\eta(n)\;=\xi(n)$

we have: $L_k^{\eta(n)}\,L_k^{\eta(m)}\,\mathrm{K}_{\eta\eta}(n,m) = \mathrm{K}_{\xi\xi}(n,m)$

$$L_k^{\eta(n)}\,L_k^{\eta(m)}\,\mathrm{K}_{\eta\eta}(n+1,m+1) = \mathrm{K}_{\xi\xi}(n+1,m+1)$$

therefore $L_k^{\eta(n)}\,L_k^{\eta(m)}\,W_{\eta\eta}(n,m) = W_{\xi\xi}(n,m)$

Consequently the steady state will be:

$$W_{\eta\eta}(n,m) = \sum_{p=-\infty}^{n-1}\sum_{q=-\infty}^{m-1}\Phi(n-1-p)\,\overline{\Phi(m-1-q)}\;W_{\xi\xi}(p,q)$$

Setting $W_{\eta\eta}(n,m) = \varphi(n)\,\overline{\varphi(m)}$ we get that:

$$W_{\eta\eta}(n,m) = \sum_{p,q=-\infty}^{n-1,m-1} \Phi(n-1-p)\ \ \overline{\Phi(m-1-q)}\ \ \varphi(p)\,\overline{\varphi(q)} = \psi(n)\ \ \overline{\psi(m)}$$

where $\psi(n) = \sum_{p=-\infty}^{n-1} \Phi(n-1-p)\ \varphi(p)$

So we have proved the following theorem.

Theorem 5

In steady state of difference equation $L_k\,\eta(n) = \xi(n)$, the correlation differences of two sequences $\eta(n), \xi(n)$ are tied by relation $W_{\eta\eta}(n,m) = \psi(n)\,\overline{\psi(m)}$

where $\psi(n) = \sum_{p=-\infty}^{n-1} \Phi(n-1-p)\ \varphi(p)$ and $\Phi(n,m)$ are Cauchy Functions of the difference operator L_k with real constant coefficients.

Definition

The non-stationary random sequence $\xi(n)$ is called a dissipative sequence if all the quadratic forms of this type

$$\sum_{n,m=0}^{N} W(n,m)\lambda_n\,\overline{\lambda_m}$$

are non-negative

From this definition, it follows that if the random sequence is dissipative then the quadratic forms will be non-increasing sequence OT p.

$$\sum_{n,m=0}^{N} K(n+p,m+p)\lambda_n\,\overline{\lambda_m}$$

In a special case, the sequence $K(n,n)$ is non-increasing sequence, from which it follows that the limit exists.

$$\lim_{n\to\infty} K(n,n) = \sigma_\infty^2$$

It is possible that $1)\sigma_\infty^2 = 0;\ \ 2)\sigma_\infty^2 > \infty$

In the first case, the sequence is called asymptotically damped, and in the second case is called asymptotically undamped

Retuning to equation (3.1), in addition to what we have supposed above, we assume that $\xi(n)$ is non-stationary random sequence of the order r, dissipative, and convergent damped [3].

Easily, we find that the correlation function can be expressed by relation:

$$\mathrm{K}_{\xi\xi}(n,m) = \sum_{\alpha=1}^{r}\sum_{r=0}^{\infty}\psi_\alpha(n+\tau)\,\overline{\psi_\alpha(m+\tau)}$$

Using (3.5) we obtain:

$$\mathrm{K}_{\eta\eta}(n,m) = \sum_{\alpha=1}^{r}\sum_{r=0}^{\infty}\sum_{j,\ell=-\infty}^{n-1,m-1} \Phi(n-1-j)\ \overline{\Phi(m-1-\ell)}\ \psi_\alpha(j+\tau)\ \overline{\psi_\alpha(\ell+\tau)}$$

where

$$= \sum_{\alpha=1}^{r}\sum_{r=0}^{\infty} \widetilde{\psi_\alpha}(n+\tau)\,\overline{\widetilde{\psi_\alpha}(m+\tau)}$$

$$\widetilde{\psi_\alpha}(n+\tau) = \sum_{j=-\infty}^{n-1} \Phi(n-1-j)\,\psi_\alpha(j+\tau)$$

By rearrangement $(j = n-1-n_1)$ $n-1-j = n_1$

we get: $$\widetilde{\psi_\alpha}(n+\tau) = \sum_{n_1=0}^{\infty} \Phi(n_1)\,(n-1-n_1+\tau)$$

Easily we can see that the non-stationarity order of sequence $\eta(n)$ equals to r

Application:

Let $x(n+1) = a\,x(n) + f(n)$ be the first order difference equation.

The solution of this equation, which meets the initial conditions.

$$x(n_0)\big|_{n=n_0} = x_0$$

is given by the relation $x(n) = a^{n-n_0}\ \ x_0 + \sum_{j=n_0}^{n-1} a^{n-(j+1)}\ f(j)$

If $|a| < 1$ the stationary solution takes this form

$$x_{stat}(n) = \sum_{j=-\infty}^{n-1} a^{n-1-j}\ f(j) = a^{n-1} \sum_{j=-\infty}^{n-1} \frac{f(j)}{a^j}$$

This series is convergent when n_0 tends to $-\infty$ if

$$\lim_{j\to-\infty} \frac{f(j+1)}{a^{j+1}} \cdot \frac{a^j}{f(j)} = \frac{1}{a} \lim_{j\to-\infty} \frac{f(j+1)}{a^{j+1}} < 1$$

which means $\lim_{j\to-\infty} \dfrac{f(j+1)}{f(j)} < \alpha$

Easily we find that the correlation function of this sequence is given by expression

$$\mathrm{K}_{xx}(n,m) = \sum_{p=-\infty}^{n-1}\sum_{q=-\infty}^{m-1} a^{n+m-p-q-2}\ \mathrm{K}_{ff}(p,q)$$

If $\mathrm{K}_{ff}(p,q) = \sum_{j=0}^{\infty} \phi(p+j)\ \overline{\phi(q+j)}$

then $\mathrm{K}_{xx}(n,m) = \sum_{j=0}^{\infty} \psi(n,j)\ \overline{\psi(m,j)}$

where $\psi(n,j) = \sum_{p=-\infty}^{n-1} a^{n-p-1}\ \phi(p+j)$

Using equality $p+j=\ell$ for substitution we get

$$\psi(n,j)=\sum_{j=-\infty}^{n+j-1} a^{n+j-\ell-1}\,\phi(\ell)$$

Consequently, if it is $\psi(n,j)=\psi(n+j)$, therefore the non-stationary order equals to 1.

Conclusion

In this study a new class of non-stationary sequences was introduced. An operator approach was applied, harmonizability of mentioned sequences was shown, and general representations are obtained for their correlation function and correlation differences. This makes possible modeling of random sequences having different nature of non-stationarity, as well as sequence restoration only by its spectrum.

One can use such sequences in investigations of transient modes of discrete control systems in both cases: when noise or useful signals are non-stationary random sequences.

The authors express deep acknowledgment to Prof Yantsevich Artem A for valued and useful guidance offered during this work.

REFERENCES

1. Astrom. K .J. Introduction to Stochastic Control Theory / K. J. Astrom. – New York: Academic press, 1970. – 320 pp.
2. Livshits M.S.; Yantsevich A.A. Operator Colligations in Hilbert Spaces / M. S. Livshits; A. A.Yantsevich. – New York: Wiley and Sons, 1979. – 211 pp.
3. Yantsevich A .A. Nonstationary Sequence in Hilbert Space I . Correlation theory // Journal of Soviet Mathematics. 1990. – vol. 48, No. 5. – P. 615 – 618.
4. A. A. Yantsevitch, A. Yu. Petrova. The spectral theory of some classes of random vector functions. // Radioelectronics and informatics, Kh. KNURE – 2007. – P. 37-40.
5. Ye. A. Kogut, Z. F. Nazyrov, A. A. Yantsevitch. About some class of linear discrete systems. // Bulletin of V. Karazin Kharkiv National University, – 2012. Series «Mathematical Modelling. Information Technology. Automated Control Systems», Issue 20. – P. 92-102.
6. A. V. Korobskaia, Z. F. Nazyrov, A. A. Yantsevitch. A class of revolutionarily representable random processes. // Bulletin of V. Karazin Kharkiv National University, – 2012. Series «Mathematical Modelling. Information Technology. Automated Control Systems», Issue 19. – P. 184-197.
7. Peterson A; Schneider J. The Cauchy Function for n^{th} Order Linear Difference Equations // Journal Of Mathematics. 1995. – Volume 25, Number 1 Winter.
8. Rooznov Ju. A. Stationary Random Processes / Ju. A. Rooznov. – Moscow: Fizmatgiz, 1963.
9. Cheremskaya N. V. Linear Transformations of Nonstationary Stochastic Sequences. // Radiotekhneka. – 2004. – vol 136. – P. 43-49 (Russian).
10. Loève M. Probability Theory / M. Loève. – Prinston, 1955. – 719 pp.

Надійшла у першій редакції 25.10.2014, в останній - 30.10.2014.

CONTENTS

Наукове видання

Вісник Харківського національного університету
№ 1131

Серія «Математичне моделювання. Інформаційні технології.
Автоматизовані системи управління»

Випуск 25

Збірник наукових праць

Українською, російською та англійською мовами

Комп'ютерне верстання О. О. Афанасьєва

Підписано до друку 03.11.2014 р.
Формат 70×108/16. Папір офсетний. Друк ризограф.
Ум. друк. арк. – 15,5
Обл.– вид. арк. – 18,0
Тираж 100 пр.
Ціна договірна

61022, м. Харків, майдан Свободи, 4
Харківський національний університет імені В.Н.Каразіна.
Видавництво

Надруковано ФОП «Петрова І.В.»
61144 Харків-144, вул. Гв. Широнінців, 79в, к.137, тел. 362-01-52
Свідоцтво про державну реєстрацію ВОО №948011 від 03.01.03