Інститут кібернетики імені В. М. Глушкова
Національної академії наук України
Кам’янець-Подільський національний університет
імені Івана Огієнка

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

Серія: Фізико-математичні науки

Збірник наукових праць

Випуск 19

Кам’янець-Подільський національний університет
імені Івана Огієнка
2019

УДК 519.6:519.7
ББК 22
М34

Свідоцтво про державну реєстрацію друкованого засобу масової інформації:
Серія КВ № 14521-3492Р від 25.06.2008 р.

Збірник наукових праць включено до Переліку наукових фахових видань ДАК Міністерства освіти і науки України з фізико-математичних наук (наказ №1021 від 07 жовтня 2015 р.)

Друкується згідно з рішенням вченої ради Кам'янець-Подільського національного університету імені Івана Огієнка, протокол № 6 від 23 травня 2019 року.

М34 **Математичне та комп'ютерне моделювання. Серія: Фізико-математичні науки** : зб. наук. праць / Інститут кібернетики імені В. М. Глушкова Національної академії наук України, Кам'янець-Подільський національний університет імені Івана Огієнка ; [редкол.: О. М. Хіміч (відп. ред.) та ін.]. — Кам'янець-Подільський : Кам'янець-Подільський національний університет імені Івана Огієнка, 2019. — Вип. 19. — 208 с.

У збірнику друкуються результати досліджень вітчизняних та закордонних науковців, що стосуються проблем застосування математичних моделей в різних галузях людської діяльності.

Для наукових та інженерно-технічних працівників, аспірантів, студентів.

УДК 519.6:519.7
ББК 22

ISSN 2308-5878
DOI: 10.32626/2308-5878.2019-19

V. M. Glushkov Institute of Cybernetics
of National Academy of Sciences of Ukraine
Kamianets-Podilskyi National Ivan Ohiienko University

MATHEMATICAL AND COMPUTER MODELLING

Series: Physical and mathematical sciences

Scientific journal

ISSUE 19

Kamianets-Podilskyi National Ivan Ohiienko University
2019

Mathematical and computer modelling. Series: Physical and mathematical sciences : scientific journal / V. M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kamianets-Podilskyi National Ivan Ohiienko University ; [Editorial Board: O. Himich (Executive Editor) and others]. — Kamianets-Podilskyi : Kamianets-Podilskyi National Ivan Ohiienko University, 2019. — ISSUE 19. — 208 p.

There are printed results of investigation of national and foreign scientists that concern to problems of practice mathematical models in different spheres of human activity.

For scientific and technical staff, postgraduate students.

ISSN 2308-5878
DOI: 10.32626/2308-5878.2019-19

УДК 519.85
DOI: 10.32626/2308-5878.2019-19.5-11

Т. М. Барболіна, канд. фіз.-мат. наук

Полтавський національний педагогічний університет
імені В. Г. Короленка, м. Полтава

ВЛАСТИВОСТІ ЕВКЛІДОВИХ ЗАДАЧ ЛЕКСИКОГРАФІЧНОЇ КОМБІНАТОРНОЇ ОПТИМІЗАЦІЇ НА РОЗМІЩЕННЯХ

Розглядаються евклідові задачі лексикографічної комбінаторної оптимізації, які передбачають знаходження лексикографічно мінімальної (для задач мінімізації) чи лексикографічно максимальної (для задач максимізації) точки серед тих, які надають екстремум цільовій функції на заданій евклідовій комбінаторній множині. Обґрунтовано властивості лінійних та дробово-лінійних задач лексикографічної комбінаторної оптимізації на загальній множині розміщень без додаткових обмежень. Отримані в роботі результати спираються на відомі раніше критерії екстремалей лінійної та дробово-лінійної функцій на розміщеннях: будь-яка екстремаль є елементом певної множини полірозміщень (для лінійних задач вигляд множини екстремалей встановлений явно, для дробово-лінійних задач множина полірозміщень формується на основі деякої відомої екстремалі). У роботі встановлено вигляд точок, які є лексикографічною мінімаллю та лексикографічною максималлю лінійної функції на загальній множині розміщень. Зокрема, якщо елементи мультимножини упорядковані за неспаданням, а коефіцієнти цільової функції — за незростанням, причому s — найменший індекс такий, що відповідний коефіцієнт цільової функції є від'ємним, то лексикографічна мінімаль формується як упорядковані за неспаданням $s-1$ перших та $k-s+1$ (k — вимірність простору) останніх елементів мультимножини. Для задач з дробово-лінійною цільовою функцією встановлений спосіб формування розв'язку задачі лексикографічної комбінаторної оптимізації на розміщеннях, якщо відома будь-яка з мінімалей (для задач мінімізації) чи максималей (для задач максимізації) цільової функції на заданій множині розміщень. Упорядкування компонент екстремалі у цьому випадку здійснюється з урахуванням упорядкування за незростанням коефіцієнтів лінійної функції спеціального вигляду.

Ключові слова: *комбінаторна оптимізація, евклідові задачі лексикографічної комбінаторної оптимізації, оптимізаційні задачі на розміщеннях.*

Вступ. Оптимізаційні задачі з обмеженнями комбінаторного характеру привертають увагу багатьох дослідників [1–6]. Такий інтерес зумовив виокремлення задач на так званих евклідових комбінаторних мно-

жинах. Важливий клас евклідових задач комбінаторної оптимізації становлять задачі на розміщеннях, дослідженню яких присвячені, серед іншого, роботи [3–6]. Зокрема, встановлено достатню [3] та необхідну [5] умови екстремалі лінійної цільової функції на загальній множині розміщень, спосіб формування множини мінімалей (максималей) дробово-лінійної функції на множині розміщень, якщо відома одна з них [5]. З іншого боку, у [6] запропоновано формулювання задач комбінаторної оптимізації, яке передбачає знаходження не довільної екстремалі, а лексикографічно максимальної (мінімальної) точки серед тих, які надають максимум (мінімум) цільовій функції. Такі задачі називаються евклідовими задачами лексикографічної комбінаторної оптимізації.

Мета роботи — обґрунтування властивостей деяких класів задач лексикографічної комбінаторної оптимізації на загальній множині розміщень.

Розглянемо спочатку необхідні позначення та факти. Позначимо $J_n = \{1, 2, \dots, n\}$, $J_n^m = \{m, m+1, \dots, n\}$. Термінологію стосовно евклідових задач комбінаторної оптимізації використовуватимемо переважно з [3]. Зокрема, під мультимножиною розуміємо сукупність елементів, серед яких можуть бути однакові. Загальною множиною розміщень з елементів мультимножини $G = \{g_1, g_2, ..., g_\eta\}$ називають множину всіх упорядкованих k-вибірок з мультимножини G вигляду $(g_{i_1}, g_{i_2}, ..., g_{i_k})$, де $g_{i_j} \in G$, $i_j \neq i_t \ \forall i_j, i_t \in J_\eta$, $\forall j, t \in J_k$.

Розглянемо упорядковане розбиття множини J_η на σ множин $N_1, N_2, \dots, N_\sigma$, яке задовольняє умови $N_i \cap N_j = \varnothing$, $N_i \neq \varnothing \ \ \forall i \in J_\sigma$, а також упорядковане розбиття числа k на σ доданків $k_1, k_2, ..., k_\sigma$, що задовольняє умови $1 \le k_i \le |N_i| \ \ \forall i \in J_\sigma$. Нехай H — множина всіх k-вибірок з множини J_η вигляду

$$\pi = (\pi(1), ..., \pi(k)) = (\pi_{11}, ..., \pi_{1k_1}, ..., \pi_{\sigma 1}, ..., \pi_{\sigma k_\sigma}) = (\pi^1, ..., \pi^\sigma), \quad (1)$$

де $\pi^i = (\pi_{i1}, ..., \pi_{ik_i})$ — довільна k_i-вибірка з множини $N_i \ \ \forall i \in J_\sigma$. Множина $E_{\eta n}^{k\sigma}(G, H) = \{(g_{\pi(1)}, ..., g_{\pi(k)}) \mid \forall \pi \in H\}$ називається загальною множиною полірозміщень.

Розглянемо розв'язування лінійної безумовної задачі лексикографічної комбінаторної мінімізації на розміщеннях, тобто задачу знаходження на множині $E_\eta^k(G)$ мінімуму та лексикографічної мінімалі (лексикографічно мінімальної точки серед тих, що надають цільовій функції мінімуму) функції

$$C(x) = \sum_{j=1}^{k} c_j x_j \,. \tag{2}$$

Вважатимемо, що елементи мультимножини G задовольняють умову

$$g_1 \le g_2 \le \ldots \le g_\eta \,, \tag{3}$$

а коефіцієнти цільової функції — умову

$$c_{t_1} = \ldots = c_{t_2-1} > c_{t_2} = \ldots = c_{t_3-1} > \ldots > c_{t_\sigma} = \ldots = c_k \,. \tag{4}$$

Теорема 1 (критерій лексикографічної мінімалі лінійної функції на загальній множині розміщень). Нехай s — найменший індекс такий, що $c_{t_s} < 0$. Лексикографічною мінімаллю функції (2) на загальній множині розміщень $E_\eta^k(G)$ є точка x^*, для якої

$$x_j^* = g_j \quad \forall j \in J_{t_s-1}\,,\ x_j^* = g_{\eta-k+j} \quad \forall j \in J_k^{t_s} \,. \tag{5}$$

Доведення. Як випливає з критерію мінімалі лінійної цільової функції на розміщеннях (теорема 3 [5]), точка x^* є мінімаллю функції (2) на множині $E_\eta^k(G)$ тоді і лише тоді, коли вона задовольняє умові $x^* \in E_{\eta\eta}^{ks}(G,H)$, де H — множина k-вибірок з множини J_η вигляду (1), при формуванні яких $k = \overline{t}_1 + \overline{t}_2 + \ldots + \overline{t}_\sigma$ ($\overline{t}_w = t_{w+1} - t_w$), а множини N_w визначаються таким чином:

$$N_w = \begin{cases} \{t_w, \ldots, t_{w+1} - 1\}, \text{ якщо } c_{t_w} > 0, \\ \{\eta - k + t_w, \ldots, \eta - k + t_{w+1} - 1\}, \text{ якщо } c_{t_w} < 0, \\ J_\eta \setminus \bigcup\limits_{v \in W} N_v, \text{ якщо } c_{t_w} = 0. \end{cases} \tag{6}$$

Очевидно, що лексикографічно мінімальною упорядкованою $\overline{t}_w$-вибіркою з множини N_w при $w \in J_\sigma^s$ є вибірка $(\eta - k + t_w, \eta - k + t_w + 1, \ldots, \eta - k + t_{w+1} - 1)$, а при $w \in J_r$ (r — найбільший індекс такий, що $c_{t_r} > 0$) — вибірка $(t_w, t_w + 1, \ldots, t_{w+1} - 1)$. Також, якщо $r \ne s-1$, тобто $r+1 = s-1$, то лексикографічно мінімальною упорядкованою $\overline{t}_{r+1}$-вибіркою з множини N_{r+1} є $(t_{r+1}, t_{r+1} + 1, \ldots, t_s - 1)$. Таким чином, лексикографічно мінімальною k-вибіркою вигляду (1) є $(1, 2, \ldots, s-1, \eta - k + s, \eta - k + s + 1, \ldots \eta)$. Оскільки елементи мультимножини G задовольняють умові (3), то лексикографічна мінімаль функції (2) на множині $E_\eta^k(G)$ задовольняє умові (5). **Теорема доведена.**

Аналогічно доводиться теорема для випадку максимізації.

Теорема 2 (критерій лексикографічної максималі лінійної функції на загальній множині розміщень). Нехай s — найбільший індекс такий, що $c_{t_s} < 0$. Лексикографічною максималлю функції (2) на загальній множині розміщень $E_\eta^k(G)$ є точка x^*, для якої

$$x_j^* = g_{\eta - j + 1} \quad \forall j \in J_{t_s - 1}, \; x_j^* = g_{k - j + 1} \quad \forall j \in J_k^{t_s}. \tag{7}$$

Розглянемо тепер безумовну задачу лексикографічної комбінаторної оптимізації на розміщеннях з дробово-лінійною цільовою функцією

$$\Phi(x) = \frac{\sum_{j=1}^{k} c_j x_j + c_0}{\sum_{j=1}^{k} d_j x_j + d_0}. \tag{8}$$

У роботі [5] встановлено спосіб формування множини мінімалей (максималей) функції (8) на загальній множині розміщень $E_\eta^k(G)$, якщо відома одна з них (для пошуку такої мінімалі може бути використаний поліноміальний метод, розроблений у [6]). Розглянемо знаходження розв'язку задачі лексикографічної комбінаторної оптимізації дробово-лінійної функції на розміщеннях. Нехай x' — мінімаль функції (8) на множині $E_\eta^k(G)$, $\Phi^* = \Phi(x')$. Перенумеруємо змінні таким чином, щоб виконувалася умова

$$c_{q_1} - \Phi^* d_{q_1} \geq c_{q_2} - \Phi^* d_{q_2} \geq \ldots \geq c_{q_k} - \Phi^* d_{q_k}, \tag{9}$$

причому якщо $c_{q_i} - \Phi^* d_{q_i} = c_{q_j} - \Phi^* d_{q_j}$ і $i < j$, то $q_i < q_j$.

Позначимо $c'_j = c_{q_j} - \Phi^* d_{q_j}$, $y_j = x_{q_j}$ для всіх $j \in J_k$. Нехай індекси p_i $i \in J_k$ такі, що виконується умова

$$c'_{p_1} = \ldots = c'_{p_2 - 1} > c'_{p_2} = \ldots = c'_{p_3 - 1} > \ldots > c'_{p_u} = \ldots = c'_k. \tag{10}$$

Позначимо $I = \left\{ i \in J_u \,\middle|\, c_{p_i} \neq 0 \right\}$, $k_i = p_{i+1} - p_i \quad \forall i \in J_u$ і розглянемо розбиття числа k на u доданків $k = \bar{p}_1 + \bar{p}_2 + \ldots + \bar{p}_u$ і розбиття множини J_η на u підмножин N'_i за правилом:

$$N'_i = \begin{cases} \left\{ l_j \,\middle|\, j \in \bar{N}_i \right\}, \text{ де } \bar{N}_i = \left\{ q_{p_i}, \ldots, q_{p_{i+1} - 1} \right\}, \text{ якщо } i \in I, \\ J_\eta \setminus \bigcup\limits_{j \in I} N'_j, \text{ якщо } c_{p_i} = 0. \end{cases} \tag{11}$$

Розглянемо множину полірозміщень $E_{\eta n}^{ku}(G,H)$, де H — множина вибірок вигляду (1), $\pi^i=(\pi_{i1},\pi_{i2},...,\pi_{ik_i})$ — довільна k_i -вибірка із множини N_i' $\forall i\in J_u$.

Теорема 3 (критерій лексикографічної мінімалі дробово-лінійної функції на загальній множині розміщень). Нехай індекси r_{ij} такі, що для всіх $i\in J_u$, $j\in J_{|N_i'|-1}$ виконуються нерівності $l_{r_{ij}}<l_{r_{u,j+1}}$. Точка x^* є лексикографічною мінімаллю функції (8) на загальній множині розміщень $E_\eta^k(G)$ тоді і лише тоді, коли $x^*_{q_j}=y_j^*$ $\forall j\in J_k$, де $\left(y_1^*,y_2^*,...,y_k^*\right)$ — такий елемент множини полірозміщень $E_{\eta n}^{ku}(G,H)$, що для всіх $i\in J_u$ вибірки $\pi^i=(\pi_{i1},\pi_{i2},...,\pi_{ik_i})$ формуються за правилом $\pi_{ij=l_{r_{ij}}}$ $\forall j\in J_{k_i}$.

Доведення. Припустимо існує така мініималь $\tilde{x}$ функції (8) на множині $E_\eta^k(G)$, що $\tilde{x}<_l x^*$ (тут і далі $<_l$ — символ лексикографічного порядку). Позначимо t — найменший індекс такий, що $\tilde{x}_t\neq x_t^*$ (тобто $\tilde{x}_t<x_t^*$).

Згідно з критерієм екстремалі дробово-лінійної функції на загальній множині розміщень (теорема 5 [5]) існує таке $\tilde{y}\in E_{\eta n}^{ku}(G,H)$, що $\tilde{y}_j=\tilde{x}_{q_j}$ $\forall j\in J_k$. Нехай v — таке, що $q_v=t$, f визначається з нерівності $p_f\le v<p_{f+1}$. З умови теореми випливає, що π^f є лексикографічно мінімальною k_f -вибіркою з множини N_f'. Враховуючи, що елементи мультимножини G задовольняють умові (3), отримуємо, що для будь-якого

$$y\in E_{\eta n}^{ku}(G,H)\ \left(y_{p_f}^*,y_{p_f+1}^*,\cdots,y_{p_{f+1}-1}^*\right)\le_l\left(y_{p_f},y_{p_f+1},\cdots,y_{p_{f+1}-1}\right).$$

Зокрема,

$$\left(y_{p_f}^*,y_{p_f+1}^*,\cdots,y_{p_{f+1}-1}^*\right)\le_l\left(\tilde{y}_{p_f},\tilde{y}_{p_f+1},\cdots,\tilde{y}_{p_{f+1}-1}\right). \qquad (12)$$

Оскільки $y_v^*=x_{q_v}^*=x_t^*>\tilde{x}_t=\tilde{y}_v$, то рівність в умові (12) місця не має і знайдеться такий індекс $p_f\le\tau<v<p_{f+1}$, що $y_\tau^*<\tilde{y}_\tau$. А тоді також $x_{q_\tau}^*<\tilde{x}_{q_\tau}$, причому зі способу упорядкування величин $c_{q_i}-\Phi^* d_{q_i}$ ви-

пливає, що $q_\tau < q_v = t$ (оскільки $c_{q_\tau} - \Phi^* d_{q_\tau} = c_{q_v} - \Phi^* d_{q_v}$ і $\tau < v$). Отримали суперечність з тим фактом, що t — найменший індекс такий, що $\tilde{x}_t \neq x_t^*$. **Теорема доведена.**

Аналогічна теорема має місце і для випадку максимізації.

Висновки. У роботі обґрунтовано критерії лексикографічної екстремалі лінійної та дробово-лінійної функцій на загальній множині розміщень. Отримані властивості можуть використовуватися при дослідженні інших класів оптимізаційних задач, у тому числі з різними видами невизначеності.

Список використаних джерел:

1. Сергиенко И. В., Шило В. П. Задачи дискретной оптимизации: проблемы, методы решения, исследования. Киев : Наук. думка, 2003. 261 с.
2. Згуровский М. З., Павлов А. А. Принятие решений в сетевых системах с ограниченными ресурсами. Киев : Наук. думка. 2010. 573 с.
3. Стоян Ю. Г., Ємець О. О. Теорія і методи евклідової комбінаторної оптимізації. Київ : Інститут системних досліджень освіти, 1993. 188 с.
4. Емец О. А., Барболина Т. Н. Комбинаторная оптимизация на размещениях. Київ : Наук. думка, 2008. 159 с.
5. Емец О. А., Барболина Т. Н. Свойства комбинаторных оптимизационных безусловных задач на размещениях с линейной и дробно-линейной целевыми функциями. *Проблемы управления и информатики.* 2017. № 1. С. 66–76.
6. Емец О. А., Барболина Т. Н. Полиномиальный метод решения безусловной дробно-линейной задачи комбинаторной оптимизации на размещениях. *Проблемы управления и информатики.* 2017. № 2. С. 27–36.

PROPERTIES OF EUCLIDEAN PROBLEMS OF LEXICOGRAPHIC COMBINATORIAL OPTIMIZATION ON ARRANGEMENTS

In the paper Euclidean problems of lexicographic combinatorial; optimization are discussed. These problems are to find lexicographically minimal (for minimization problems) or lexicographically maximal (for maximization problems) points among those that give the extremum of the objective function on a given Euclidean combinatorial set. The properties of linear and linear-fractional problems of lexicographic combinatorial optimization on a general set of arrangements are substantiated. The results obtained in the work are based on the previously known criteria of extremals of linear and linear-fractional functions on arrangements: any extremal is an element of certain set of polyarrangements (for linear problems the form of the extremal set is established explicitly, for linear-fractional problems the polyarrangement set is formed on the basis of some known extremal). In the paper we substantiate the form of points that are an lexicographic minimal and lexicographic maximal of linear function on the general set of arrangements. In particular if elements of multiset are in nondecreasing order, coefficients of objective function are in nonincreasing order and *s* is the least index such that corresponding coefficient of objective function is

negative, tehn lexicographic minimal if formed as $s-1$ first and $k-s+1$ (k is the dimension of space) last elements of multiset which are in nondecreasing order. For problems with linear-fractional function we obtain the method of forming solution of lexicographic combinatorial problem on arrangements, if any minimal (for minimization problems) or any maximal (for maximization problems) of objective function on given set of arrangements is known. In this case ordering of components of the extremal is carried out taking into account ordering for nonincreasing of coefficients of special linear function.

Key words: *combinatorial optimization, Euclidean problems of lexicographic combinatorial optimization, optimization problems on arrangements.*

Одержано 31.01.2019

УДК 512.61:519.61
DOI: 10.32626/2308-5878.2019-19.11-17

Є. Ф. Галба, д-р фіз.-мат. наук,
Н. А. Варенюк, канд. фіз.-мат. наук,
Н. І. Тукалевська, канд. фіз.-мат. наук

Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ЗВАЖЕНЕ СИНГУЛЯРНЕ РОЗВИНЕННЯ МАТРИЦЬ ТА МЕТОДИ РОЗВ'ЯЗУВАННЯ ЗАДАЧ ЗВАЖЕНОЇ ПСЕВДОІНВЕРСІЇ З ВИРОДЖЕНИМИ ВАГАМИ

На основі зваженого сингулярного розвинення матриць з виродженими вагами отримано зображення зважених псевдообернених матриць з виродженими вагами та їх розвинення в матричні степеневі ряди і матричні степеневі добутки. Отримано граничні зображення зважених псевдообернених матриць, на основі яких побудовано та досліджено регуляризовані методи обчислення зважених нормальних псевдорозв'язків з виродженими вагами.

Ключові слова: *зважене сингулярне розвинення матриць з виродженими вагами, зважені псевдообернені матриці.*

Вступ. Вперше сингулярне розвинення матриць отримано у монографії [1]. У роботах [2, 3] отримано зважене сингулярне розвинення матриць з додатно-означеними вагами. У ряді робіт (див. [4, 5] та наявну там бібліографію) отримано зважене сингулярне розвинення матриць з виродженими вагами на основі ортогональних, зважених ортогональних та зважених псевдоортогональних матриць. У цих роботах визначені достатні умови існування запропонованих варіантів зваженого сингулярного розвинення матриць з виродженими вагами, а у роботі [4] визначено необхідні та достатні умови, при яких існує побудоване зважене сингулярне розвинення матриць з виродженими вагами на основі ортогональних матриць.

У роботі [6] вперше дано визначення одного із варіантів зважених псевдообернених матриць з виродженими вагами і визначені необхідні та достатні умови існування та єдиності розглянутого варіанта зважених псевдообернених матриць. У ряді робіт (див. [5, 7] та наявну там бібліографію) досліджені інші варіанти зважених псевдообернених матриць з виродженими вагами. Визначені необхідні та достатні умови існування та єдиності розглянутих зважених псевдообернених матриць та встановлено їх зв'язок із зваженими нормальними псевдорозв'язками.

У представленій статті використовується зважене сингулярне розвинення матриць з виродженими вагами, запропоноване та досліджене у роботі [4], для встановлення властивостей, розглянутих в роботі [7] зважених псевдообернених матриць з виродженими вагами. Отримано розвинення зважених псевдообернених матриць з виродженими вагами в матричні степеневі ряди і матричні степеневі добутки. Отримано многочленні граничні зображення цих матриць, на основі яких побудовано та досліджено регуляризовані методи обчислення зважених нормальних псевдорозв'язків з виродженими вагами.

Позначення, визначення. Зазначимо, що в подальшому скрізь припускається дійсність використовуваних скалярів, векторів, матриць та просторів. Позначимо R^n — n-вимірний векторний простір над полем дійсних чисел, де вектори є матриці розміру $n \times 1$. Нехай H — додатно-означена або ж додатно-напіввизначена матриця. $R^n(H)$ позначатимемо евклідів простір у випадку додатно-означеної метрики або ж псевдоевклідів у випадку невід'ємної метрики, що введена скалярним добутком $(u, v)_H = (Hu, v)_E$, де $(u, v)_E = u^T v$. Норму (напівнорму) в $R^n(H)$ введемо формулою $\|u\|_H = (u, u)_H^{1/2}$. У випадку додатно-напіввизначеної матриці H через $\overline{R}^n(H) \subset R^n(H)$ і $\overline{R}^n(H_{EE}^+) \subset R^n(H_{EE}^+)$ позначатимемо підпростір векторів u, що задовольняють умову $HH_{EE}^+ u = H^{1/2} H_{EE}^{+1/2} u = u$, де позначено $H_{EE}^{+1/2} = (H^{1/2})_{EE}^+$, H_{EE}^+ — псевдообернена матриця Мура–Пенроуза до матриці H, а E — одинична матриця.

Зазначимо, що напівнорми $\|\cdot\|_H$, $\|\cdot\|_{H_{EE}^+}$ для векторів у $R^n(H)$, $R^n(H_{EE}^+)$ будуть нормами в $\overline{R}^n(H)$, $\overline{R}^n(H_{EE}^+)$ [7].

Визначимо норму прямокутної матриці. Нехай $A \in R^{m \times n}$, а $H \in R^{m \times m}$ і $V \in R^{n \times n}$ — додатно-напіввизначені матриці, x — довільний вектор із R^n. Припускаємо, що виконуються умови $rk(HA) =$

$= rk(A)$, $rk(AV) = rk(A)$. Для множини матриць A, що задовольняють ці умови, норму введемо співвідношенням

$$\|A\|_{HV} = \sup_{x \neq 0} \frac{\left\| H^{1/2} AVx \right\|_{E_m}}{\|x\|_{E_n}}. \qquad (1)$$

Нехай $A \in R^{m \times n}$, $X \in R^{n \times m}$, а $B \in R^{m \times m}$ і $C \in R^{n \times n}$ — симетричні додатно-напіввизначені матриці. Тоді зважена псевдообернена до A матриця в роботі [7] визначається як матриця $X = A^{+}_{BC}$, що задовольняє умови

$$AXA = A,\ XAX = X,\ (BAX)^T = BAX,\ (CXA)^T = CXA. \qquad (2)$$

У роботі [7] визначені необхідні та достатні умови існування єдиного розв'язку системи матричних рівнянь (2) з виродженими вагами.

Теорема 1. Для того, щоб система матричних рівнянь (2) з виродженими вагами мала єдиний розв'язок $X = A^{+}_{BC}$, необхідно та достатньо виконання умов

$$rk(BA) = rk(A), \quad AC^{+}_{EE}C = A. \qquad (3)$$

Нехай

$$Ax = f,\ x \in R^n,\ f \in R^m, \qquad (4)$$

— система лінійних алгебраїчних рівнянь (СЛАР) з $A \in R^{m \times n}$.

У роботі [7] встановлено зв'язок між зваженими псевдооберненими матрицями з виродженими вагами, визначеними умовами (2), (3) і зваженими нормальними псевдорозв'язками.

Теорема 2. Вектор $x^{+} = A^{+}_{BC} f$, де матриця A^{+}_{BC} визначена умовами (2), (3), є в $\overline{R}^n(C)$ зваженим нормальним псевдорозв'язком СЛАР (4) з додатно-напіввизначеними вагами B і C, а саме, єдиним розв'язком задачі: знайти

$$\min_{x \in \overline{R}^n(C) \cap \Omega} \|x\|_C, \quad \Omega = Arg \min_{x \in R^n} \|Ax - f\|_B. \qquad (5)$$

Зважене сингулярне розвинення та зважене псевдообернення матриць. В роботі [4] отримано зважений сингулярний розклад матриць с виродженими вагами на основі ортогональних матриць. Визначено необхідні і достатні умови існування цього розвинення.

Теорема 3. Нехай $A \in R^{m \times n}$ і виконуються умови

$$B^{+}_{EE} BA = A,\ AC^{+}_{EE}C = A, \qquad (6)$$

тоді:

1) для матриці A існують ортогональні матриці $U \in R^{m \times m}$ і $V \in R^{n \times n}$ такі, що

$$U^T B^{1/2} A C_{EE}^{+1/2} V = \Sigma = \begin{cases} \left\| diag(\sigma_1, \sigma_2, ..., \sigma_r, 0, ..., 0) \mathrm{O}_m^{n-m} \right\|, & m \le n \\ \left\| \begin{matrix} diag(\sigma_1, \sigma_2, ..., \sigma_r, 0, ..., 0) \\ \mathrm{O}_{m-n}^{n} \end{matrix} \right\|, & m \ge n \end{cases}$$

і

$$A = B_{EE}^{+1/2} U \Sigma V^T C^{1/2}, \tag{7}$$

де r — ранг матриці A, стовпчики матриці U — ортонормовані в $R^m(E)$ власні вектори матриці $B^{1/2} A C_{EE}^{+} A^T B^{1/2}$, стовпчики матриці V — ортонормовані в $R^n(E)$ власні вектори матриці $C_{EE}^{+1/2} A^T B A C_{EE}^{+1/2}$, σ_i, $i = 1,\ldots, r$ — квадратні корені із ненульових власних значень матриці $B^{1/2} A C_{EE}^{+} A^T B^{1/2}$, $\mathrm{O}_k^l \in R^{k \times l}$ — нульова матриця;

2) умови (6) є необхідними і достатніми для існування зваженого сингулярного розвинення матриці A вигляду (7).

Отримано розвинення зважених псевдообернених матриць на основі зваженого сингулярного розвинення матриць.

Теорема 4. Зважена псевдообернена матриця для матриці A, визначена умовами (2), при виконанні умов (6) має розклад

$$A_{BC}^{+} = C_{EE}^{+1/2} V \Sigma_{EE}^{+} U^T B^{1/2}, \tag{8}$$

де матриці V, U, B, C визначені в теоремі 3, а матриця Σ_{EE}^{+} — псевдообернена матриця Мура–Пенроуза до матриці Σ, визначеної в теоремі 3.

Розвинення в матричні степеневі ряди і добутки зважених псевдообернених матриць з виродженими вагами. На основі зваженого сингулярного розвинення зважених псевдообернених матриць з виродженими вагами (див. теореми 3, 4) отримано і досліджено [4] розвинення в матричні степеневі ряди і добутки з від'ємними показниками степенів зважених псевдообернених матриць, визначених умовами (2), (6).

Теорема 5. Для довільної матриці $A \ne 0 \in R^{m \times n}$, симетричних додатно-напіввизначених матриць $B \in R^{m \times m}$ і $C \in R^{n \times n}$, що задовольняють умови (2), (6), і для дійсного числа $0 < \delta < \infty$ має місце наступний розклад зважених псевдообернених матриць, визначених умовами (2), (6), в матричні степеневі ряди

$$A_{BC}^{+} = \sum_{k=1}^{\infty} \delta^{k-1} (C_{EE}^{+} A^T B A + \delta E)^{-k} C_{EE}^{+} A^T B, \tag{9}$$

причому,

$$\left\| A_{BC}^{+} - A_{\delta, p}^{+} \right\|_{C B_{EE}^{+1/2}} \le \sigma_*^{-1} \delta^p (\delta + \sigma_*^2)^{-p}, \tag{10}$$

де $A^{+}_{\delta,p} = \sum_{k=1}^{p} \delta^{k-1} (C^{+}_{EE} A^T BA + \delta E)^{-k} C^{+}_{EE} A^T B, \quad p = 1, 2, \ldots,$ σ_* — мінімальний ненульовий діагональний елемент матриці Σ, визначеної в теоремі 3.

При виконанні припущень теореми 5 маємо наступний розклад зваженої псевдообернені матриці з виродженими вагами в матричний степеневий добуток

$$A^{+}_{BC} = \prod_{k=0}^{\infty} \{E + \delta^{2^k} (C^{+}_{EE} A^T BA + \delta E)^{-(2^k)}\} (C^{+}_{EE} A^T BA + \delta E)^{-1} C^{+}_{EE} A^T B. \quad (11)$$

Позначимо

$$A^{+}_{\delta,n} = \prod_{k=0}^{n-1} \{E + \delta^{2^k} (C^{+}_{EE} A^T BA + \delta E)^{-(2^k)}\} (C^{+}_{EE} A^T BA + \delta E)^{-1} C^{+}_{EE} A^T B, \; n = 1, 2 \ldots$$

Тоді в силу оцінки (10) отримаємо

$$\left\| A^{+}_{BC} - A^{+}_{\delta,n} \right\|_{CB^{+1/2}_{EE}} \le \sigma_*^{-1} \delta^{2^n} (\delta + \sigma_*^2)^{-(2^n)}. \quad (12)$$

Граничні зображення зважених псевдообернених матриць і регуляризація задач. Із оцінки (10) випливає, що для довільного p = 1, 2,… маємо наступне граничне представлення зваженої псевдооберненої матриці:

$$A^{+}_{BC} = \lim_{\delta \to +0} \sum_{k=1}^{p} \delta^{k-1} (C^{+}_{EE} A^T BA + \delta E)^{-k} C^{+}_{EE} A^T B, \quad (13)$$

а із оцінки (12) для довільного n = 1, 2, … маємо

$$A^{+}_{BC} = \lim_{\delta \to +0} \prod_{k=0}^{n-1} \{E + \delta^{2^k} (C^{+}_{EE} A^T BA + \delta E)^{-(2^k)}\} (C^{+}_{EE} A^T BA + \delta E)^{-1} C^{+}_{EE} A^T B. \quad (14)$$

Розвинення зважених псевдообернених матриць у матричні степеневі ряди і добутки можна використовувати для побудови регуляризованих ітераційних методів їх обчислення [4].

На основі граничних представлень зважених псевдообернених матриць можна також запропонувати регуляризовані задачі для обчислення зважених нормальних псевдорозв'язків. Згідно з граничним представленням (13) наближення до розв'язку задачі (5) при достаньо малому δ можна отримати розв'язуючи СЛАР

$$(C^{+}_{EE} A^T BA + \delta E)^{p-m} x = \sum_{k=1}^{p} \delta^{k-1} (C^{+}_{EE} A^T BA + \delta E)^{p-m-k} C^{+}_{EE} A^T Bf, \quad (15)$$

де $m = 0, 1, \ldots, p-1$.

Оцінку похибки наближення розв'язку задачі (5) розв'язком однієї із систем (15) дає наступна теорема.

Теорема 6. Нехай x^{+} — розв'язок задачі (5), а $x_{\delta,p}$ — розв'язок однієї із систем (15), тоді має місце оцінка

$$\left\| x^{+} - x_{\delta,p} \right\|_{C} \le \sigma_{*}^{-1} \delta^{p} (\delta + \sigma_{*}^{2})^{-p} \left\| f \right\|_{B}, \qquad (16)$$

де σ_* — мінімальний ненульовий діагональний елемент матриці Σ, визначеної в теоремі 3.

На основі формули (14) наближення до розв'язку задачі (5) при досить малому δ можна отримати розв'язуючи СЛАР

$$(C_{EE}^{+} A^{T} BA + \delta E)^{n-m} x =$$
$$= \prod_{k=0}^{n-1} \{ (C_{EE}^{+} A^{T} BA + \delta E)^{n-m-1} + \delta^{2^{k}} (C_{EE}^{+} A^{T} BA + \delta E)^{n-m-(2^{k})-1} \} C_{EE}^{+} A^{T} Bf, \qquad (17)$$

де $m = 0,1,...,n-1$.

Оцінку похибки наближення розв'язку задачі (5) розв'язком систем (17) дає наступна теорема.

Теорема 7. Нехай x^{+} — розв'язок задачі (5), а $x_{\delta,n}$ — розв'язок однієї із систем (17), тоді має місце оцінка

$$\left\| x^{+} - x_{\delta,n} \right\|_{C} \le \sigma_{*}^{-1} \delta^{2^{n}} (\delta + \sigma_{*}^{2})^{-(2^{n})} \left\| f \right\|_{B}.$$

Висновки. У роботі встановлено зв'язок зваженого сингулярного розвинення матриць з виродженими вагами із зваженими псевдообернеными матрицями з виродженими вагами. Досліджено фундаментальні властивості зважених псевдообернених матриць з виродженими вагами. Побудовано регуляризовані методи обчислення зважених нормальних псевдорозв'язків з виродженими вагами. Перспективними з даного напряму є дослідження зважених псевдообернених матриць з індефінітними ваговими матрицями.

Список використаних джерел:

1. Forsythe G., Moler C. Computer solution of linear algebraic systems. Englewood Cliffs, N.J. : Prentice-Hall, 1967. 148 p.
2. Van Loan C.F. Generalizing the singular value decomposition. SIAM *J. Numer. Anal.* 1976. Vol. 13, № 1. P. 76–83.
3. Галба Е. Ф. Взвешенное сингулярное разложение и взвешенное псевдообращение матриц. *Укр. мат. журн.* 1996. 48, № 10. С. 1426–1430.
4. Сергиенко И. В., Галба Е. Ф., Дейнека В. С. Необходимые и достаточные условия существования взвешенного сингулярного разложения матриц с вырожденными весами. *Укр. мат. журн*. 2015. Вип. 67. № 3. С. 406–426.
5. Сергиенко И. В., Галба Е. Ф. Взвешенная псевдоинверсия с вырожденными весами. *Кибернетика и системный анализ*. 2016. № 5. С. 56–80.
6. Ward J. F., Boullion T. L., Lewis T. O. Weighted pseudoinverses with singular weights. *SIAM J. Appl. Math.* 1971. Vol. 21. № 3. P. 480–482.
7. Галба Е.Ф., Дейнека В.С., Сергиенко И.В. Взвешенные псевдообратные матрицы и взвешенные нормальные псевдорешения с вырожденными весами. *Журн. вычисл. математики и мат. физики.* 2009. Вип. 49. № 8. С. 1347–1363.

WEIGHTED SINGULAR-VALUED DECOMPOSITION OF MATRICES AND METHODS OF SOLVING PROBLEMS WEIGHTED PSEUDOINVERSE WITH SINGULAR WEIGHTS

Weighted pseudoinverse matrices with singular weights and their expansions into matrix power series and matrix power products are obtained based on weighted singular-valued decomposition of matrices with singular weights. Boundary representations of weighted pseudoinverse matrices with singular weights are obtained. Regularization methods for the calculation of weighted normal pseudosolutions with singular weights are constructed and investigated.

Key words: *weighted singular-valued decomposition of matrices with singular weights, weighted pseudoinverse of matrices.*

Одержано 31.02.2019

UDC 330.341
DOI: 10.32626/2308-5878.2019-19.17-21

J.-F. Emmenegger, Dr.,
D. Chable, Dipl. math.,
H. A. Nour Eldin, Prof. Dr.,
H. Knolle, Dr. math.

University of Fribourg, Switzerland

SRAFFA AND LEONTIEF REVISITED. MATHEMATICAL METHODS AND MODELS OF A CIRCULAR ECONOMY (BOOK PRESENTATION)

The main purpose of the present book is to reveal, elucidate and illustrate the mathematical background of Sraffa's theory didactically in detail with the means of modern *matrix algebra* and the corresponding fundamental theorems. Our book is also a contribution to the increasing call for alternative approaches to the understanding of the realities of today economic activity.

Key words: *economics, productive Leontief model, Frobenius number.*

Economics[1] *is a decision-based and number-based science.*
«Currency and market decisions in a decision-based economy» [1]
«All what we are doing should be based theoretically».
Andrei Broder, scientist, Google (~2017)

[1] The Webster New Collegiate Dictionary, p. 260, defines the term «Economics» as follows: *«The science that investigates the conditions and laws affecting the production, distribution and consumption of wealth, or material means of satisfying human desires; political economy»*.

Preliminary Remarks. This book focuses mainly on Sraffa's theoretical model Production of Commodities by Means of Commodities ([2], PCMC, 1960). It interprets and extends PCMC following the initial footsteps undertaken by Newman, Pasinetti and especially Schefold in the German version of PCMC [3]. Matrix algebra is used applying the mathematical and notational standards set By Miller and Blair [4] and the standards defined by EUROSTAT [5]. The dominating importance of the Perron-Frobenius Theorem, ensuring the existence of a solution to Sraffa's model of production, reformulated as an eigenvalue problem, is stressed, together with an important result due to Ashmanov's book ([6], Theorem 1.5, p. 3)] concerning *Leontief models*, *productive Leontief model*s and their Frobenius number. These techniques are supplemented by elements of graph theory.

Piero Sraffa (1898–1983), a classical economist, reformulated in his book *Production of Commodities by Means of Commodities (PCMC) «the theory of value and distribution».* Wassili Leontief (1906–1999) [7] made early contributions to *input-output analysis* and earned the Nobel Prize in Economics in 1973. Sraffa and Leontief are concerned with the whole structure of production, considered in its totality as a cyclic process. A matrix describes quantitatively the exchange between the branches of the economy.

On less than 100 pages Sraffa uses in PCMC mathematical concepts and theorems which he has mainly hidden. He just presents calculus and numerical results. The main purpose of the present book is to reveal, elucidate and illustrate the mathematical background of Sraffa's theory didactically in detail with the means of modern *matrix algebra* and the corresponding fundamental theorems. A large place is given for computed examples.

Our book is also a contribution to the increasing call for alternative approaches to the understanding of the realities of today economic activity. In writing this book we have stood on the shoulders of eminent Sraffa connoisseurs: P. Newman, 1962; B. Schefold, 1976 and, 1989; L. L. Pasinetti, 1977, 1980 and (1986); H. D. Kurz and N. Salvadori [8], 2007; A. Roncaglia, 2009.

Wassili Leontief (1906–1999) models the economic activity within the context of a circular economy of production and exchange, today expressed in *Input-Output Tables* in *monetary terms*. Leontief proposed to divide the economy into sectors, each one producing a group of products. There is a highly technical process to achieve this partition described by the NACE Rev. 2 report [9], and the CAP nomenclature, leading to *Input-Output Tables* and *Input-Output Analysis*, see the *European Union* [5].

Independently of Leontief, Sraffa in PCMC [2], linearly modeled the English production of single commodities, like *wheat*, *iron* or *pigs*, considering the circularity of these production processes expressed in *physical terms*. He solved the distribution problem of David Ricardo (1772–1823), determined the production prices and the labour values. Sraffa's and Leontief's approach both need *matrix algebra* and the fundamental theorem of Frobenius [10].

Summary of the Chapters. This summary begins with the second chapter, the first one being the *Introduction.*

Chapter 2 gives a rigorous, detailed and ahead presentation of the set of matrices and vectors used in Input-Output Analysis. The notations and matrix algebra involved permit an advanced presentation of the material. The elements of the Leontief Input-Output Tables (IOT) are then presented. The principles of the system of *Classification of products by activities (CAP)*, respectively the *Nomenclature des activits économiques dans la communauté européenne (NACE)* are explained. They are at the basis of the determination of the *homogenous branches*, constituting the IOTs. A selection of Leontief Input-Ouput models are presented.

Chapter 3 is a complete discussion of the three elementary examples figuring at the beginning of PCMC, described now in terms of matrix algebra and introducing the **Perron-Frobenius Theorem** [10] as the centre piece of the algebraic structure of Sraffa's models.

Chapter 4 develops the complete theory of Sraffa's price model for single commodity production processes, examining in particular the distinction between basic and non-basic commodities. Examples are calculated, determining all involved economic variables and economic ratios.

Novelties: the general relationship between the rate of profits, the surplus ratio and the ratio of total wages to national income, valid for all Sraffa systems; the introduction of directed graphs and bipartite networks as tools for the analysis of all types of Sraffa production processes.

Chapter 5 presents the complete theory of the Standard system of production for single commodity processes, including the famous linear relationship between the rate of profits, the Standard ratio and the share of total wages to national income, valid for all Standard systems.

Novelties: explicit formulation of the fundamental relations forming the basis of a Standard system; the introduction of the notion of the *commodity space* and the *orthogonal Euler mapping* (Euler affinity) central for the transformation of an actual non standard System into a *Standard system*.

Chapter 6 is an introduction to *joint production systems*, where the same commodities may be produced by more than one industry.

Novelties: output polyhedrons; a compact algebraic methodology for the distinction between basic and non-basic commodities which completes the approaches of Manara-Pasinetti-Schefold to the case of *joint production systems*; a matrix introduced by Pasinetti [11] shows to be pertinent to determine the number of *basics*.

Chapter 7. This chapter has been proposed and developed by H. Knolle.

Joint production systems are considered with industries producing in parallel several commodities, a typical situation with ecological conse-

quences. Then, new approaches and examples treating waste problems and presenting situations involving ecological economics and taxation are treated. The main proposition is to show that Sraffa's price model offers an approach to treat ecological problem, involving waste in the whole economic system. The consequences on the prices are studied.

Chapter 8 is entirely concentrated to novel extensions of Sraffa's price models as indicated in the corresponding item of the Table of Contents.

Chapter 9 is a complete formal algebraic analysis of the *interindustrial economy*, developed by H. A. Nour Eldin. Tables of matrices presenting synoptically the aspect of *value*, *quantities*, *prices* and *objects* of Leontief's and Sraffa's concepts. It cumulates in the statement that the *Interindustrial Market* together with the *Consumption Market* is unified to the *Leontief-Sraffa economy*. Each one of these three entities is described by a proper sets of matrices and vectors.

Chapter 10 goes beyond simplified educational examples and gives a presentation of how Sraffa's approach, together with the IOT apparatus, can be applied to official IOTs. In this case we apply the developed methodology to the official Swiss IOTs 2008 and 2014 and the German IOT 2013. We also compute the productiveness of these economies. We perform some aggregations of the official IOTs and show the limits of these calculations.

Chapter 11 summarises the results obtained in this book and indicates proposed avenues of research in an extended Sraffa context.

Chapter 12 (Appendix I) contains all the necessary mathematical tools required for a complete understanding of the present text.

Chapter 13 (Appendix II) is a summary of Schefold's historical contribution to the understanding of Sraffa's PCMC.

Chapter 14 Glossary of terms.

Acknowledgments go to the publishing House de Gruyter-Oldenbourg and the encouraging and professional support of Mrs. Kristin Berber-Nerlinger, lecturer, and Mrs. Nadja Schedensack, technical support, then to the artist Karim Noureldin, who permitted to use his splendid artwork for the cover of the book. Without the competent organizational work accomplished by Tamara Bardadym, encouraging and animating the colleagues of the Kiev group, our book could not have been written. Tamara Bardadym contributed to the book with competent proof-reading of the Lemmas, Theorems and the proofs.

References:

1. Nour Eldin, H. A., Emmenegger, J.-F., Nabout, A. A. Currency and Market Decisions. *Decesion-Based Economy.* 2019.
2. Sraffa P. Warenproduktion mittels Waren (aus dem Englischen ubersetzt mit einem Anhang von B. Schefold), Edition Suhrkamp 780, Erste Auage. 1976.

3. Miller R. E., Blair P. D. Input-Output Analysis, Foundations and Extensions. Cambridge University Press, 2nd edition. 2009.
4. Carré H. NACE Rev. 2, Statistical classification of economic activities in the European Community. *General and regional statistics*, Methodologies and working papers, European Community. 2008.
5. Norlund L. Eurostat Manual of Supply, Use and Input-Output Tables. *Luxembourg: Office for Official Publications of the European Community*. 2008.
6. Ashmanov S. A. Introduction in Mathematical Economics (in Russian language). Moscow : Nauka, 1984. 296 p.
7. Leontief W. W. Die Wirtschaft als Kreislauf. *Archiv für Sozialwissenschaft und Sozialpolitik*. 1928. Vol. 60. P. 577–623.
8. Sraffa P. Production of Commodities by means of Commodities. Cambridge : Cambridge University Press. 1960.
9. Kurz H. D., Salvadori N. Theory of Production. *A Long-Period Analysis*. Cambridge University Press, paperback. 1995. 2007.
10. Frobenius G. Über Matrizen aus nicht negativen Elementen. Berliner Bericht. 1912. P. 456–477.
11. Pasinetti L. L. (ed.). Essays on the Theory of Joint Production. New York : Columbia Univ. Press 1980.

СРАФФА И ЛЕОНТЬЕВ. ПЕРЕСМОТР. МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ КРУГОВОЙ ЭКОНОМИКИ (ПРЕЗЕНТАЦИЯ КНИГИ)

Основная цель настоящей книги — детально раскрыть, прояснить и проиллюстрировать математическое обоснование теории Сраффы с помощью современной теории матриц и соответствующих фундаментальных теорем. Наша книга также является ответом на появившийся запрос на альтернативные подходы к пониманию реалий современной экономической деятельности.

Ключевые слова: *экономика, продуктивная модель Леонтьева, число Фробениуса.*

Date received 12.03.2019

УДК 519.64;519.65
DOI: 10.32626/2308-5878.2019-19.22-28

В. К. Задірака, д-р фіз.-мат. наук, професор, академік НАН України,
Л. В. Луц, канд. фіз.-мат. наук

Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ЕЛЕМЕНТИ ТЕОРІЇ ОПТИМАЛЬНОГО ІНТЕГРУВАННЯ ШВИДКООСЦИЛЮЮЧИХ ФУНКЦІЙ НА КЛАСАХ ФУНКЦІЙ

Наведені елементи теорії побудови (при даній інформації про підінтегральну функцію) оптимальних за точністю квадратурних формул обчислення інтегралів від швидкоосцилюючих функцій для певних класів підінтегральних функцій.

Як осцилюючі функції розглядаються: $e^{-i\omega x}$, $e^{i\omega g(x)}$, $\sin \omega x$, $\cos \omega x$, вейвлет-функція $\psi(x)$ з компактним носієм, $J_m(\omega x)$ — функції Бесселя першого роду порядку m.

Отримані результати для перелічених осцилюючих функцій дали змогу створити теорію оптимального інтегрування швидкоосцилюючих функцій як у класичній постановці, так і для інерполяційних класів функцій.

Значна увага приділена виявленню та уточненню апріорної інформації про підінтегральну функцію та її використання для звуження звичайних (класичних) класів підінтегральних функцій до інтерполяційних [1]. Функції, які входять у такі (інтерполяційні) класи, не розрізняються квадратурними формулами (для всіх них наближене значення інтегралу буде одне і те ж саме).

Друга особливість результатів полягає (на відміну від результатів усіх інших авторів) в припущенні наближеного задання вхідної інформації про підінтегральну функцію. Розгляд інтерполяційних класів дозволяє підвищити потенційну спроможність квадратурних формул.

Аналізуються комп`ютерні технології (КТ) інтегрування швидкоосцилюючих функцій з заданою точністю.

Ключові слова: *квадратурна формула, метод капелюхів, метод граничних функцій, апріорна інформація, функції Бесселя, комп`ютерна технологія.*

Вступ. При розв`язанні ряду важливих класів задач, таких як задачі математичної фізики, цифрової обробки сигналів і зображень, прикладної статистики, моделювання оптичних систем та синтезованих голограм, аналіз і синтез мовних сигналів, математичного моделювання, інформаційної безпеки, виникає необхідність в обчисленні швидкоосцилюючих інтегралів вигляду

$$I(\omega)=\int_a^b f(x)\left\{\begin{array}{c} e^{-i\omega x} \\ e^{i\omega g(x)} \\ \sin\omega x \\ \cos\omega x \\ \psi\left(\dfrac{x-t}{s}\right) \\ J_m(\omega x) \end{array}\right\}dx \qquad (1)$$

у припущенні, що $f(x)\in F$, F — деякий[1] клас функцій, заданих на відрізку $[a,b]$, $|\omega|\geq 2\pi(b-a)$, інформація про значення $f(x)$ задається не більше ніж в N вузлових точках $\{x_i\}_0^{N-1}$ з $[a,b]$.

Відмінність отриманих результатів від результатів Л. Файлона, Л. Коллатца, В. І. Крилова, М. В. Ніколаєвої, В. Л. Рвачова, М. С. Бахвалова, Я. М. Жилейкіна, Б. Ейнарсона та інших дослідників полягає у наступному:

- інформація про $f(x)$ задана наближено;
- $f(x)$ занурюється у більш вузькі інтерполяційні класи функцій F_N, $F_{N,\varepsilon}$;
- використовуються алгоритми виявлення та уточнення апріорної інформації (гладкість, області монотонності, кількість точок перегину, опуклість, константа та показник Гельдера та ін.) [3];
- для деяких класів вдалося побудувати оптимальні за точністю або асимптотично оптимальні квадратурні формули;
- запропоновані КТ розв'язання задач інтегрування швидкоосцилюючих функцій.

Це дозволяє зменшити чебишевський радіус області невизначеності значень інтегралу та покращити потенційну спроможність квадратурних формул.

Постановка задачі. Позначимо $R=R(f,A,\omega)$ — результат наближеного обчислення $I(\omega)$ за допомогою квадратурної формули A.

Введемо характеристики:

$$V(f,A,\omega)=\rho\left(I(\omega),R\right),\ V(F,A,\omega)=\sup_{f\in F}V(f,A,\omega),$$
$$V=V(F,\omega)=\inf_{A}V(F,A,\omega), \qquad (2)$$

[1] Результати отримані для 45 класів підінтегральних функцій F, F_N, $F_{N,\varepsilon}$ [2].

де $\rho\left(I(\omega), R\right) = \left|I - R\right|$ — похибка чисельного інтегрування. Квадратурну формулу A^*, на якій досягається $V = V\left(F, \omega\right)$, назвемо оптимальною за точністю. Якщо для квадратурної формули $\bar{A}$ виконується $V(F, \bar{A}, \omega) \le V\left(F, \omega\right) + \eta$, то $\bar{A}$ назвемо оптимальною з точністю до η. Якщо $\eta = o\left[V\left(F, \omega\right)\right]$ або $\eta = O\left[V\left(F, \omega\right)\right]$, то $\bar{A}$ назвемо відповідно асимптотично оптимальною або оптимальною за порядком точності. Для отримання оцінок похибки чисельного інтегрування $I\left(\omega\right)$ на класах підінтегральних функцій F використовується метод «капелюхів» [1]. Він дозволяє отримати оцінку *V*, а не її саму. Це обумовлює застосовність даного методу для побудови лише оптимальної з точністю до η квадратурної формули.

Підвищення «потенційної спроможності» квадратурних формул може бути здійснене шляхом «звуження» відповідного класу *F* на інтерполяційні класи F_N, які визначаються належністю класу *F* та ще 2*N* фіксованими значеннями інформаційного оператора: $\left\{x_i\right\}_0^{N-1}$ і $\left\{f_i\right\}_0^{N-1}$. В цьому випадку можна ввести за аналогією з (2) характеристики:

$$\delta(f, A, \left\{f_i\right\}_0^{N-1}, \omega) = \rho(I, R),$$

$$\delta(F_N, A, \left\{f_i\right\}_0^{N-1}, \omega) = \sup_{f \in F_N} \delta(f, A, \left\{f_i\right\}_0^{N-1}, \omega),$$

$$\delta = \delta(F_N, \left\{f_i\right\}_0^{N-1}, \omega) = \inf_{A} \delta(F_N, A, \left\{f_i\right\}_0^{N-1}, \omega), \qquad (3)$$

і визначити оптимальні за точністю, асимптотично оптимальні і оптимальні за порядком точності квадратурні формули на класі F_N.

Для побудови й обґрунтування оптимальних за точністю і близьких до них квадратурних формул обчислення $I(\omega)$ в класах F_N застосовується метод граничних функцій [1].

Для прикладу розглянемо побудову оптимальних за точністю квадратурних формул обчислення інтегралу $I_1(\omega) = \int_a^b f(x) \sin \omega x dx$ та оцінок їх похибки на класі $C_{L,N}$ ($C_{L,N}$ — інтерполяційний клас визначених на [*a*, *b*] функцій, що задовольняють умові Ліпшиця $\left|f(x_1) - f(x_2)\right| \le L\left|x_1 - x_2\right|$, $x_1, x_2 \in \left[a, b\right]$).

Теорема 1. Нехай $f\left(x\right) \in C_{L,N}$, $\left[\left|\omega\right|\left(b-a\right)/\pi\right] + 1$ нулів функції $\sin \omega x$ на $\left[a, b\right]$ входять в число вузлів x_i, $i = \overline{0, N-1}$, і $\omega > 0$. Тоді для похибки обчислення $I_1(\omega)$ справедлива наступна оцінка знизу:

$$\delta\left(C_{L,N},\{f_\upsilon\}_0^{N-1},\{x_\upsilon\}_0^{N-1},\omega\right)\geq$$

$$\geq\begin{cases}\dfrac{4L}{\omega^2}\displaystyle\sum_{\upsilon=0}^{N-1}\left\{\sin^2\frac{\omega\Delta x_\upsilon}{4}-\sin^2\frac{\omega|\Delta f_\upsilon|}{4L}\right\}\left|\sin\left(\frac{\omega}{2}\left(x_\upsilon+x_{\upsilon+1}\right)\right)\right|+P(\omega),\ N\geq|\omega|,\\ \dfrac{L}{\omega}\left\{\dfrac{2}{\pi}+\dfrac{\pi}{\omega+\pi}-\dfrac{4}{\omega}\displaystyle\sum_{\upsilon=0}^{\omega/\pi}\sin^2\frac{\omega|\Delta f_\upsilon|}{4L}\right\},\ N=\left[\frac{\omega}{\pi}\right]+1,\end{cases}\quad(4)$$

де $P(\omega)=\dfrac{L}{2}\left|\dfrac{2}{\omega}\sin\dfrac{\omega\Delta x_{N-1}}{2}\cos\left(\omega\left(1-\dfrac{\Delta x_{N-1}}{2}\right)\right)-\left(1-x_{N-1}\right)\cos\omega\right|$, $\Delta f_\upsilon=$ $=f_{\upsilon+1}-f_\upsilon$, причому квадратурна формула

$$R_1(\omega)=\sum_{\nu=0}^{N-1}\int_{x_\nu}^{x_{\nu+1}}f_1^*(x)\sin\omega x dx,\qquad(5)$$

$$f_1^*(x)=\begin{cases}f_\upsilon\ , & x_\upsilon\leq x\leq\bar{x}_\upsilon,\\ f_\upsilon+L\left(x-x_\upsilon\right)sign\left(\Delta f_\upsilon\right), & \bar{x}_\upsilon\leq x\leq\overline{=}{x}_\upsilon,\\ f_{\upsilon+1}\ , & \overline{=}{x}_\upsilon\leq x\leq x_{\upsilon+1},\\ f_{N-1}\ , & x_{N-1}\leq x\leq x_N,\end{cases}\quad x\notin\left[x_{N-1},x_N\right]\ ,$$

$$\bar{x}_\upsilon=\frac{x_\upsilon+x_{\upsilon+1}}{2}-\frac{|\Delta f_\upsilon|}{2L},\ \overline{=}{x}_\upsilon=\frac{x_\upsilon+x_{\upsilon+1}}{2}+\frac{|\Delta f_\upsilon|}{2L},$$

є оптимальною за точністю при $N\geq|\omega|$ і $N=\left[|\omega|/\pi\right]+1$.

Доведення проводиться методом граничних функцій. Спочатку будуються мажоранта і міноранта $f^{\pm}(x)$ класу $C_{L,N}$. Очевидно, що клас функцій $C_{L,N}$ на кожному $[x_\upsilon\ ,x_{\upsilon+1}]$ обмежений прямими $f_\upsilon\pm L\left(x-x_\upsilon\right)$ і $f_{\upsilon+1}\pm L\left(x_{\upsilon+1}-x\right)$, які попарно перетинаються в точках $\bar{x}_\upsilon$ и $\overline{=}{x}_\upsilon$ (рисунок).

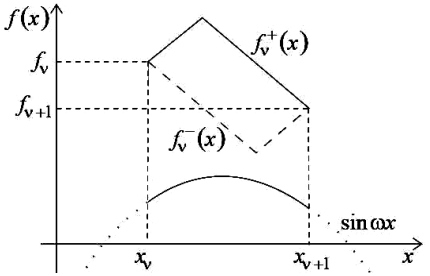

Рисунок. *Загальний вигляд функцій* $f^+(x),\ f^-(x)\in C_{L,N}$

Враховуючи знак функції $\sin \omega x$ на відрізку $\left[x_\upsilon, x_{\upsilon+1}\right]$ та окремо розглядаючи відрізки $\left[x_\upsilon, \bar{x}_\upsilon\right]$, $\left[\bar{x}_\upsilon, \bar{\bar{x}}_\upsilon\right]$, $\left[\bar{\bar{x}}_\upsilon, x_{\upsilon+1}\right]$, знаходимо аналітичні вирази для $f^{\pm}(x)$. Скориставшись ними, будуємо чебишевський центр і чебишевський радіус області невизначеності розв'язку задачі (1), які є відповідно квадратурною формулою $R_1(\omega)$ і оптимальною оцінкою δ її похибки на класі $C_{L,N}$.

Аналогічний результат отриманий для $f(x) \in C_{L,N,\varepsilon}$, де $C_{L,N,\varepsilon}$ — клас функцій $C_{L,N}$ з наближено заданою апріорною інформацією про значення $f(x)$: $\left|\tilde{f}_i - f_i\right| \le \varepsilon_i$, $i = \overline{0, N-1}$. Для побудови граничних функцій $f^{\pm}(x) \in C_{L,N,\varepsilon}$ попередньо проводиться згладжування вхідних даних і уточнення ε_i шляхом розв`язання деякої системи нерівностей [1]. В роботі [4] отримані оцінки повної похибки побудованих квадратурних формул.

При неточно заданій апріорній інформації про клас пропонуються близькі до оптимальних за точністю квадратурні формули обчислення $I(\omega)$ на основі методів нев'язки, квазірозв`язків. У роботі [3] розроблений алгоритм виявлення та уточнення порядку похідної функції, показника Гельдера та константи Ліпшиця, яким вона задовольняє, при сітково заданому інформаційному операторі.

Для випадку чисельного інтегрування перетворення Бесселя

$$I_m(\omega) = \int_a^b f(x) J_m(\omega x)\, dx,$$

де $J_m(\omega x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(m+k+1)} \left(\frac{\omega x}{2}\right)^{2k+m}$ — функція Бесселя першого роду порядку m, в [2] при $[a,b] = [0,1]$ запропоновано квадратурні формули типу Файлона

$$I_m(\omega) \approx R_{m,1}(\omega) = \int_0^1 P(x) J_m(\omega x)\, dx, \qquad (6)$$

$$I_m(\omega) \approx R_{m,2}(\omega) = \int_0^1 S(x) J_m(\omega x)\, dx, \qquad (7)$$

де $P(x)$ — параболічний сплайн, що будується на $[i/N, (i+1)/N]$ у вузлах $c_{1,i} = i/N$, $c_{2,i} = (2i+1)/2N$, $c_{3,i} = (i+1)/N$, $S(x)$ — ерміто-

вий кубічний сплайн, що будується на $[i/N,(i+1)/N]$ у вузлах $c_{1,i}=i/N$, $c_{2,i}=(i+1)/N$, $i=\overline{0,N-1}$. Для квадратурних формул (6) та (7) справедливі такі оцінки похибки відповідно:

$$V\left(f,R_{m,1},\omega\right)\frac{\sqrt{3}Ah^3}{36\sqrt[3]{\omega}}\max_{0\le x\le 1}\left|f^{(3)}(x)\right|,$$

$$V\left(f,R_{m,2},\omega\right)\le\frac{Ah^4}{384\sqrt[3]{\omega}}\max_{0\le x\le 1}\left|f^{(4)}(x)\right|,$$

де $h=1/N$, $\omega\ge 1$, A — деяка константа.

У монографії [2] обгрунтовано і покроково описано інформаційно-комп`ютерну технологію обчислення інтеграла (1) з заданими значеннями характеристик якості розв`язку, яка дозволяє залучати і ефективно використовувати для розв`язання задачі (1) необхідні резерви оптимізації обчислень.

Висновки. Наведені елементи теорії побудови оптимальних за точністю квадратурних формул обчислення інтегралів від швидкоосцилюючих функцій (1) на класах функцій, яка, на відміну від інших результатів, відбувається за умов найповнішого врахування апріорної інформації про $f(x)$. Поряд з класами F розглядаються інтерполяційні класи F_N, та класи $F_{N,\varepsilon}$ з наближеним заданням інформації.

Список використаних джерел:

1. Задирака В. К. Теория вычисления преобразования Фурье. Киев : Наук. думка, 1983. 215 с.
2. Сергієнко І. В., Задірака В. К., Литвин О. М. та ін. Оптимальні алгоритми обчислення інтегралів від швидкоосцилюючих функцій та їх застосування: у 2 т. Київ : Наук. думка, 2011. Т. 1: Алгоритми. 448 с.; Т. 2.: Застосування. 348 с.
3. Луц Л. В., Задірака В. К. Наближене інтегрування швидкоосцилюючих функцій з виявленням і уточненням апріорної інформації. *Математичне та комп`ютерне моделювання.* 2017. Вип. 15. С. 100–106.
4. Луц Л. В. Оцінка якості деяких квадратурних формул обчислення інтегралів від швидкоосцилюючих функцій. *Штучний інтелект*. 2008. №.4. С. 671–682.

THE ELEMENTS OF THE THEORY OF THE OPTIMAL INTEGRATION OF HIGHLY OSCILLATING FUNCTIONS ON CLASSES OF FUNCTIONS

The elements of the theory of construction (with given information on the integral function) of the optimal quadrature formulas for calculating the integrals of highly oscillatory functions for certain classes of integral functions are presented.

As oscillatory functions are considered: $e^{-i\omega x}$, $e^{i\omega g(x)}$, $\sin\omega x$, $\cos\omega x$, wavelet-function $\psi(x)$ with compact carrier, $J_m(\omega x)$ — Bessel functions of the first kind of order *m*.

The obtained results for the listed oscillatjry functions allowed us to create a theory of optimal integration of highly oscillatory functions both in classical formulation and for interpolation classes of functions.

Considerable attention is paid to the identification and refinement of a priori information about the integral function and its use for narrowing the usual (classical) classes of integral functions to interpolation classes [1]. The functions included in such (interpolation) classes do not differ in quadrature formulas (the approximate integral value will be the same for them all).

The second feature of the results is (in contrast to the results of all other authors) in the assumption of an approximate input of information about the integral function. Examining interpolation classes can increase the potential of quadrature formulas.

Computer technologies (CTs) of the integration of highly oscillatory functions with given accuracy are analyzed.

Key words: *quadrature formula, cap method, method of boundary functions, a priori information, Bessel functions, computer technology.*

Одержано 29.01.2019

УДК 004.056.55
DOI: 10.32626/2308-5878.2019-19.28-34

О. Г. Качко*, канд. техн. наук,
С. О. Кандій,** студент,
Є. В. Острянська,** студент

*АТ «Інститут інформаційних технологій», м. Харків,
**Харківський національний університет імені В. Н. Каразіна, м. Харків

ОПТИМІЗАЦІЯ ФУНКЦІЇ МНОЖЕННЯ ПОЛІНОМІВ ДЛЯ ЗВИЧАЙНОЇ ТА PRODUCT ФОРМИ ЗАДАННЯ ОДНОГО З ПОЛІНОМІВ

У роботі проведено дослідження та виконано розробку ефективного алгоритму множення тернарного полінома у кільці $\mathrm{Z}_3[x]\left(x^n - x - 1\right)$ з урахуванням його структури. Розглядаються варіанти для поліномів із звичайною структурою з фіксованою кількістю ненульових елементів «1» («–1») та у PRODUCT-формі, у якій поліном є результатом обчислення $F_1 * F_2 + F_3$, де $F_1, F_2, F_3 \in \mathrm{Z}_3[x]\left(x^n - x - 1\right)$ та мають відповідно d_1, d_2, d_3 елементів із значеннями «1» та «–1». Приводяться результати оптимізації за допомогою векторизованих наборів інструкцій (а саме, набір інструкцій AVX2), розпаралелювання та спеціальних засобів для мінімізації та компенсування використання не

вирівняної пам'яті. Критичний код написано на асемблері під мікропроцесорну архітектуру x86-64, яка є однією з найрозповсюдженіших на сьогоднішній день. Отримані часові показники оптимізованої реалізації алгоритму для наборів параметрів для 256, 384 та 512 біт класичної безпеки та зроблено порівняння ефективності з алгоритмом множення поліномів, що був запропонований у асиметричній постквантовій криптосистемі на алгебраїчних решітках NTRU Prime. Тестування здійснено на процесорі Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz на операційній системі Linux 4.15.0-44-generic #47-Ubuntu SMP x86-64. Результати, що отримані, на наш погляд — надзвичайно актуальні. Вони можуть бути корисними для криптологів та інших фахівців, що займаються розробкою нових, ефективних криптографічних алгоритмів та протоколів для постквантового періоду. Це пояснюється тим, що для класів криптосистем, у яких використовуються перетворення у кільцях поліномів, як базові операції, а саме операцій множення, займає найбільше часу та тому потребує значної оптимізації. Значна перевага розробленого алгоритму — можливість його розпаралелювання на багатопроцесорних системах, що є значною перевагою порівняно з алгоритмом, що представлені в NTRU Prime.

Ключові слова: *NTRU Prime, PRODUCT- форма , алгебраїчна решітка, кільце поліномів, множення поліномів, постквантова криптографія, тернарний поліном.*

Вступ. На сьогоднішній день квантові комп'ютери розвиваються з великою швидкістю — це значна загроза для існуючих асиметричних криптосистем. Один з найперспективніших напрямків — це криптографія на алгебраїчних решітках. В таких системах більшість обчислень зводиться до множення поліномів, один з яких є тернарний. Існує багато алгоритмів для множення поліномів, але ці алгоритми або не враховують спеціальну структуру поліномів [1], або не забезпечують константний час множення. У даній роботі запропоновано алгоритм, що використовує особливості тернарних поліномів, та забезпечує константний час. Функція множення поліномів обчислює $c = a * b$, де a — тернарний поліном, b — поліном у полі $Z_q[x](x^n - x - 1)$. Для завдання поліному a використовується 2 форми.

Форма 1. a задається як поліном у полі $Z_3[x](x^n - x - 1)$ з фіксованою кількістю ненульових елементів;

Форма 2. a задається у формі $F_1 * F_2 + F_3$, де кожний з поліномів F_1, F_2, F_3 — це поліноми у полі, $Z_3[x](x^n - x - 1)$ містять однакову кількість «1» та «–1». Позначимо цю кількість відповідно d_1, d_2, d_3.

Множення для звичайної форми. У самому найпростішому «шкільному» методі послідовно виконується множення коефіцієнтів одного поліному на фіксований коефіцієнт іншого і накопичення суми для коефіцієнтів з однаковими номерами. Номер коефіцієнта визначається сумою номерів коефіцієнтів, які перемножаться. Формально це можна представити у вигляді псевдокоду 1.

Псевдокод 1

```
1.  for (i = 0; i < n; i++)
2.  for (j = 0; j < n; j++)
3.          c[i + j]+= a[i]∗b[j].
```

Оскільки основна кількість коефіцієнтів A — нульові елементи, то більшість операцій множення $a[i]$ на $b[j]$ не мають сенсу. Основна ідея нового алгоритму полягає у тому, щоб запам'ятати індекси ненульових елементів та замість множення $a[i]$ на $b[j]$ виконувати лише віднімання або додавання $b[j]$ в залежності від значення $a[i]$. Формально це можна представити у вигляді псевдокоду 2.

Псевдокод 2

```
1. c = 0;
2  for (i = 0; i < n; i++){
3.          if (a[i] == 1){
4.            for (j = 0; j < n; j++) {
5.              c[i+j] += b[j];
6.               if (c [i + j] >=q) c [i + j]-=q;
7.          }}else if (a[i] == -1){
8.               for (j = 0; j < n; j++){
9.                   c[i+j] -= b[j];
10.                  if (c [i + j] <0) c [i + j]+=q;
11.       }}}
```

Крім того, перед виконанням циклу усі елементи результату обнуляються, що гарантує їх наявність в кеші. При цьому загальний розмір не перевищує розміру кешу 1 рівня, тобто механізм витіснення за рахунок недостатнього розміру не буде спрацьовувати. Використання структури забезпечує неможливість перекриття адрес.

Передбачається використання SIMD операцій (AVX-2). Значення коефіцієнтів завжди цілі, тому використовується тип __m256i. Згідно значень параметрів [2] значення $c[i + j] < 2^{15} - 1$, тобто компоненти блоку 16 бітні знакові числа, а блок містить 16 таких компонентів і одночасно обробляється 16 коефіцієнтів.

Вищенаведений оптимізований варіант для псевдокоду передбачає:

- мінімізацію звернень до не вирівняної пам'яті (за рахунок перед обчислень) і компенсацію решти звернень за рахунок використання спеціальних операцій процесора;
- видалення операцій умовного переходу (за рахунок використання AVX-2 команд);
- мінімізацію промахів кешу (за рахунок обрання ефективної структури співмножників);
- паралельну обробку порцій коефіцієнтів. Кількість одиниць та мінус одиниць може не співпадати, тому порція для паралельного виконання містить і одиниці і мінус одиниці, в цьому разі їх кількість приблизно однакова для усіх порцій. Для збільшення навантаження на паралельну гілку кожна гілка приводить частковий результат не тільки по модулю q, а і по модулю p.

Результати тестування функції множення після оптимізації для набору параметрів з [2] наведені у табл. 1 (тактів).

Таблиця 1

Результати тестування множення для звичайної форми

K	**Linux**
256 (N = 761; q = 4591; T = 143)	7564
384 (N = 1031; q = 8297; T = 172)	11864
512 (N = 1301; q = 10427; T = 217)	16347

Для порівняння, час виконання функції множення з роботи [1] для того ж процесору, тих же режимів компіляції і для тих же ключів дорівнює 12321 тактів (K = 256, N = 761, q = 4591, t = 143). Прискорення в порівнянні з [1] складає 38.6 %

Множення для PRODUCT-форми. Для забезпечення ефективного використання кешу 1 рівня ключ задається як структура:

```
struct{
    unsigned short ones1[ d1 ], minusones1[ d1 ];
    unsigned short ones2[ d2 ], minusones2[ d2 ];
    unsigned short ones3[ d3 ], minusones3[ d3 ];
}
```

де d_1, d_2, d_3 — кількість 1 та –1 в F_1, F_2, F_3 відповідно; ones1, ones2, ones3 — масив з індексами відповідних коефіцієнтів, які дорівнюють 1; minusones1, minusones2, minusones3 — масиви з індексами відповідних коефіцієнтів, які дорівнюють –1.

Є 2 варіанти обчислення:

1) спочатку обчислити $F = F_1 * F_2 \bmod p + F_3$, а потім $F * h$;

2) спочатку обчислити $F' = \left(\left(F_1 * h\right) \bmod q\right) \bmod p$. Далі обчислити $F'' = \left(\left(F' * F_2\right) \bmod q\right) \bmod p$, а потім $F''' = F'' + \left(\left(\left(F_3 * h\right) \bmod q\right) \bmod p\right)$. Еквівалентність $F = F'''$ забезпечується на етапі генерації ключів.

Перевірка показала, що відсоток ключів, для яких ця умова не виконується, дуже незначний. Результати експериментальної перевірки для 100000 ключів наведені в табл. 2.

Аналіз показав, що перший варіант не дозволяє використовувати операції додавання та віднімання замість операції множення, тому що коефіцієнти обчисленого поліному можуть відрізнятися від 1, –1. Другий варіант дозволяє це зробити, тому далі будемо використовувати другий варіант.

Таблиця 2

Кількість відбракованих ключів (всього 100000 ключів)

K	**Кількість**	**%**
256 (N = 787; q = 7307; D_1 = 12; D_2 = 12; D_3 = 15)	102	0.1
384 (N = 1019; q = 8867; D_1 = 13; D_2 = 13; D_3 = 31)	60	0,06
512 (N = 1301; q = 11959; D_1 = 15; D_2 = 15; D_3 = 48)	50	0,05

Для обчислення часткових добутків F_1', F_2', $F_3 * h$ використовуються засоби оптимізації, які наведені вище. Для цього варіанту кількість одиниць та мінус одиниць у кожній з трьох функцій співпадає, крім того, кількість ненульових елементів у кожній з функцій значно менше, ніж при звичайному завданні, тому принцип розподілу обчислень між паралельними гілками відрізняється від попереднього.

Аналіз параметрів (див. табл. 3) показує що кількість ненульових елементів $D_1 + D_2$ в більшості випадків не перевищує D_3, більше того F_1 та F_2 пов'язані між собою знаком множення.

Пропонується паралельно обчислювати:

$$\left(\left(\left(\left(\left(F_1 * h\right) \bmod q\right) \bmod p\right) * F_2\right) \bmod q\right) \bmod p\Big)$$

та

$$\left(\left(\left(F_3 * h\right) \bmod q\right) \bmod p\right).$$

Значення D_1, D_2, D_3 — параметри алгоритму, тому різниця в часі для паралельних гілок не залежить від місця ненульових коефіцієнтів.

Пам'ять під внутрішні значення подвійної довжини може бути локальною, що забезпечує використання свого кешу для їх зберігання.

Результати тестування функції множення для product-форми після оптимізації для набору параметрів з [2] наведені у табл. 3 (тактів).

Таблиця 3

Результати тестування множення для PRODUCT — форми

K	**Linux**
256 (N = 787; q = 7307; D_1 = 12; D_2 = 12; D_3 = 15)	6499
384 (N = 1019; q = 8867; D_1 = 13; D_2 = 13; D_3 = 31)	8468
512 (N = 1301; q = 11959; D_1 = 15; D_2 = 15; D_3 = 48)	14718

Для порівняння, час виконання функції rq_mult [1] для того ж процесору і для найближчих параметрів (K = 256, N = 761, q = 4591, t = 143) дорівнює 12321 тактів. Прискорення в порівнянні з [1] практично в 2 рази. Використання multiply форми замість звичайної забезпечує прискорення.

Висновки. Визнано, що одним з перспективних напрямків у пост квантовій криптографії є криптографія в кільцях поліномів (на алгебраїчних решітках). Реалізації таких криптосистем вимагають ефективних алгоритмів множення поліномів. Запропоновані алгоритми враховують особливості структури тернарних поліномів у таких NTRU — подібних системах та забезпечують константність часу виконання операції множення. Для реалізації використовувалися кеш ефективні структури даних, AVX2 інструкції, розпаралелювання та спеціальні засоби для мінімізації та компенсування використання не вирівняної пам'яті. Критичний код написано на асемблері. Показано, що він працює вдвічі швидше за алгоритм, запропонований у [1] як для звичайної форми, так і для product-форми. Для звичайної форми відсоток прискорення — 38.6%, а для product-форми в 2 рази.

Список використаних джерел:

1. Daniel J. Bernstein, Chitchanok Chuengsatiansup, Tanja lange, and Christine van Vredendaal. NTRU Prime. URL: https://ntruprime.cr.yp.to/ntruprime-20160511.pdf.
2. Gorbenko I. D., Alekseychuk A. N., Kachko O. H., Yesina M. V., Bobukh V. A., Kandyi S. O., Ponomar V. A. Calculation of general parameters form NTRU Prime Ukraine of 6-7 levels of stability. *Radiotekhnika* : All-Ukr. Sci. Indep. Mag. Kharkiv : KNURE. 2019. № 195. P. 17–25.
3. Agner Fog. Optimizing subroutines in assembly language. URL: https://www.agner.org/optimize/ optimizing_assembly.pdf.

OPTIMIZATION OF THE MULTIPLY FUNCTION OF POLYNOMIALS FOR GENERAL AND PRODUCT FORMS OF THE REPRESENTATION OF ONE POLYNOMIAL

The research was carried out and the development of an effective practical algorithm for multiplying ternary polynomials in a ring $Z_3[x]\left(x^n - x - 1\right)$ was performed taking into account their structure. Variants for polynomials

with a normal structure with a fixed number of nonzero elements and in the PRODUCT-form in which the polynomial is the result of the calculation $F_1 * F_2 + F_3$, where $F_1, F_2, F_3 \in \mathrm{Z}_3[x]\left(x^n - x - 1\right)$ and have d_1, d_2, d_3 elements with values «1» and «–1» respectively, were considered. The results of optimization are given using vectorized instructions (AVX2 instructions), parallelization and special tools to minimize and compensate for the use of unbalanced memory. The critical code was written on the assembler under the microprocessor architecture x86-64, which is one of the most widely spread for today. Optimized version's time values for the parameter sets for 256, 384 and 512 bit classical of security were obtained and a comparison of the efficiency with the polynomial multiplication algorithm, which was proposed in an asymmetric post-quantum cryptosystem on algebraic NTRU Prime grids, was proposed. Testing was done on the Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz processor and on operating system Linux 4.15.0-44-generic #47-Ubuntu SMP x86-64. The results obtained in this paper are extremely relevant. They can be useful for cryptologists and other professionals involved in the development of new, effective cryptographic scheme and protocols for the post-quantum period, because for cryptosystem classes that use the modification in polynomial rings as basic operations, multiplication operation is operation that takes most of the time and requires significant optimization. The biggest advantage of the developed algorithm is the possibility of its parallelization on multiprocessor systems, which is a significant advantage over the algorithm presented in NTRU Prime.

Key words: *NTRU, multiplication of polynomials, ternary polynomial, PRODUCT-form, algebraic lattice, post quantum.*

Одержано 12.02.2019

УДК 519.8
DOI: 10.32626/2308-5878.2019-19.35-41

И. В. Козин, д-р физ.-мат. наук,
С. И. Полюга, канд. физ.-мат. наук,
В. И. Сардак, аспирант

Запорожский национальный университет, г. Запорожье

ФРАГМЕНТАРНАЯ МОДЕЛЬ РАЗМЕЩЕНИЯ ПРОИЗВОДСТВА

Рассмотрена двумерная задача размещения производственных объектов в дискретной постановке. Показано, что дискретная задача размещения производства сводится к задаче покрытия графа звездами и имеет фрагментарную структуру. Для поиска приближенного решения задачи предложены модификация эволюционного алгоритма на перестановках с геометрическим оператором кроссовера и алгоритм муравьиной колонии на фрагментарной структуре. Приводятся результаты численного эксперимента по сравнению алгоритмов.

Ключевые слова: *фрагментарная модель, задача размещения производства, эволюционный алгоритм, геометрический кроссовер, алгоритм муравьиной колонии.*

Введение. Задача размещения производственных объектов часто возникает в экономике, производстве, строительстве. В последнее время, учитывая децентрализацию управления страны, эта задача актуальна при выборе мест объектов коллективного пользования в территориальных образованиях. Существует много различных постановок задачи [1–3]. Несмотря на относительную простоту описания, задача размещения производства сложная практически в любой из постановок. Существует множество подходов к поиску оптимальных решений задачи [3]. Но для больших размерностей задача, как правило, решается приближенно с помощью метаэвристик различного вида. Большое количество дополнительных условий в конкретных постановках приводит к необходимости использовать вероятностные методы и эвристические процедуры.

Постановка задачи. Рассмотрим одну из наиболее простых дискретных постановок задачи размещения производства на эвклидовой плоскости. Имеется конечное множество точек плоскости, каждая из которых характеризуется двумя эвклидовыми координатами. Каждая из точек представляет собой или будущего потребителя продуктов производства, или возможную точку размещения производства. Все объекты в этой постановке считаются точечными. В качестве целевой функции задачи принимаются затраты на открытие производства в заданной точке либо за доставку продукции к определен-

ному клиенту. Все точки-потребители распределяются между точками производства таким образом, что:

а) каждый потребитель приписан одной и только одной точке производства;
б) распределение потребителей неизменно;
в) каждой точке приписана стоимость открытия производства в этой точке;
г) стоимость доставки продукции из точки производства к точке — потребителю прямо пропорциональна эвклидову расстоянию между точками.

Таким образом, каждое допустимое решение задачи может быть представлено на плоскости графом, который имеет вид объединения непересекающихся в вершинах звезд. Центром каждой звезды являются точки производства, а лучи связывают точки производства с точками — потребителями продукта (рисунок).

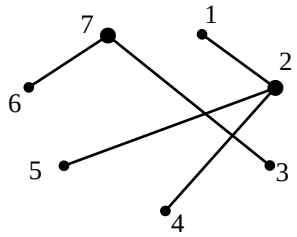

Рисунок. *Одно из допустимых решений задачи размещения производства*

Покажем, что рассматриваемая задача может рассматриваться как задача на фрагментарной структуре и, соответственно, к ней могут быть применены универсальные алгоритмы на фрагментарной структуре, в основе которых лежат метаэвристики [4].

Фрагментарная структура. В соответствии с [5] фрагментарной структурой (X, E) на конечном множестве X будем называть семейство его подмножеств $E = \{E_1, E_2, ..., E_n\}$ такое, что $\forall E_i \in E$, $E \neq \varnothing \quad \exists e \in E_i : E_i \setminus \{e\} \in E$.

Элементы из множества E будем называть допустимыми фрагментами. Таким образом, для любого допустимого фрагмента E_i существует нумерация его элементов $E_i = \{e_{i1}, e_{i2}, ...e_{is_i}\}$ такая, что $\forall k = 1, 2, ...s_i \quad \{e_{i1}, e_{i2}, ...e_{ik}\} \in E$. Элементарным фрагментом будем называть допустимый фрагмент, состоящий из одного элемента. Максимальный фрагмент — допустимый фрагмент, который не является подмножеством никакого другого фрагмента.

Максимальный фрагмент может быть построен с помощью следующего «жадного» алгоритма:

а) элементы множества X линейно упорядочиваются;

б) на начальном шаге выбирается пустое множество $X_0 = \varnothing$;

в) на шаге с номером $k+1$ выбирается первый по порядку элемент $x \in X \setminus X_k$, такой, что $X_k \cup \{x\} \in E$;

г) алгоритм заканчивает работу, если на очередном шаге не удалось найти элемент $x \in X \setminus X_k$ с требуемым свойством.

Результат работы алгоритма определяется заданным линейным порядком на множестве X. Таким образом, любой максимальный фрагмент может быть описан некоторой перестановкой элементов множества X. Пусть $A \in E$. Условие для элемента $x \in X$, при котором $A \cup \{x\} \in E$, будем называть условием присоединения элемента x.

Пусть теперь каждому фрагменту приписан вес, то есть задана функция $\rho : E \to R^1$. Будем предполагать, что функция ρ монотонна по включению (возрастающая или убывающая). Если $A, B \in E$ и $A \subseteq B$, то $\rho(A) \le (\ge) \rho(B)$. Задача оптимизации на фрагментарной структуре, это задача отыскания допустимого фрагмента максимального (минимального) веса. Очевидно, что для монотонных весов оптимальное решение будет являться максимальным фрагментом.

Фрагментарная модель. Покажем, что задача размещения производства в вышепредложенной постановке может быть представлена как задача оптимизации на фрагментарной структуре. В качестве множества элементарных фрагментов рассмотрим множество ребер полного графа с вершинами в заданных точках локации. Каждый допустимый фрагмент будем строить, соблюдая следующее условие присоединения. Очередное ребро присоединяется к выбранному набору ребер, если после присоединения полученный подграф представляет собой объединение непересекающихся в вершинах звезд. Если очередное ребро присоединить не удается, то переходим к следующему по порядку ребру. Алгоритм заканчивает работу, если список ребер исчерпан. Множества ребер, которые последовательно будут построены в результате работы такого алгоритма (множество E), образуют фрагментарную структуру. Центры полученных в результате работы алгоритма звезд и изолированные вершины (если в них локализованы получатели продукции) являются точками размещения производственных мощностей. Целевая функция задачи $F : E \to R^1$ — стоимость локализации организации производства в выбранных точках размещения производства плюс стоимость доставки продукции потребителям по лучам звезд. Очевидно, целевая функция является монотонной.

Любой максимальный фрагмент определяется заданным линейным порядком просмотра элементарных фрагментов. Этот порядок определяет результат работы фрагментарного алгоритма, который и построит требуемый максимальный фрагмент.

Каждый линейный порядок определяется некоторой перестановкой $s \in S_n$ укладываемых объектов (n — число ребер графа). Сопоставим каждой перестановке максимальный фрагмент, который ей порождается. Обозначим это отображение через $\varphi : S_n \to E$. Таким образом, имеет место естественная коммутативная диаграмма отображений

$$\begin{array}{ccc} & S_n & \\ \varphi \downarrow & & \searrow F \circ \varphi, \\ & E \to R^1 & \end{array}$$

которая превращает задачу оптимизации на фрагментарной структуре в задачу оптимизации на множестве перестановок. Причем любая перестановка является допустимой. Для больших значений n задача поиска оптимальной перестановки, как правило, является трудной. Предлагается использовать для поиска приближенных решений этой задачи варианты муравьиного и эволюционный алгоритмов на перестановках определенного вида [4].

Эволюционный алгоритм. Базовое множество X эволюционной модели — это множество $S_n = \{i_1, i_2 ..., i_n\}$ всех перестановок чисел $1, 2, \ldots, n$. Оператор построения начальной популяции выделяет произвольное подмножество заданной мощности Q из множества X.

Правило вычисления критерия селекции устроено следующим образом: по заданной перестановке фрагментов с помощью фрагментарного алгоритма строится максимальный допустимый фрагмент и вычисляется значение целевой функции задачи для этого фрагмента.

Опишем теперь оператор кроссовера. Пусть $U = (u_1, u_2, ..., u_n)$ и $V = (v_1, v_2, ..., v_n)$ — две произвольные перестановки. Перестановка-потомок строится следующим образом: последовательности U и V просматриваются в порядке следования элементов. На k-м шаге выбирается наименьший из первых элементов последовательностей и добавляется в новую перестановку-потомок. Затем этот элемент удаляется из двух последовательностей-родителей. Например, результатом кроссовера перестановок (2, 3, 6, 1, 7, 8, 4, 5) и (4, 6, 7, 1, 3, 2, 8, 5) будет перестановка (2, 3, 4, 6, 1, 2, 8, 5). В работе [5] показано, что определенный таким образом оператор кроссовера является геометрическим в инверсной метрике на перестановках [6]. Оператор мутации M выполняет случайную транспозицию в перестановке. Оператор селекции выбирает случайным образом набор пар из текущей популяции для последующего скрещивания.

Оператор эволюции упорядочивает элементы промежуточной популяции в последовательность по убыванию значения критерия селекции. В качестве новой текущей популяции выбираются первые Q элементов последовательности.

Обычное правило остановки — количество поколений достигло предельного значения. Лучшая по значению критерия селекции перестановка из последней построенной популяции определяет приближенное решение задачи.

Алгоритм муравьиной колонии [7]. Процедура вычисления оптимальной перестановки будет состоять из ряда циклов расчета. Каждый путь муравья между позициями 1, 2, ..., n будет определяться перестановкой $s = (i_1, i_2, ...i_n)$. Муравьи имеют собственную «память». У каждого муравья есть список уже посещенных позиций — список запретов. Обозначим $J_{i,k}^t$ список позиций, которые на цикле t необходимо посетить k-му муравью, находящемуся в позиции i.

Количество феромона в цикле с номером t при переходе из позиции i в позицию j определяется величиной $\tau_{ij}(t)$. На начальном этапе это количество можно задавать произвольно.

Вероятность перехода k-го муравья из позиции i в позицию j на цикле с номером t определяется следующим соотношением:

$$P_{ij,k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}}{\sum\limits_{l \in J_{i,k}^t} [\tau_{il}(t)]^{\alpha}}, & j \in J_{i,k}^t, \\ 0, & j \notin J_{i,k}^t, \end{cases}$$

где α — параметр, задающий вес следа феромона.

Количество откладываемого феромона составляет величину:

$$\Delta\tau_{ij,k}(t) = \begin{cases} \dfrac{Q}{L_k(t)}, (i, j) \in T_k(t), \\ 0, (i, j) \notin T_k(t), \end{cases}$$

где Q — положительный параметр, $L_k(t)$ — значение накрывающего отображения на перестановке, соответствующей маршруту k-го муравья на цикле с номером t. Изменение количества феромона определяется следующим выражением:

$$\tau_{ij}(t+1) = (1-\rho)\cdot\tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij,k}(t),$$

где m — количество муравьев, ρ — коэффициент «испарения» $(0 < \rho < 1)$.

Алгоритм прекращает работу, когда выполнено некоторое правило остановки, например, достигнута граница числа циклов. Минимальная по значению накрывающего отображения перестановка, найденная на последнем цикле, преобразуется в решение исходной задачи.

Результаты работы. Для проверки качества предлагаемых метаэвристик было сгенерировано 100 задач со случайным размещение точек на плоскости и со случайными оценками стоимости организации производства. Использовались три вида эвристик: локальный алгоритм со случайным выбором начальной точки, муравьиный и эволюционный алгоритмы на фрагментарных структурах. В каждом подходе выполнялось примерно одинаковое количество вычислений значения целевой функции.

Решения, полученные в результате применения различных алгоритмов, сравнивались по значению целевой функции.

Результаты работы алгоритмов сравнивались по числу первых мест и по рейтингу, который рассчитывался как сумма набранных алгоритмом очков. За первое место предлагалось три очка, за второе — два, за третье — одно очко (таблица).

Таблица

Результаты тестирования алгоритмов

Алгоритм	**Кол-во задач**	**Рейтинг**
Локальный	100	269
Эволюционный	100	210
Муравьиный	100	121

Выводы. Теоретические результаты и результаты численных экспериментов показали достаточно высокую эффективность эволюционного алгоритма и алгоритма муравьиной колонии при решении задачи размещения производства. Учитывая простоту реализации и возможность учета дополнительных ограничений, рассматриваемый в статье подход может быть предложен для практического решения задач размещения объектов коллективного пользования с различными ограничениями.

Список использованной литературы:

1. Khumawala B. M. An Efficient Branch-Bound Algorithm for the Warehouse Location Problem. *Management Science*. 1972. Vol. 18. P. 718–731.
2. Krarup J., Pruzan P.M. The simple plant location problem: Survey and synthesis. *European Journal of Operational Research*. 1983. Vol. 12. P. 36–81.
3. Береснев В. Л., Гимади Э. Х., Дементьев В. Т. Экстремальные задачи стандартизации. Новосибирск : Наука, 1978. 333 с.
4. Козин И. В., Перепелица В. А., Максишко Н. К. Фрагментарные структуры в задачах дискретной оптимизации. *Кибернетика и системный анализ.* 2017. № 6. С. 125–131.

5. Козин И. В. Фрагментарные структуры и эволюционные алгоритмы. *Питання прикладної математики і математичного моделювання.* Дніпропетровськ, 2008. С. 138–146.
6. Dorigo M. Optimization, Learning, and Natural Algorithms. PhD Thesis, Dipartimento di Elettronica, Politechnico Di Milano, Italy. 1992. 140 p.
7. Moraglio A., Poli R. Inbreeding Properties of Geometric Crossover and Non-geometric Recombinations. *Foundations of Genetic Algorithms.* 2007. P. 1–14.

FRAGMENTAL MODEL PLACEMENT PRODUCTION

A two-dimensional problem of locating production objects in a discrete formulation is considered. It is shown that the discrete problem of locating production reduces to the problem of covering a graph with stars and has a fragmentary structure. To search for an approximate solution of the problem, a modification of the evolutionary algorithm on permutations with a geometric crossover operator and an ant colony algorithm on a fragmentary structure are proposed. The results of numerical experiment comparison of algorithms are given.

Key words: *fragmented model, task of locating production, evolutionary algorithm, geometric crossover, ant colony algorithm.*

Получено 29.01.2019

УДК 519.8
DOI: 10.32626/2308-5878.2019-19.41-46

О. М. Коломис, канд. фіз.-мат. наук
Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ОЦІНКА ПОХИБКИ ЗАОКРУГЛЕННЯ АЛГОРИТМУ ОБЧИСЛЕННЯ ОЦІНКИ СПЕКТРАЛЬНОЇ ЩІЛЬНОСТІ

У роботі розглянуто ефективні за швидкодією алгоритми обчислення оцінок спектральних щільностей стаціонарних ергодичних випадкових процесів із нульовим середнім значенням. Найчастіше для їх обчислення використовують метод прямого перетворення Фур'є з використанням алгоритму швидкого перетворення Фур'є (ШПФ). Стаття продовжує дослідження і обґрунтування цього методу в напрямку отримання більш якісних оцінок похибок заокруглення. Наведена оцінка похибки заокруглення алгоритму обчислення оцінки спектральної щільності.

Ключові слова: *оцінка спектральної щільності, похибка заокруглення, швидке перетворення Фур'є.*

Вступ. Швидкі алгоритми розв'язання задач спектрального і кореляційного аналізу випадкових процесів почали з'являтися, в основному, після 1965 року, коли в обчислювальну практику увійшов алгоритм ШПФ [1, 2]. З його появою розроблено ряд обчислювальних

алгоритмів прискореного розв'язання деяких задач цифрової обробки сигналів, побудовані ефективні за швидкодією алгоритми обчислення таких оцінок імовірнісних характеристик об'єктів керування, як оцінок згорток, кореляційних функцій, спектральних щільностей стаціонарних і деяких типів нестаціонарних випадкових процесів [2, 3].

Постановка задачі та алгоритм розв'язання. Нехай $x(t)$ — випадковий стаціонарний ергодичний процес з нульовим середнім значенням і задана вибірка $x_\nu = x(t_\nu)$, $\nu = \overline{0, N-1}$. Для отримання оцінки спектральної щільності використовуємо співвідношення [2]

$$S_x(k) = S_x(\omega_k) = \frac{h}{N}\left|\hat{X}_k\right|^2, \; k = \overline{0, N-1}, \tag{1}$$

де h — крок часу, $\hat{X}_k = \hat{X}(\omega_k)$ — дискретне перетворення Фур'є (ДПФ) початкового сигналу $x(t)$, $\omega_k = k/(Nh)$, $k = \overline{0, N-1}$. Для обчислення $\hat{X}_k$, $k = \overline{0, N-1}$ будемо використовувати алгоритм ШПФ

$$\hat{X}_k = \hat{X}(\omega_k) = \sum_{\nu=0}^{N-1} x_\nu W_N^{\nu k}, \tag{2}$$

де $k, \nu = \overline{0, N-1}$, $W_N = e^{-i\frac{2\pi}{N}}$.

Відомо [2], що евклідова норма оцінки похибки заокруглення алгоритму ШПФ обчислення ДПФ $\hat{X} = \left\{\hat{X}_k\right\}_0^{N-1}$ сигналу $x = \left\{x_\nu\right\}_0^{N-1}$, для $N = 2^\gamma$, $\gamma > 0$ — ціле, і класичному правилу заокруглення має вигляд

$$\left\|E_{\hat{X}}\right\|_E < 8 \cdot 1,06 \cdot \gamma \cdot 2^{-\tau} \cdot \left\|\hat{X}\right\|_E. \tag{3}$$

Величину $S_x(k)$, $k = \overline{0, N-1}$, що визначається у вигляді співвідношення (1), називають первинною оцінкою спектральної щільності. Ця оцінка є досить «грубою», оскільки відбувається велике «просочування енергії» через бокові пелюстки за рахунок розширення головного пелюстка. Зазвичай прагнуть локалізувати енергію на центральній частоті, зменшуючи її «витік» у бокові пелюстки. Цього можна досягти, помноживши часову послідовності x_ν на деяке вікно даних (вагові послідовності) $d_\nu = d(\nu)$, $\nu = \overline{0, N-1}$. Тоді отримаємо [2]:

$$x_{d,\nu} = x_d(\nu) = x(\nu) \cdot d(\nu) = x_\nu \cdot d_\nu, \; \nu = \overline{0, N-1}. \tag{4}$$

Спектральна щільність визначається виразом

$$S_x^*(k) = S_x^*(\omega_k) = S_{x_d}(k) = \frac{h}{N}\left|\hat{X}_{d,k}\right|^2, \tag{5}$$

де

$$\hat{X}_{d,k} = \hat{X}_d(k) = \sum_{\nu=0}^{N-1} x_{d,\nu} W_N^{\nu k} = \sum_{\nu=0}^{N-1} x_\nu \cdot d_\nu \cdot W_N^{\nu k}, \qquad (6)$$

$\hat{X}_{d,k}$ — ДПФ сигналу $x_{d,\nu}$, $\nu = \overline{0, N-1}$.

Вікна даних зменшують відхилення амплітуди і дисперсію сигналу. Це призводить до погіршення оцінки спектра. Для отримання асимптотично незміщеної первинної оцінки спектральної щільності застосовується ваговий множник Q, який дає спектральне вікно одиничної площі. Отже, якщо застосовується згладжування даних, то оцінка спектральної щільності визначається виразом

$$\hat{S}_x^*(k) = \hat{S}_x^*(\omega_k) = S_x^*(k) \big/ Q, \qquad (7)$$

де $Q = \dfrac{1}{N} \sum_{\nu=0}^{N-1} d^2(\nu)$.

Таким чином, зміна спектра, викликана вживанням вікон даних компенсується введенням множника $1/Q$.

Для отримання остаточних згладжених оцінок спектральної щільності $\hat{S}_x(k) = \hat{S}_x(\omega_k)$, $k = \overline{0, N-1}$, які мають кращі статистичні властивості і придатні для практичного використання, слід провести подальше згладжування в частотній області, використовуючи згладжуючі функції [2, 3]

$$\hat{S}_x(k) = \hat{S}_x(\omega_k) = P\left(\hat{S}_x^*(k), j\right), \qquad (8)$$

де $P\left(\hat{S}_x^*(k), j\right)$ — функція, що здійснює згладжування первинної спектральної щільності $\hat{S}_x^*(k)$ в частотній області, j — деякий параметр.

У роботах [2, 3] наведені сімейства вікон даних і вікон частот, які найчастіше застосовуються на практиці. Наприклад, наведені чотири сімейства вікон даних: алгебраїчне, узагальнене косинусоїдальне, гауссове і подібне до кореляційних вікон Хеммінга. Як згладжуючі функції $P = P\left(\hat{S}_x^*(k), j\right)$ в частотній області застосовують наступні: просте згладжування, згладжування з трикутною вагою, згладжування з косинусоїдальним вікном, усереднення за відрізками реалізації та комбінований спосіб усереднення за частотами та за відрізками.

Розглянемо алгоритм, який дозволяє (за ознакою) обчислювати або $S_x(k)$, або $S_x^*(k)$, або $\hat{S}_x^*(k)$, або $\hat{S}_x(k)$, $k = \overline{0, N-1}$, причому в ньому можуть бути використані вікна як для даних, так і для частот [2, 4]. Основні кроки алгоритму.

Крок 1. «Набивка» L нулями вихідної послідовності x_ν, $\nu = \overline{0, N_1 - 1}$, так щоб $N_1 = N + L = 2^\gamma$, $\gamma > 0$ — ціле, у випадку, якщо $N \neq 2^\gamma$.

Крок 2. Обчислення (за ознакою) ДПФ $\hat{X}_k$, $k = \overline{0, N_1 - 1}$ отриманої послідовності x_ν, $\nu = \overline{0, N_1 - 1}$, згідно співвідношення (2) з використанням алгоритму ШПФ.

Крок 3. Обчислення $S_x(k)$, $k = \overline{0, N_1 - 1}$ згідно співвідношення (1). Перехід на крок 8.

Крок 4. Згладжування (за ознакою) отриманої на кроці 1 послідовності x_ν, за допомогою вікон даних d_ν, $\nu = \overline{0, N_1 - 1}$ (див. співвідношення (3)) та отримання послідовності $x_{d,\nu} = x_\nu \cdot d_\nu$, $\nu = \overline{0, N_1 - 1}$.

Крок 5. Обчислення ДПФ $\hat{X}_{d,k}$ послідовності $x_{d,\nu}$, $\nu, k = \overline{0, N_1 - 1}$ з використанням алгоритму ШПФ та отримання $S_x^*(k)$, $k = \overline{0, N_1 - 1}$ згідно співвідношення (4).

Крок 6. Обчислення оцінки спектральної щільності $\hat{S}_X^*(\omega_k)$ згідно співвідношення (6), $k = \overline{0, N_1 - 1}$. Перехід за ознакою на крок 7 або 8.

Крок 7. Згладжування $\hat{S}_x^*(k)$, $k = \overline{0, N_1 - 1}$ (за ознакою) за допомогою вікон частот та отримання згладженої оцінки спектральної щільності $\hat{S}_x(k)$ згідно співвідношення (7), $k = \overline{0, N_1 - 1}$.

Крок 8. Кінець.

Основними характеристики наведеного алгоритму обчислення оцінок спектральної щільності є точність та обчислювальна складність. В роботі основна увага приділена аналізу точності, тобто, отриманню оцінок похибок, що супроводжують процес обчислення оцінки спектральної щільності. Для оцінки точності запропонованого алгоритму обчислення спектральних щільностей дослідимо лише оцінку похибки заокруглення E_3, що виникає при реалізації обчислювального алгоритму на комп'ютері для класичного правила заокруглення, для обчислень у режимі плаваючої коми з τ розрядами в мантисі числа.

Оцінка похибки заокруглення алгоритму обчислення $S_x(k)$, $k = \overline{0, N-1}$. Нехай $N = 2^\gamma$, $\gamma > 0$ — ціле, $x_\nu = x(t_\nu)$, $\nu = \overline{0, N-1}$ — вибірка стаціонарного ергодичного випадкового процесу $x(t)$ з нульовим середнім значенням, $fl(*)$ — результат обчислення виразу,

який стоїть у дужках на ЕОМ у режимі з плаваючою комою з τ розрядами у мантисі числа, $\|*\|_E$ — евклідова норма вектора. Справедлива така теорема [5].

Теорема 1. Оцінка евклідової норми похибки заокруглення обчислення $S_x(k)$, $k=\overline{0,N-1}$, згідно виразу (1) за допомогою алгоритму ШПФ для режиму з плаваючою комою з τ розрядами у мантисі числа має вигляд

$$\left\| E_{З,S_x} \right\|_E \sim 8\cdot 1,06\cdot\gamma\cdot 2^{-\tau}\cdot h\cdot \| x \|_E^2 \left[1+1,06\cdot 2^{-\tau}(16\gamma+\sqrt{N})\right]. \quad (9)$$

Оцінка похибки заокруглення алгоритму обчислення $\hat{S}_x^*(k)$, $k=\overline{0,N-1}$**.** Нехай $\varepsilon_{d,\nu}=\varepsilon_d(\nu)$, $e_{x_{d,\nu}}=e_{x_d}(\nu)$ — похибки заокруглення, які виникають при обчисленні відповідно d_ν та $x_{d,\nu}$ за допомогою співвідношення (4) в режимі з плаваючою комою з τ розрядами у мантисі числа, $\hat{X}_{d,k}$ та $\hat{d}_k$ — ДПФ відповідно послідовностей $x_{d,\nu}$ та d_ν, $\nu=\overline{0,N-1}$, які обчислюються за допомогою ШПФ з похибками заокруглення відповідно $E_{\hat{X}_{d,k}}=E_{\hat{X}_d}(k)$ та $E_{\hat{d},k}=E_{\hat{d}}(k)$, E_Q — похибка заокруглення обчислення масштабуючого множника Q. Справедлива наступна теорема [6].

Теорема 2. Оцінка евклідової норми похибки заокруглення обчислення оцінки спектральної щільності $\hat{S}_x^*(k)$, $k=\overline{0,N-1}$, згідно співвідношень (4)–(7) за допомогою алгоритму ШПФ, для режиму з плаваючою комою з τ розрядами у мантисі числа має вигляд

$$\left\| E_{З,\hat{S}_x^*} \right\|_E \sim \frac{h}{Q} 8\cdot 1,06\cdot\gamma\cdot 2^{-\tau}\|x\|_E^2\|d\|_E\left\{\|d\|_E\left(1+\left\|E_Q\right\|_E+1,06\cdot 2^{-\tau+2}\times\right.\right.$$
$$\left.\left.\times\left[\sqrt{N}+4\gamma\right]\right)+2\sqrt{N}\left\|\varepsilon_d\right\|_E(8\cdot 1,06\cdot 2^{-\tau}\gamma+1)\right\}. \quad (10)$$

Із більш детального аналізу алгоритму ШПФ випливає, що виграш у кількості арифметичних операцій у порівнянні зі стандартним способом виходить ще більшим, оскільки багато множників вигляду W_N^{kr} мають в «метеликах» значення ± 1, $\pm i$, завдяки чому виключаються відповідні операції множення.

Доведено, що при $N=2^\gamma$ оцінка знизу (серед всіх алгоритмів обчислення ДПФ) кількості операцій додавання рівна $\frac{N}{2}\log_2 N$. Існують також алгоритми і програми обчислення багатовимірного ДПФ за допомогою ШПФ.

Список використаних джерел:

1. Cooley J. W., Tukey J. W. An algorithm for the machine calculation of complex Fourier Series. *Math. Comput.,* 1965, Apr. P. 257–301.
2. Задирака В. К. Теория вычисления преобразования Фурье. Киев : Наук. думка, 1983. 216 с.
3. Сергієнко І. В., Задірака В. К., Литвин О. М., Мельникова С. С., Нечуйвітер О. П. Оптимальні алгоритми обчислення інтегралів від швидкоосцилюючих функцій та їх застосування. Т. 2. Застосування. Київ : Наук. думка, 2011. 348 с.
4. Коломис О. М., Луц Л. В. Алгоритм обчислення оцінок спектральної щільності. *Питання оптимізації обчислень* (*ПОО-XLII*)*: праці міжнар. наук. школи-семінару, присвяченої 85-річчю від дня народження академіка В.С. Михалевича* (21–25 вересня 2015 р.). Київ : Інститут кібернетики імені В.М. Глушкова НАН України, 2015. С. 45–46.
5. Коломис О. М. Оцінка похибки заокруглення алгоритму обчислення первинної оцінки спектральної щільності. *Математичне та комп'ютерне моделювання.* Серія: Фізико-математичні науки. 2017. Вип. 15. С. 80–84.
6. Коломис О. М. Ефективні за точністю та швидкодією алгоритми визначення оцінок динамічних та імовірнісних характеристик неперервних виробничих процесів. Дисертація на здобуття наукового ступеня канд. фіз.-мат. наук. Київ, 2015. 172 с.

ESTIMATION OF THE ROUNDING ERROR OF THE ALGORITHM FOR CALCULATING THE ESTIMATE OF THE SPECTRAL DENSITY

The estimation of the rounding error of the algorithm for calculating the estimate of the spectral density are developed.

Key words: *the estimate of the spectral density, rounding error, fast Fourier Transform.*

Одержано 14.02.2019

УДК 517.958:532.5;517.957;681.513.8;001.891.57:53
DOI: 10.32626/2308-5878.2019-19.47-53

А. М. Крот, д-р техн. наук, профессор

Объединенный институт проблем информатики
НАН Беларуси, г. Минск, Республика Беларусь

МОДЕЛЬ ЭВОЛЮЦИИ ХАОТИЧЕСКИХ ВОЛНОВЫХ ПРОЦЕССОВ В СЛОЖНЫХ ДИНАМИЧЕСКИХ СИСТЕМАХ НА ОСНОВЕ ТЕОРИИ МАТРИЧНОЙ ДЕКОМПОЗИЦИИ

В работе разработана общая модель возникновения и эволюции хаотических волновых процессов в сложных системах на основе предложенного метода матричной декомпозиции операторов нелинейных систем. Предложенная модель показала, что эффект самоорганизации в сложных системах различной физической природы (на примерах гидродинамической, электронной и космогонической систем) заключается во взаимодействии нелинейных процессов высших порядков, приводящей к стабилизации (к конечной величине) амплитуды хаотического волнового процесса. Математически это выражается в синхронном «противодействии» нелинейных процессов чётных и нечётных порядков в общей векторно-матричной модели сложной системы, находящейся в хаотическом режиме. Реализация векторно-матричной декомпозиции посредством вычислительных экспериментов показала, что модель Л. Д. Ландау достаточно хорошо описывает сценарий возникновения хаотических режимов в сложных системах. Отмечено, что режим жесткого самовозбуждения нелинейных колебаний в сложных системах приводит к появлению хаотического аттрактора в пространстве состояний. Вместе с тем предложенная векторно-матричная модель позволила найти более общие условия возникновения и эволюции хаотических волновых процессов и, как следствие, объяснить возникновение согласованных нелинейных явлений в сложных системах.

Ключевые слова: *сложная динамическая система, пространство состояний, хаотический аттрактор, матричный ряд в пространстве состояний, общая векторно-матричная модель хаотических волновых процессов, режим жесткого самовозбуждения нелинейных колебаний, стабилизация амплитуды хаотического процесса.*

Введение. Развитие теории хаотических волновых процессов (в частности, теории турбулентности в аэрогидродинамических потоках) важно с точки зрения понимания процессов самоорганизации в сложных динамических системах. Л. Д. Ландау в своей статье [1] разработал теорию *начальной турбулентности*, в рамках которой показал, что

первоначальная неустойчивость нестационарного движения не растет неограниченно, а стремится к некоторому конечному пределу. В работе [2] предложена модель дискретной *квазистационарной* линейной динамической системы на основе обобщенного спектрального представления в базисе собственных функций, соответствующих собственным значениям оператора этой системы. Э. Лоренц [3], исследуя динамическое поведение вязкой жидкости в условиях конвенкции (течение Рэлея–Бенара), предложил модель турбулентности, для построения которой использовался метод Галёркина с целью редуцирования системы уравнений Навье–Стокса и теплопроводности. В результате *редуцированная модель Лоренца*, описываемая тремя обыкновенными нелинейными дифференциальными уравнениями, позволила выявить хаотическое поведение системы, приведшее к открытию *хаотического* (странного) *аттрактора* в пространстве состояний. Математически понятие «хаотического аттрактора» сформулировано Д. Рюэлем и Ф. Такенсом [4] как ключевой элемент в интерпретации иррегулярного поведения, описываемого детерминистскими уравнениями для понимания главным образом турбулентности. Тем самым было положено начало исследованиям того, что теперь именуется *детерминированным хаосом* [5, 6]. Несмотря на достигнутые успехи, вместе с тем остаются не до конца выясненными вопросы, касающиеся *стабилизации* хаотических волновых процессов, позволяющей достаточно долго поддерживать незатухающие хаотические колебания в сложных системах при неизменности их *управляющих параметров.*

Построение общей модели возникновения хаотических волновых процессов с использованием метода матричной декомпозиции. Известно [5, 6], что хаотические волновые процессы возникают в сложных системах самой различной физической природы (например, в планетарных, электродинамических, химических и физиологических системах). В этой связи попытаемся построить общую модель возникновения и стабилизации хаотических волновых процессов с использованием теории матричной декомпозиции в пространстве состояний сложной системы [7–10]. В векторно-матричном виде система обыкновенных дифференциальных уравнений может рассматриваться как задача Коши в N - мерном пространстве состояний U сложной НДС:

$$\dot{\vec{u}} = \vec{f}\left(\vec{u}(t),\ \vec{u}_0, \{c_l\}\right),\ \vec{u}(0) = \vec{u}_0,\ \vec{u}(t) \in U, \qquad (1)$$

где $\vec{u}(t) = (u_1(t), \ldots, u_N(t))^T$, T — символ транспонирования, $\vec{u}_0$ — вектор начальных данных, $\{c_l\}$ — множество параметров системы. Решение $\vec{u}(t)$ уравнения (1) задаёт некоторую кривую в пространстве состояний (фазовом пространстве) $U = \Re^N$, называемую *фазовой траекторией*. Для исследования поведения решения уравнения (1)

вблизи конкретного стандартного состояния $\vec{u}^*$ рассматриваем невозмущённое решение (1), постоянно возмущаемое внешними воздействиями (или внутренними флуктуациями) на величину $\vec{v} = \vec{v}(t)$ [5]. В результате вместо $\vec{u}^*$ возникает новое решение

$$\vec{u} = \vec{u}^* + \vec{v}(t) \,. \tag{2}$$

С учетом (2) запишем систему (1) относительно $\vec{v}(t)$:

$$\dot{\vec{v}} = \Delta\vec{f}\,(\vec{v}(t),\ \vec{u}^*, \{c_l\}) \,, \tag{3}$$

где $\vec{v}(t) = (v_1(t), \ldots, v_N(t))^T$, $\Delta\vec{f}$ — приращение векторной функции, $\vec{u}^*$ — вектор невозмущённого (стандартного) состояния, $\{c_l\}$ — набор параметров системы. Согласно теории матричной декомпозиции, приращение векторной функции $\Delta\vec{f}$ сложной НДС в пространстве состояний описывается матричным рядом вида [7–10]:

$$\begin{aligned} \Delta\vec{f}\,(\vec{v},\vec{u}^*) = \vec{f}\,(\vec{u}^* + \vec{v}) - \vec{f}\,(\vec{u}^*) = L^{(1)}_{N\times N}\vec{v} + \frac{1}{2!}L^{(2)}_{N\times N^2}(\vec{v}\otimes\vec{v}) + \\ + \frac{1}{3!}L^{(3)}_{N\times N^3}(\vec{v}\otimes\vec{v}\otimes\vec{v}) + \ldots = \sum_{k=1}^{\infty}\frac{1}{k!}L^{(k)}_{N\times N^k}\cdot\vec{v}^{\otimes k}, \end{aligned} \tag{4}$$

где $L^{(k)}_{N\times N^k} = \underbrace{(\frac{\partial}{\partial\vec{v}^T}\otimes(\frac{\partial}{\partial\vec{v}^T}\otimes\ldots\otimes(\frac{\partial}{\partial\vec{v}^T}}_{k}\otimes\vec{f}\,)\ldots))_{\vec{u}^*}$ — матричные ядра однородных нелинейных операторов системы, $\vec{v}^{\otimes k} = (\underbrace{\vec{v}\otimes\vec{v}\otimes\ldots\otimes\vec{v}}_{k})$ — k-я кронекеровская степень вектора возмущений $\vec{v}$.

Применяя матричное разложение (4) к правой части уравнения (3), получаем:

$$\dot{\vec{v}} = L^{(1)}_{N\times N}\vec{v} + \frac{1}{2!}L^{(2)}_{N\times N^2}(\vec{v}\otimes\vec{v}) + \frac{1}{3!}L^{(3)}_{N\times N^3}(\vec{v}\otimes\vec{v}\otimes\vec{v}) + \ldots \tag{5}$$

Нетрудно видеть, что полученное уравнение (5) обобщает как модель Ландау начальной турбулентности [1], так и модель конвекционной турбулентности Лоренца [3], поэтому его можно рассматривать в качестве общей модели возникновения и эволюции хаотических волновых процессов в сложных НДС.

Реализация общей модели возникновения и амплитудной стабилизации незатухающих хаотических волновых процессов. Проведенные вычислительные эксперименты с использованием общей модели (5) возникновения хаотических волновых процессов для конкретных типов сложных НДС указывают на тот факт, что конечные незатухающие хаотические колебания наблюдаются лишь при определённых соотношениях между вкладами линейного $L^{(1)}_{N\times N}$, квадратичного $L^{(2)}_{N\times N^2}$,

кубического $L^{(3)}_{N\times N^3}$ и т. д. ядер матричного ряда в общую динамику сложной системы. В случае электронной схемы Чжуа модель (5) хаотических волновых процессов принимает следующий вид:

$$\dot{\vec{v}} = L^{(1)}_{3\times 3}\vec{v} + \frac{1}{2!}L^{(2)}_{3\times 9}(\vec{v}\otimes\vec{v}) + \frac{1}{3!}L^{(3)}_{3\times 27}(\vec{v}\otimes\vec{v}\otimes\vec{v}), \qquad (6)$$

поскольку динамика сложной НДС Чжуа точно описывается на основе только линейного, квадратичного и кубического ядер:

$$L^{(1)}_{3\times 3}(\vec{u}^*) = \left[\begin{array}{c|c|c} -(3A\alpha u_1^{*2} + C\alpha) & \alpha & 0 \\ \hline 1 & -1 & 1 \\ \hline 0 & -\beta & 0 \end{array}\right];$$

$$L^{(2)}_{3\times 9}(\vec{u}^*) = \left[\begin{array}{c|c|c|c|c|c|c|c|c} -6A\alpha u_1^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right];$$

$$L^{(3)}_{3\times 27}(\vec{u}^*) = \left[\begin{array}{c|c} -6A\alpha & 000000000000000000000000000 \\ \hline 0 & 000000000000000000000000000 \\ \hline 0 & 000000000000000000000000000 \end{array}\right].$$

Более того, вычислительное моделирование сигналов на выходах кубического и квадратичного ядер в хаотических режимах работы схемы Чжуа при выборе параметров системы, равных $\alpha = 15,6$, $\beta = 28$, $A = 0,002$, $C = -1,3$, $u_1^* = -1,5$, выявило подобие их эволюции во времени, но с противоположным знаком и неодинаковой амплитудой (рис. 1).

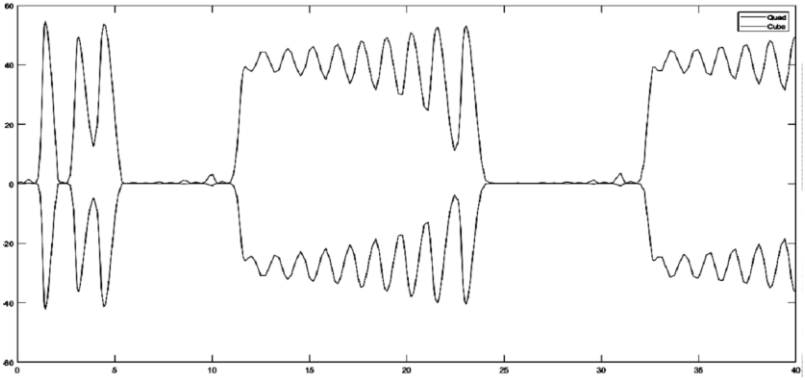

Рис. 1. *Вид сигналов, порождёнными квадратичным* $L^{(2)}_{N\times N^2}$ *и кубическим* $L^{(3)}_{N\times N^3}$ *ядрами в общей модели* (5) *возникновения хаотических волновых процессов в сложной НДС типа схемы Чжуа*

Другими словами, когда электронная схема Чжуа функционирует в хаотическом режиме, выходные сигналы от кубического и квадратичного ядер, находясь в противофазе, частично компенсируют друг друга, что в целом приводит к *стабилизации амплитуды хаотического волнового процесса* к конечной величине. В этом проявляется *эффект самоорганизации процессов* в сложной НДС типа схемы Чжуа, заключающийся во взаимодействии нелинейностей 2-го и 3-го порядков с последующей их синхронизацией.

При таких условиях наблюдается скачкообразный переход от стационарного режима сложной НДС к нестационарному, сопровождающийся возникновением двух частот ω_1 и ω_2, определяющих два цикла в пространстве состояний схемы Чжуа. Полученный результат находит свое объяснение и с точки зрения теории Рюэля–Такенса [4, 6]. При появлении дополнительной частоты ω_3 незначительные возмущения могут разрушить регулярные циклы, образующие тор T^3, и преобразовать его в хаотический аттрактор [5] (например, типа «двойной завиток» в пространстве состояний схемы Чжуа), что и было подтверждено вычислительными экспериментами, результат которых показан на рис. 2.

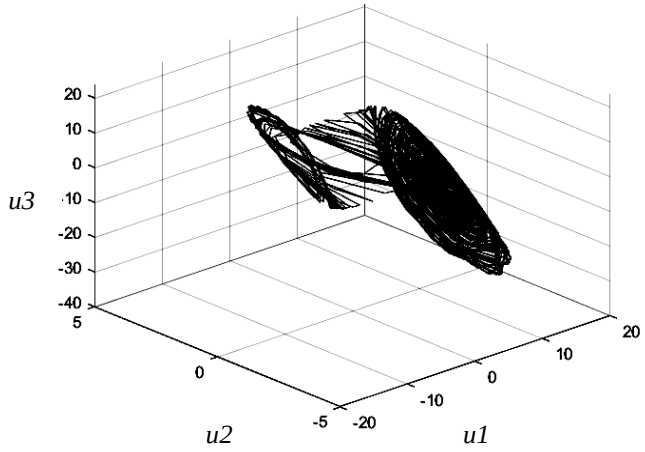

Рис. 2. *Вид хаотического аттрактора типа «двойной завиток» в пространстве состояний схемы Чжуа*

Модель возникновения хаотических волновых процессов в сложной НДС Фитц–Хью [9] имеет вид

$$\dot{\vec{v}} = L^{(1)}_{2\times 2}\vec{v} + \frac{1}{2!}L^{(2)}_{2\times 4}(\vec{v}\otimes\vec{v}) + \frac{1}{3!}L^{(3)}_{2\times 8}(\vec{v}\otimes\vec{v}\otimes\vec{v}), \qquad (7)$$

т. е. аналогично системе Чжуа динамика сложной НДС Фитц–Хью также описывается на основе лишь линейного, квадратичного и кубического ядер [8, 9]:

$$L_{2x2}^{(1)}(\vec{u}^*) = \left[\begin{array}{c|c} c - c\,u_1^{*2} & c \\ \hline -\dfrac{1}{c} & \dfrac{b}{c} \end{array}\right],\ L_{2x4}^{(2)}(\vec{u}^*) = \left[\begin{array}{c|c|c|c} -2cu_1^* & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right],$$

$$L_{2x8}^{(3)}(\vec{u}^*) = \left[\begin{array}{c|c|c|c|c|c|c|c} -2c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

В этом случае наблюдается переход от стационарного режима сложной НДС к нестационарному с возникновением сначала частоты ω_1, а затем второй частоты ω_2, в последующем определяющих два цикла хаотического аттрактора в пространстве состояний системы Фитц–Хью.

Выводы. Предложенная модель показала, что эффект самоорганизации в сложной НДС различной физической природы (на примерах гидродинамической, электронной и физиологической систем) заключается во взаимодействии нелинейных процессов высших порядков, приводящей к стабилизации (к конечной величине) амплитуды хаотического волнового процесса.

Благодарность. Работа выполнена в рамках предоставленного гранта Президента Республики Беларусь в науке на 2019 год.

Acknowledgements. This work has been supported by the grant of President of Republic of Belarus in science (2019).

Список использованных источников:

1. Ландау Л. Д. К проблеме турбулентности. *Доклады АН СССР.* 1944. Т. 44, № 8. С. 339–342.
2. Крот А. М. О классе дискретных квазистационарных линейных динамических систем. *Доклады АН СССР*. 1990. Т. 313, № 6. С. 1376–1380.
3. Lorenz E.N. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences.* 1963. Vol.20, March. P. 130–141.
4. Ruelle D. On the nature of turbulence. *Communications in Mathematical Physics*. 1971. Vol. 20. P. 167–192.
5. Николис Г., Пригожин И. Самоорганизация в неравновесных системах: От диссипативных структур к упорядоченности через флуктуации. М. : Мир, 1979. 512 с.
6. Берже П. Порядок в хаосе: о детерминистском подходе к турбулентности. М. : Мир, 1991. 368 с.
7. Krot A. M. The decomposition of vector functions in vector-matrix series into state-space of nonlinear dynamic system. *EUSIPCO–2000: Proc. X European Signal Processing Conf.*, Tampere, Finland, Sep. 4–8, 2000. Tampere, 2000. Vol. 3. P. 2453–2456.
8. Krot A. M. Matrix decompositions of vector functions and shift operators on the trajectories of a nonlinear dynamical system. *Nonlinear Phenomena in Complex Systems*. 2001. Vol. 4, № 2. P. 106–115.

9. Крот А. М. Анализ аттракторов сложных нелинейных динамических систем на основе матричных рядов в пространстве состояний. *Информатика*. 2004. № 1 (1). С. 7–16.
10. Krot A.M. The development of matrix decomposition theory for nonlinear analysis of chaotic attractors of complex systems and signals. *DSP–2009: Proc. 16th IEEE Intern. Conf. on Digital Signal Processing*, Thira, Santorini, Greece, July 5–7, 2009. Santorini, 2009. P. 1–5. https://doi.org/10.1109/icdsp.2009.5201123

A MODEL OF EVOLUTION OF CHAOTIC WAVE PROCESSES IN COMPLEX DYNAMICAL SYSTEMS ON THE BASIS OF THE MATRIX DECOMPOSITION THEORY

A general model of the origin and evolution of chaotic wave processes in complex systems based on the proposed method of matrix decomposition of operators of nonlinear systems is developed in the article. The proposed model shows that the effect of self-organization in complex systems of different physical nature is based on the interaction of nonlinear processes of higher orders leading to stabilization (to the finite value) of the amplitude of chaotic wave process. Mathematically, this means the synchronous «counteraction» of nonlinear processes of even and odd orders in a general vector-matrix model of a complex system being in a chaotic mode. The implementation of the vector-matrix decomposition by means of computational experiments shows that the model of L. D. Landau describes the scenario of the occurrence of chaotic modes in complex systems quite well. It is noted that the regime of hard self-excitation of nonlinear oscillations in complex systems leads to the appearance of a chaotic attractor in the state-space. Moreover, the proposed vector-matrix model permits to find more general conditions for the origin and evolution of chaotic wave processes and, as a result, to explain the appearance of coherent nonlinear phenomena in complex systems.

Key words: *complex nonlinear dynamical system, state-space, chaotic attractor, matrix series in state-space, general vector-matrix model of chaotic wave processes, mode of hard self-excitation of nonlinear oscillations, stabilization of the amplitude of chaotic process.*

Получено 21.01.2019

УДК 519.853
DOI: 10.32626/2308-5878.2019-19.54-60

Ю. П. Лаптин, д-р физ.-мат. наук,
Т. А. Бардадым, канд. физ.-мат. наук
Институт кибернетики имени В. М. Глушкова НАН Украины, г. Киев

О ПРИБЛИЖЕННОМ ВЫЧИСЛЕНИИ КОЭФФИЦИЕНТОВ ТОЧНЫХ ШТРАФНЫХ ФУНКЦИЙ

Предложены упрощенные процедуры уточнения штрафных коэффициентов. Приводятся результаты вычислительных экспериментов на случайно генерируемых задачах линейного программирования.

Ключевые слова: *точные штрафные функции, структурированные задачи оптимизации, методы декомпозиции.*

Введение. Точные штрафные функции давно уже стали общепринятым инструментом решения задач оптимизации. Однако при решении задач большой размерности с использованием схем декомпозиции допустимым подходом оказывается приближенное оценивание штрафных коэффициентов. Для построения упрощенных процедур уточнения штрафных коэффициентов предлагаются вспомогательные задачи. Данный подход развивает результаты, полученные в [1–3].

Рассматриваются задачи, представимые в виде: найти

$$f_0^* = \min\{ f_0(x) :\ x \in C,\ x \in M \}, \qquad (1)$$

где $C = \{x : f_i(x) \le 0,\ i = 1,...,m,\ x \in R^n\}$, $f_i : R^n \to R$, $i = 0,...,m$ — выпуклые функции, M — некоторое (простое) выпуклое множество, $M \subseteq R^n$. В качестве множества M обычно используется положительный ортант пространства R^n, другие множества простой структуры. Положим

$$\Phi_\beta(x) = f_0(x) + \sum_{i=1}^{m} \beta_i f_i^+(x),\ F_\lambda(x) = f_0(x) + \lambda \cdot h^+(x),$$

$$\lambda, \beta_i \ge 0,\ i = 1,..., m,$$

где $f^+(x) = \max\{0, f(x)\}$, $h(x) = \max\{ f_i(x),\ i = 1,..., m\}$,

$$\Phi_\beta^* = \min\{\Phi_\beta(x) : x \in M\}, \qquad (2)$$

$$F_\lambda^* = \min\{F_\lambda(x) : x \in M\}. \qquad (3)$$

Далее предполагается, что задача (1) имеет решение. Функцию $\Phi_\beta(x)$ (или $F_\lambda(x)$) будем называть точной штрафной функцией, ес-

ли решения задач (1) и (2) (соответственно (3)) совпадают. Условия, при которых функции $\Phi_\beta(x)$ $F_\lambda(x)$ являются точными, исследовались, например, в [1].

Оценивание коэффициентов функции $\Phi_\beta(x)$. Сформулируем вспомогательные оптимизационные задачи, приближенные решения которых позволяют оценивать значения коэффициентов точных штрафных функций. Под приближенным решением понимается точка допустимого множества вспомогательной задачи, получаемая в результате применения некоторых упрощенных процедур поиска.

Существующие методы негладкой оптимизации, например, r-алгоритм Н. З. Шора [1, 4], мало чувствительны к завышенным значениям коэффициентов штрафных функций, поэтому высокая точность приближенных решений вспомогательных задач не требуется.

Лемма. [5**]** Пусть M — компактное множество, $\tilde{x}$ — решение задачи (2), значения штрафных коэффициентов фиксированы. Заданы числа $\varepsilon > 0$, $\delta > 0$ и последовательность точек $x_k \in M$, $k = 1, 2, \ldots$, сходящаяся к $\tilde{x}$. Пусть каждой x_k по некоторому правилу R_i поставлены в соответствие точки $z_{ki} = R_i(x_k) \in M$, $i = 1, \ldots, m$, такие, что

$$f_i(z_{ki}) \le \left(f_i(x_k) - \delta \right)^+,$$

выполняются неравенства

$$\Phi_\beta(x_k) \ge \Phi_\beta(z_{ki}) + \varepsilon \left\| z_{ki} - x_k \right\|, \text{ если } f_i(x_k) > 0. \qquad (4)$$

Тогда $\tilde{x} \in C$.

Лемма позволяет формулировать процедуры уточнения штрафных коэффициентов при решении задачи (2) каким-либо сходящимся алгоритмом. Пусть правила R_i, $i = 1, \ldots, m$ фиксированы, а на некоторой итерации k для индекса $p \in \{1, \ldots, m\}$ неравенство (4) нарушено. Тогда значение коэффициента β_p необходимо увеличить, чтобы это неравенство выполнялось. Обозначив $\chi_p(x) = f_0(x) + \sum_{i \ne p} \beta_i f_i^+(x)$,

$\zeta_p(z, x_k) = \dfrac{\chi_p(z) + \varepsilon \left\| z - x_k \right\|^2 - \chi_p(x_k)}{f_p^+(x_k) - f_p^+(z)}$, можно получить условие, которому должен удовлетворять коэффициент β_p:

$$\beta_p \ge \zeta_p(z_{kp}, x_k). \qquad (5)$$

При нарушении этого условия коэффициент β_p уточняется, т. е. можно полагать $\beta_p = \zeta_p(z_{kp}, x_k)$.

Значение коэффициента β_p существенно зависит от правил R_i, $i = 1,...,m$. Для выбора правил построения точки $z_{kp} = R_p(x_k)$ в [5] рассматривались различные оценочные задачи, в частности,

$$\min_z \left\{ \chi_p(z) + \varepsilon \left\| z - x_k \right\|^2 : z \in \bar{C}_p(x_k) \right\}, \tag{6}$$

$$\min_z \left\{ f_0(z) + \varepsilon \left\| z - x_k \right\|^2 : z \in \bar{C}_p(x_k) \right\}, \tag{7}$$

где

$$\bar{C}_p(x_k) = \{ z : f_p(z) \le (f_p(x_k) - \delta)^+, \ f_i(z) \le f_i^+(x_k), \ i \ne p, \\ i \ne p, \ i = 1,...,m, z \in M \}.$$

В качестве точки z_{kp} может использоваться любое допустимое решение задачи (6) или (7). Заметим, что поиск допустимого решения этих задач может представлять существенную проблему. Сужение допустимой области задач (6) или (7) приводит к увеличению соответствующих оптимальных значений (и к увеличению оценок штрафных коэффициентов) и может использоваться для упрощения получаемых задач.

Процедура упрощенного направленного поиска. Будем рассматривать исходную задачу (1) и штрафную функцию $F_\lambda(x)$. Предполагается, что $M = R^n$ и для множества C выполняется условие Слейтера. Для функции $F_\lambda(x)$ и фиксированной точки $x_k \in R^n$ задачу (7) можно записать как

$$\min_z \left\{ f_0(z) + \varepsilon \left\| z - x_k \right\|^2 : h(z) \le \left(h(x_k) - \delta \right)^+ \right\}, \tag{8}$$

неравенство (5), которому в данном случае должен удовлетворять коэффициент λ, принимает вид

$$\lambda \ge \zeta(z_k, x_k) = \frac{f_0(z_k) - f_0(x_k) + \varepsilon \left\| z_k - x_k \right\|^2}{h^+(x_k) - h^+(z_k)}, \tag{9}$$

где в качестве точки z_k будем использовать приближенное решение задачи (8).

Будем считать заданной базовую точку $y_0 \in C$, такую что $h(y_0) < 0$. Для $y \in R^n$ положим

$$z(y) = \arg\min_{t \in R} \left\{ f_0(z) + \varepsilon \left\| z - x_k \right\|^2 : z = y + t(y_0 - y), h(z) \le \left(h(x_k) - \delta \right)^+ \right\}. \tag{10}$$

Задача одномерного поиска (10) имеет решение при любом y, а точка $z(x_k)$ — это приближенное решение задачи (8). В дальнейшем

при определении точки z_k будем использовать приближенное решение $\tilde{z}(x_k)$ задачи (10), для которого выполняется

$$h(\tilde{z}(x_k)) = \left(h(x_k) - \delta\right)^+, \qquad (11)$$

т. е. $z_k = \tilde{z}(x_k)$. Такое правило будем называть **процедурой упрощенного направленного поиска** относительно функции $h(x)$ и точки x_k. Особенности таких процедур рассматривались в [5].

Процедура проектирования и направленного поиска. Рассмотрим случай, когда допустимое множество исходной задачи (1) имеет вид $C = \{x : Ax = b,\ f_i(x) \le 0,\ i = 1,...,m,\ x \in R^n\}$, где A — невырожденная $m_e \times n$-матрица, $b \in R^{m_e}$, m_e — число ограничений-равенств, $m_e < n$. Положим $h_1(x) = \max\left\{\left|A_i x - b_i\right|,\ i = 1,...,m_e\right\}$, $h_2(x) = \max\left\{f_i(x),\ i = 1,...,m\right\}$, $h(x) = \max\left\{h_1(x), h_2(x)\right\}$. Будем использовать штрафные функции вида $F_\lambda(x) = f_0(x) + \lambda \cdot h^+(x)$.

Обозначим $X_e = \{x : Ax = b,\ x \in R^n\}$, и рассмотрим задачу проектирования произвольной точки x_k на множество X_e

$$y_e(x_k) = \Pi_{X_e}(x_k) = \arg\min\left\{\frac{1}{2}\left\|x - x_k\right\|^2 : Ax = b,\ x \in R^n\right\}. \qquad (12)$$

Решение задачи (12) может быть представлено в виде

$$y_e(x_k) = A^T\left(AA^T\right)^{-1}\left(b - Ax_k\right) + x_k = \bar{b} + \bar{A}x_k,$$

где $\bar{b} = A^T\left(AA^T\right)^{-1} b$, $\bar{A} = I - A^T\left(AA^T\right)^{-1} A$. Будем предполагать, что задана начальная точка $y_0 \in X_e$, для которой выполняется $h_2(y_0) < 0$. Для уточнения коэффициента λ функции $F_\lambda(x)$ предлагается следующая ***процедура проектирования и направленного поиска:***

Шаг 1. Если $h_1(x_k) \le \left(h(x_k) - \delta\right)^+$, выполнить процедуру упрощенного направленного поиска относительно функции $h_2(x)$ и точки x_k, перейти на шаг 5.

Шаг 2. Определить точку $y_e(x_k) = \Pi_{X_e}(x_k)$ (решить задачу (12)).

Шаг 3. Определить точку y_1 отрезка $\left[y_e(x_k), x_k\right]$, ближайшую к x_k, для которой выполняется $h_1(y_1) = \left(h(x_k) - \delta\right)^+$.

Шаг 4. Выполнить процедуру упрощенного направленного поиска относительно функции $h_2(x)$ и точки y_1.

Шаг 5. Вычислить $\zeta(z_k, x_k)$ в соответствии (9) и, если условие $\lambda \geq \zeta(z_k, x_k)$ не выполняется, увеличить значение λ.

Действия, выполняемые на шагах 2, 3, 4, можно рассматривать как приближенное решение аналога задачи (8) при дополнительном ограничении $z \in aff\{a_1, a_2, a_2\}$.

Теорема 1. [5**]** Пусть множество C ограничено, функции f_0 и h_2 удовлетворяют условию Липшица на C. Заданы:

- точка $y_0 \in X_e = \{x : Ax = b\}$, для которой выполняется $h_2(y_0) < 0$,
- последовательность точек, сходящаяся к решению $\tilde{x}$ задачи (3).

Пусть для каждого $k = 1, 2, \ldots,$ величина $\zeta(z_k, x_k)$ определяется в соответствии с процедурой проектирования и направленного поиска. Тогда существует $\tilde{\zeta} < \infty$, такое, что при выборе $\lambda > \tilde{\zeta}$ штрафная функция $F_\lambda(x)$ — точная.

Для точки y_0 можно полагать $y_0 = \arg\min\{h_2(x) : Ax = b\}$.

Вычислительные эксперименты проводились на задачах ЛП:

$$\min\{\langle c, x\rangle : x \in C\},$$

где $C = \{x : Ax \leq b, A_e x = b_e, -100 \leq x^i \leq 100, i = 1, \ldots, n, x \in R^n\}$, $b \in R^m$, $b_e \in R^{m_e}$, $A \in R^{m \times n}$, $A_e \in R^{m_e \times n}$, $m_e < n$, x^i — i-я компонента вектора x. Множество $\{x : Ax \leq b,\ x \in R^n\}$ определялось набором m случайных опорных плоскостей к n-мерной сфере с центром в случайной точке x_0 с нормально распределенными координатами и радиусом $r = 1$. Для генерирования таких многогранников использовалась Octave-программа Е. А. Нурминского, приведенная в [6]. Ограничения $A_e x = b_e$ определялись случайными плоскостями, проходящими через точку x_0. Целевая функция задавалась вектором со случайными коэффициентами. Для решения задачи ЛП использовалась Octave–программа GLPK, что позволяло определять значения двойственных переменных. Для сгенерированной задачи ЛП формировалась штрафная функция $F_\lambda(x)$, для решения задачи безусловной оптимизации (3) использовался r-алгоритм Н. З. Шора [4] (Octave–программа из [7]).

Результаты вычислительных экспериментов по использованию процедуры проектирования и направленного поиска приведены в таблице, где n — число переменных задачи ЛП; m — число ограничений–неравенств; m_e — число ограничений–равенств, λ^* — сумма значений

(по абсолютной величине) двойственных переменных, полученных при решении задачи ЛП; $\tilde{\lambda}$ — значение коэффициента λ, полученное при использовании применяемой процедуры уточнения значения штрафного коэффициента; τ — величина сдвига из точки x_0 в направлении случайного вектора p при построении базовой точки $y_0 = x_0 + \tau p$.

В качестве базовой выбиралась точка $y_0(\tau) = x_0 + \tau p$, где $p \in R^n$ — случайный вектор такой, что $Ay_0(\tau) < b$, $A_e y_0(\tau) = b_e$, если $\tau < 1$, при $\tau = 1$ вектор $y_0(\tau)$ принадлежит границе множества C.

Таблица

Результаты вычислительных экспериментов

$m \times n$	m_e	λ^*	$\tilde{\lambda}$			
			$\tau = 0$	$\tau = 0.3$	$\tau = 0.6$	$\tau = 0.9$
20 × 10	3	2.1796	4.5815	6.6127	6.6127	6.6127
100 × 10	3	1.8820	4.6494	4.6494	4.6494	4.6494
100 × 20	7	2.6429	4.6072	4.6072	4.6072	4.6072
100 × 50	17	7.9559	13.476	13.476	16.403	16.403
100 × 100	33	4.8329	27.909	31.711	45.899	45.899
200 × 100	33	8.3726	11.528	11.528	14.239	38.517

Из таблицы следует, что для рассматриваемого класса задач с ограничениями равенствами предложенная процедура проектирования и направленного поиска генерирует приемлемые значения штрафных коэффициентов.

Выводы. Предложены упрощенные процедуры уточнения коэффициентов точных штрафных функций. Приведенные результаты вычислительных экспериментов для задач линейного программирования показали, что штрафные коэффициенты, полученные с использованием предложенных процедур, не слишком сильно отличались от оценок, определяемых оптимальными значениями двойственных переменных.

Рассмотренные процедуры построения приближенных решений вспомогательных задач могут также использоваться для блочных задач со связывающими переменными.

Список использованной литературы:

1. Shor N. Z. Nondifferentiable Optimization and Polynomial Problems. Amsterdam ; Dordrecht ; London : Kluwer Academic Publishers. 1998. 381 p.
2. Лаптин Ю. П. Вопросы построения точных штрафных функций. *Вестн. С.-Петерб. ун-та. Сер. 10: Прикладная математика*. 2013. Вып. 4. С. 21–31.
3. Лаптин Ю. П. Точные штрафные функции и выпуклые продолжения функций в схемах декомпозиции по переменным. *Кибернетика и системный анализ*. 2016. № 1. С. 96–108.

4. Шор Н. З., Журбенко Н. Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов. *Кибернетика*. 1971. № 3. С. 51–59.
5. Лаптин Ю. П., Бардадым Т. А. Проблемы определения коэффициентов точных штрафных функций. *Кибернетика и системный анализ.* 2019.
6. Нурминский Е. А. Проекция на внешне заданные полиэдры. *Вычисл. матем. и матем. физ.* 2008. Т. 48. № 3. С. 387–396.
7. Стецюк П. И. Программа ralgb5 для минимизации овражных выпуклых функций. *Математичне та програмне забезпечення інтелектуальних систем.* 2016. С. 185–197.

ON APPROXIMATE CALCULATION OF THE COEFFICIENTS OF EXACT PENALTY FUNCTIONS

Simplified procedures to specify the penalty coefficients more exactly are proposed. The results of computational experiments on randomly generated linear programming problems are given.

Key words: *exact penalty functions, structured optimization problems, decomposition methods.*

Получено 30.01.2019

УДК 519.9
DOI: 10.32626/2308-5878.2019-19.60-64

О. М. Литвин*, д-р фіз.-мат. наук,
О. О. Литвин*, д-р фіз.-мат. наук,
О. В. Ткаченко**, канд. фіз.-мат. наук

*Українська інженерно-педагогічна академія м. Харків,
**ДП «Івченко-Прогрес», м. Запоріжжя

МЕТОД ОДНОЧАСНОГО РІВНОМІРНОГО НАБЛИЖЕННЯ СПЛАЙНАМИ ТРИГОНОМЕТРИЧНИХ ФУНКЦІЙ ТА ЇХ ПОХІДНИХ

Наведені теореми про найкраще наближення сплайнами тригонометричних функцій та їх похідних, з дотриманням ізогеометричних властивостей.

Ключові слова: *найкраще наближення сплайнами, розривні періодичні сплайни.*

Вступ. На даний час наближення функцій із збереженням ізогеометричних властивостей досліджувалося в багатьох працях Відмітимо, зокрема, роботи [1–5] присвячені наближенню функцій і збереження ізогеометричних властивостей. Використовуються також узагальнені сплайн-функції. У роботах [4, 5] ізогеометричні власти-

вості забезпечувалися автоматично формулами інтерлінації із автоматичним збереженням потрібного класу диференційовності.

Зупинимось детальніше на використанні узагальнених сплайнів для наближення функції однієї змінної із збереженням її ізогеометричних властивостей. Як відомо [1], якщо у вузлах сітки $\Delta: a = x_0 < x_1 < \ldots < x_n = b$ задані значення функції $y = f(x)$: $y_i = f(x_i)$ $(i = \overline{0, n})$, то кубічним інтерполяційним сплайном на сітці Δ називається функція $S_3(f, x)$, неперервна на $[a, b]$ разом із своєю першою та другою похідними, яка є кубічним поліномом на кожному з відрізків $[x_0, x_1]$, $[x_1, x_2], \ldots, [x_{n-1}, x_n]$ і задовольняє умови $S_3(f, x_i) = y_i$, $i = \overline{0, n}$.

Доведено, що такий сплайн єдиний. Наведемо алгоритм його чисельної побудови. Позначимо $S_3''(f, x_i) = M_i$, $i = \overline{0, n}$. Побудуємо сплайн на відрізку $[x_i, x_{i+1}]$, $i = \overline{0, n-1}$. Оскільки на цьому відрізку $S_3(f, x)$ — це поліном третього степеня, то його друга похідна є поліномом першого степеня, тобто це лінійна функція, графік якої проходить через точки (x_{i-1}, M_{i-1}) і (x_i, M_i), $(i = 1, n)$. Користуючись формулою для рівняння прямої, яка проходить через дві задані точки, можна написати

$$S_3''(f, x) = M_{i-1}\frac{x_i - x}{h_i} + M_i\frac{x - x_{i-1}}{h_i}, h_i = x_i - x_{i-1}. \quad (1)$$

Проінтегруємо двічі рівність (1) і, знайшовши невідомі довільні сталі, що виникають під час інтегрування на відрізку $[x_{i-1}, x_i]$, отримуємо:

$$S'(f, a) = f_a', S'(f, b) = f'_b \quad (2)$$

або

$$S_3''(f, a) = f_a'', S_3''(f, b) = f_b'' \quad (3)$$

Кубічний сплайн на кожному проміжку $[x_i, x_{i+1}]$ сітки є кубічним поліномом, який може бути представлений через значення наближуваної функції і другі похідні на кінцях цього проміжку [a, b]

$$S_3(f, x) = \sigma(x) + \varphi(1-t)h_i^2 M_i + \varphi(t)h_i^2 M_{i+1}, \quad (4)$$

де

$$\sigma(x) = (1-t)f_i + tf_{i+1}, \quad \varphi(t) = \frac{t^3 - t}{6}, \quad t = \frac{x - x_i}{h_i}, \quad M_i = S''(x_i). \quad (5)$$

Невідомі вузлові значення M_i других похідних знаходяться шляхом розв'язання системи лінійних алгебраїчних рівнянь, яка отримується з умови гладкості сплайна $S_3(f, x)$, відповідних умовам гладкості сплайна класу $S_3(f, x) \in C^2[a, b]$. Відомо [1–3], що у пове-

дінці класичного кубичного сплайна виникають проблеми внаслідок використання функцій *у формулі* $\varphi(t)$ *вище написаної*. Тому в [1–3], використовується функція $\varphi(p,t) \in C^2[0,1]$, яка має вільний параметр p для управління поведінкою сплайна $S_3(f,x)$ і при $p=0$ співпадає з $\varphi(t) = \dfrac{(t^3-t)}{6}$, що породжує класичний кубічний сплайн.

У роботі пропонується доповнення до методу інтерлінації з автоматичним збереженням ізогеометричних властивостей (монотонного спадання, зростання, опуклості тощо), який досліджується детальніше лише для тригонометричних функцій $\sin 2\pi t$, $\cos 2\pi t$, $0 \le t \le 1$.

Наведені основні твердження методу, основані на використанні основної теореми про найкраще наближення монотонної на інтервалі $[a,b]$ неперервної функції $f(x) \in C[a,b]$ сталою.

Крім того, якщо ми наближуємо $f(x)$ кусково-сталими сплайнами найкращого наближення в нормі $C[a,b]$, то інтеграл від такого сплайна буде сплайном 1-го степеня, який теж дасть найкраще наближення для інтеграла. Тобто наближуючи невідому $f''(x)$ кусково-сталими сплайнами з невідомими значеннями сталих на підінтервалах монотонності, після інтегрування ми отримаємо сплайн 1-го степеня, який найкраще наближує $f(x)$ на $[a,b]$. Далі, інтегруючи цей сплайн 1-го степеня, отримаємо сплайн 2-го степеня, у якому можна вибрати невідомі сталі так, що цей сплайн буде найкраще наближувати $f(x) = \int f'(x)dx + C$.

Наведено аналіз результатів обчислювального експерименту, який підтверджує ефективність запропонованого методу одночасного наближення похідних і функції у припущенні, що нам відомі значення $y_k = f(x_k), k = \overline{1,Q}$, а також інтервали монотонності $f'(x)$ або $f''(x)$.

Означення. Кусково-сталий сплайн [6]

$$S(r,m,t) = \begin{cases} \dfrac{i}{m}, \ if \ \ t = \dfrac{1}{2\pi}\arcsin\dfrac{i}{m}, & i = \overline{0,m}, \\ \dfrac{i+0,5}{m}, \ if \ \ T_i < t < T_{i+1}, & i = \overline{0,m-1}. \end{cases}$$

Теорема 1. Сплайн $S(-1,n,t)$ будемо називати кусково-сталим сплайном (сплайном степеня 0), який найкраще наближує функцію $y = \sin 2\pi t$ на інтервалі $0 \le t \le 0,25$.

Доведення цієї теореми базується на відомому твердженні: найкраще наближення монотонної на $[a,b]$ функції $f(x)$ сталою C в нормі $C[a,b]$ існує і єдине, і визначається формулою

$$C = \frac{\max\limits_{a\le x\le b} f(x) + \min\limits_{a\le x\le b} f(x)}{2} = \frac{f(a)+f(b)}{2}. \qquad (6)$$

При цьому похибка наближення

$$\max\left|f(x)-C\right| \le \frac{M-m0}{2}, \qquad (7)$$

де $m0 = \min\limits_{a\le x\le b} f(x)$, $M = \max\limits_{a\le x\le b} f(x)$.

З теореми 1 витікає, що для зменшення похибки наближення в m разів достатньо розбити інтервал монотонності $[a,b]$ на m підінтервалів $\left[x_i,\ x_{i+1}\right]$, $i = \overline{0,m-1}$, $x_0 = a$, $x_m = b$, $x_i = a + \frac{b-a}{m} i$ і в кожному з цих підінтервалів наближувати $f(x)$ сталою $C_i = \frac{f(x_i)+f(x_{i+1})}{2}$. Тоді результуючий сплайн кусково-сталий буде наближувати $f(x)$ в нормі $C[a,b]$ з похибкою $\varepsilon = \frac{\left|f(b)-f(a)\right|}{2m}$. Зауважимо, що точки x_i цього розбиття знаходяться як розв'язки рівнянь $f(x) = f(a) - \frac{f(b)-f(a)}{m} i$, $i = \overline{0,m}$, якщо $f(a) < f(b)$ і $f(x) = f(a) - \frac{f(a)-f(b)}{m} i$, $i = \overline{0,m-1}$, якщо $f(a) > f(b)$.

Теорема 2. Сплайн 1-го степеня $S(0,m,t) \in C\left[0,0.25\right]$, який найкраще наближує $\int f(t)dt + C = -\frac{\cos 2\pi t}{2\pi} + C$, можна представити у вигляді

$$S1(t) = S(0,m,t) = \begin{cases} \int\limits_x^{T_m} S0(\tau)d\tau, & T_{m-1} \le t \le T_m, \\ \int\limits_x^{T_{m-1}} S0(\tau)d\tau + S1(T_{m-1}), & T_{m-2} \le t \le T_{m-1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \int\limits_x^{T_1} S0(\tau)d\tau + S1(T_1), & T_0 < t \le T_1. \end{cases} \qquad (8)$$

Таким чином, з формули (8) випливає, що $S1(T_m) = 0$, що означає автоматичне виконання відомого співвідношення

$$\left(-\frac{\cos 2\pi t}{2\pi}\right)_{t=0,25} = -\frac{\cos\frac{\pi}{2}}{2\pi} = 0.$$

На рисунку показано графік функції $y = S1(t)$ сплайна 1 степеня, $t \in [0,1]$, який отримується шляхом інтегрування кусково-сталого сплайна найкращого наближення функції $y'(t) = \sin 2\pi t$, $t \in [0,1]$.

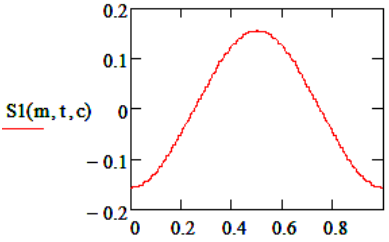

Рисунок.

Висновки. Наведені твердження, які описують метод одночасного рівномірного наближення сплайнами тригонометричних функцій та їх похідних. Показано графік функції, яка отримується шляхом інтегрування кусково-сталого сплайна її похідної.

Список використаних джерел:

1. Квасов Б. И. Методы изогеометрической аппроксимации сплайнами. М. : ФИЗМАТЛИТ, 2006.
2. Spath H. Spline Algorithms for Curves and Surfaces. *Winnipeg: Uni-tas Mathematica Publ*. 1974.
3. Богданов В. В., Волков Ю. С. Выбор параметров обобщённых кубических сплайнов при выпуклой интерполяции. *Сиб. журн. вычисл. математики*. Новосибирск, 2006. Т. 9. № 1. С. 5–22.
4. Литвин О.М. Інтерлінація функцій та деякі її застосування. Харків : Основа, 2002. 544 с.
5. Сергиенко И. В., Литвин О. Н., Литвин О. О., Ткаченко А. В., Грицай О. Л. Интерлинация эрмитового типа на системе непересекающихся линий (Обзор). *Кибернетика и системный анализ.* 2015. Т. 51, № 2. С. 134–144.
6. Литвин О. М. Періодичні сплайни і новий метод розв'язання плоскої задачі рентгенівської комп'ютерної томографії. *Системний аналіз, управління і інформаційні технології:* Вісн. Харків. держ. політех. ун-ту : зб. наук. праць. 2000. № 125. С. 27–35.

BY THE METHOD OF SIMULTANEOUS UNIFORM APPROXIMATION BY SPLINES OF TRIGONOMETRIC FUNCTIONS AND THEIR DERIVATIVES

Method is presented on the best approximation by splines of trigonometric functions and their derivatives, with respect to isogeometric properties.

Key words: *best approximation by splines, discontinuous periodic splines.*

Одержано 05.02.2019

УДК 518.9
DOI: 10.32626/2308-5878.2019-19.65-71

А. С. Макаренко, д-р физ.-мат. наук, профессор
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского», г. Киев

НОВЫЕ ПОДХОДЫ К СОКРАЩЕНИЮ ИСКУССТВЕННЫХ КОЛЕБАНИЙ В ЧИСЛЕННЫХ РЕШЕНИЯХ. АНТИДИФФУЗИЯ, АНТИДИСПЕРСИЯ И ЛАНГОЛЬЕРЫ

Две наиболее известные ошибки — это искусственное сглаживание решения и колебания в решениях вблизи мест с большими производными решений (вблизи фронтов решения). Некоторые методы улучшения численных решений эволюционных уравнений предложены на основе теоретических соображений. В качестве первых примеров предложены искусственная вязкость и искусственная дисперсия для разностных схем газовой динамики. Предлагается новый класс инструментов для улучшения численных решений «лангольеры». «Лангольеры» — это специальные операторы разностей, которые должны применяться на каждом временном шаге после запуска оригинальных разностных схем. Конструкция «лангольеров» позволяет снизить диссипативные и дисперсионные ошибки схем. Примерами являются антидиффузионные, антидисперсионные и специальные схемы.

Ключевые слова: *численные схемы, дисперсия, дисперсия, негладкие растворы, антидисперсия, «Langoliers», нелинейные задачи.*

Вступление. Хорошо известно, что разностные схемы для приближенных решений эволюционных уравнений обычно имеют некоторые ошибки в пределах теоретической точности схем [1–4]. Две наиболее известные ошибки — это искусственное сглаживание решения и колебания в решениях вблизи мест с большими производными решений (вблизи фронтов решения). Для предотвращения таких эффектов предложено множество специальных инструментов: искусственная вязкость в схемах [1], искусственная дисперсия в схемах [3, 5, 6], антидиффузионные [7], ENO (по существу не-колебательные) схемы [8] и т. д. Но проблема остается открытой, особенно в разработке специальных схем.

Из-за усложнения уравнений, которые следует использовать для моделирования развивающихся сред и систем в гидродинамике, газодинамике, плазме, реологии, проблема разработки более точных разностных схем очень важна. Для достижения этой цели необходимо знать особенности поведения числовых схем, признаки «артефактов» в численных решениях и лучшее теоретическое понимание разностных схем как объектов.

Поэтому тут некоторые методы улучшения численных решений эволюционных уравнений предлагаются на основе теоретических соображений. В случае линейных уравнений предлагаемые инструменты могут повысить порядок точности. В качестве первых примеров предложены искусственная вязкость и искусственная дисперсия для разностных схем газовой динамики. Предлагается новый класс инструментов для улучшения численных решений «лангольеры». Конструкция «лангольеров» позволяет снизить диссипативные и дисперсионные ошибки схем. Таким образом, «лангольеры» — это реализация новой идеи повышения точности (с использованием дополнительного шага по времени), которая является вспомогательной для идеи использования пространственно расширенного шаблона. Примерами являются антидиффузионные, антидисперсионные и специально построенные разностные схемы.

Различные иллюстративные примеры таких инструментов рассматриваются для уравнений газовой динамики и для волнового уравнения. Также в качестве примеров рассмотрены некоторые вычисления новых многомасштабных задач: гиперболическая модификация уравнения Бюргерса и взрывные решения.

1. Диссипация и дисперсия конечно-разностных схем. Термины «диссипация» и «дисперсия» разностных схем имеют строгий смысл в случае, когда исходные уравнения в частных производных являются линейными и имеют постоянные коэффициенты. В таком случае разностные гармоники разностной схемы (или гармоника непрерывного аналога разностной схемы) являются адекватным инструментом для исследования свойств численных схем. Такой подход хорошо известен и предлагается в любом учебнике по численным методам (например, см. [1–3, 9]).

Для каждой проблемы или процесса результаты анализа имеют конкретный вид. Но общая схема исследований остается прежней, то есть анализ гармоник и их дисперсионного соотношения проводится для начального условия и для метода его аппроксимации. Поэтому в дальнейшем мы проиллюстрируем схему методов, а также средств улучшения на простейшем примере — уравнении переноса или адвекции.

Как пример, мы рассмотрим задачу Коши для уравнения переноса:

$$Lu = \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0, \;\; a = \text{const} \tag{1}$$

с начальными условиями. Для иллюстрации рассмотрим также общий класс явных численных схем для уравнения (1)

$$\Lambda y = \frac{y_j^{n+1} - y_j^n}{\tau} + \sum_{l=-m_1}^{m_2} a_l y_{j+l}^n = 0 \; . \tag{2}$$

Для проведения анализа гармоник схемы (2) рассмотрим решения схем специального вида (числовая гармоника):

$$y_j^n = q_k^n \exp(ikx_j). \tag{3}$$

Коэффициент перехода может быть представлен в виде также как

$$q_k = \rho_k \exp(i\varphi_k). \tag{4}$$

В формуле (4) $\rho_k = \mathrm{mod}\, q_k = [(\mathrm{Re}\, q_k)^2 + (\mathrm{Im}\, q_k)^2]^{1/2}$ модуль коэффициента перехода $\varphi_k = -\arg q_k = \mathrm{arctg}(-\mathrm{Im}\, q_k / \mathrm{Re}\, q_k)$. Назовем $v_k = q_k / k\tau$ фазовую скорость *k*-й гармоники. Вводятся непрерывные аналоги модуля для коэффициента перехода $\rho(\zeta) = a\varphi(\zeta)/\gamma\zeta$, от аргумента ζ, такого, что $\rho(\zeta_k) = \rho_k$, а также фазовой скорости $v(\zeta_k) = v_k$, где $\zeta_k = kh$. Рассмотрим численные схемы, для которых

$$\begin{gathered}\rho(\zeta) = 1 - \omega(\zeta),\ 0 \le \omega(\zeta) \le 2,\ \zeta \le 1,\\ \omega(\zeta) = c\zeta^s + O(\zeta^{s+2}),\ c = \mathrm{const},\ s = 2p.\end{gathered} \tag{5}$$

Из работ Рихтмайера [4] известно, что такая схема имеет *s*-й порядок диссипации. Согласно [10] схема имеет *m*-й порядок дисперсии, если дисперсионная функция может быть записана в виде

$$\upsilon(\zeta) = a[1 + \theta\zeta^m + O(\zeta^{m+2})]),\ \theta = \mathrm{const}. \tag{6}$$

Теперь воспользуемся некоторым результатом из [11]. Там характер схемы (коэффициент перехода схемы) был представлен в виде

$$q(\zeta) = \exp[-i\gamma\zeta + \Psi(\zeta)]. \tag{7}$$

Тогда *s* интерпретируется как порядок диссипации, а *r* как порядок аппроксимации. Можно получить очень важную корреляцию между порядком аппроксимации, диссипацией и дисперсией схемы:

$$r = \min(s, m+1) - 1. \tag{8}$$

Очень важный вывод состоит в том, что порядок аппроксимации может быть определен либо порядком диссипации, либо порядком дисперсии. Отметим, что порядок аппроксимации также определяет порядок сходимости (в зависимости от гладкости решений). Из такого анализа можно обнаружить, что, например, для схем четного порядка схемы имеют четный порядок аппроксимации, а скорость сходимости определяется дисперсионными эффектами. Это означает, что большие нефизические колебания, которые обычно наблюдаются в схемах четного порядка аппроксимации при вычислении негладких решений, обусловлены именно дисперсией разностных гармоник. Отметим, что в работах [10, 12–14] рассмотрены другие уравнения и многомерный случай. Применение результатов по порядку диссипации и дисперсии позволяет понять «артефакты» в численных решениях эволюционных уравнений и предложить новые инструменты для их подавления или уменьшения.

2. Некоторые существующие инструменты для уменьшения «артефактов» в расчетах. Здесь опишем конструкцию некоторых более или менее известных инструментов для улучшения качества численных решений и дадим описание их механизмов.

Выбор новой схемы с возрастающей точностью. Первый подход к уменьшению «артефактов» заключается в использовании другой схемы с повышенной точностью. Но обычно это дорого и сложно в теоретических аспектах, особенно для моделирования по нелинейным уравнениям в многомерных случаях. Далее обсудим методы улучшения «базовых» оригинальных схем с помощью специальных инструментов.

Метод искусственной вязкости. В соответствии с этим подходом к разностной схеме для подавления искусственных колебаний следует добавить специальные члены путем добавления нефизической вязкости [1, 3, 9] и много других работ.

Искусственная дисперсия. Особые члены следует добавить в разностную схему для подавления искусственных колебаний путем добавления нефизической дисперсии [3, 5, 6].

Анти-диффузия. Идея антидиффузии развита, возможно, со времен работ Бориса Дж. и Д. Бука [1, 7, 15]. В антидиффузии специальный оператор фильтрации применяется к числовому решению после выполнения шага условной схемы для уменьшения колебаний путем применения специальных правил к решению. Показано, что действие такого фильтра эквивалентно некоторой части искусственной сглаживающей вязкости. Антидиффузия уже имеет множество применений, особенно в газовой динамике. Но трудности приложений заключаются в нелинейном характере фильтра.

3. Новые инструменты для улучшения численных схем. Здесь мы кратко опишем некоторые инструменты для улучшения решений, которые имеют в качестве базы концепции из раздела 2.

Композитные схемы высшего порядка. Исследования фазовой скорости и переходных модулей показывают, что такие функции могут иметь как положительную, так и отрицательную дисперсию (то есть гармоника разностной схемы может быть медленнее, чем гармоника исходного дифференциального уравнения); также переходные модули измеряют уровень уменьшения (увеличения) амплитуды гармоники и могут быть меньше или больше, чем в гармонике исходного уравнения. Таким образом, последовательное применение двух схем с различными свойствами для перехода с одного временного уровня на следующий уровень эквивалентно прикладной составной схеме с различными свойствами [16].

Анти-дисперсия. Цель антидисперсии состоит в том, чтобы уменьшить искусственную дисперсию численных схем и обеспечить

применение специального разностного оператора, который имеет дисперсию, противоположную дисперсии базовой разностной схемы [17]. В качестве такого оператора полезно взять приближение простейших дифференциальных уравнений с необходимой дисперсией.

'Langoliers. Полезно ввести специальное название для нового класса инструментов, которое следует применять после применения базовой схемы на каждом временном шаге расчета. Мы назвали его «лангольерами», потому что такие инструменты применяются в каждой точке пространственной сетки разностной схемы на заданном временном уровне, и действие таких «лангольеров» заключается в «съедании» «искусственных» дефектов численного решения в каждой точке решения. После применения базовой схемы решение имеет много искусственных колебаний (если истинное решение — ступенчатая функция). Применение «Langolier» существенно уменьшает количество ошибок. Антидиффузионный фильтр можно рассматривать как символ «лангольер» типа «анти-вязкость». Также могут существовать другие случаи проектирования "лангольеров". Мы можем использовать не один «Langolier» между временными уровнями, а последовательность разных «Langoliesr». Например, как следует из теории дисперсии и диссипации схем, мы можем для линейных уравнений теоретически получить любой порядок аппроксимации составных «базовых схем» + серии специально построенных «лангольеров». Одна из конструкций состоит из «лангольеров» «антидисперсионного» и «антидиффузионного» характера (но, конечно, возрастающего порядка диссипации или дисперсии и, следовательно, увеличивающейся структуры).

4. Нелинейный случай. Как мы уже отмечали, вышеупомянутые подходы уже были разработаны и проверены в случае некоторых уравнений (линейное уравнение переноса, волновое уравнение, уравнение Кадомцева–Петвиашвили). Но и опыт численного решения нелинейных уравнений ведет к заключению о применимости инструментов выше. Ключевой подход заключается в применении двух идей: 1) линеаризации нелинейного уравнения вокруг «базового» решения для исходного нелинейного уравнения и 2) идеи «замороженных» коэффициентов полученного линеаризованного уравнения [1–3, 9]. Тогда анализ гармоник должен проводиться локально. В таком случае коэффициенты таких инструментов должны зависеть от значений решений в данной точке в данный момент времени. Результаты такого анализа для случая нелинейного уравнения Клайна–Гордона опубликованы в [13]. Другой интересный пример применения предложенной концепции к нелинейным уравнениям описан в [16]. Предложенный подход также весьма перспективен для численных расчетов коллапсов, взрывных решений или решений с осо-

бенностями. Область применения «лангольеров» во время вычислений может быть сосредоточена вблизи точек сингулярностей.

Выводы. Таким образом, в статье описаны специальные методы и теория их применения для уменьшения искусственных ошибок типа «расплывания» и «колебаний» при расчете решений эволюционных уравнений. Очень важно, что предлагаемые средства также пригодны для вычисления решений односторонних физических процессов с памятью, потому что эволюционные уравнения с памятью входят компонентом в математической постановка соответствующих математических задач. Кроме того, предлагаемые методы становятся особенно перспективными в связи с современной разработкой средств для параллельных вычислений (GRID-вычислений) решений. Это связано с тем, что «лангольеры» можно использовать параллельно в каждом узле числовой решетки, что приводит к повышению точности всех методов аппроксимации. Обратим внимание на то, что методы глубокого обучения могут быть использованы для построения «лангольеров».

Список использованной литературы:

1. Roache P. J. Computational Fluid Dynamics [Russian translation]. Moscow : Mir, 1980
2. Kalitkin N. N. Numerical methods [in Russian], Moscow : Nauka, 1990.
3. Samarskii A. A., Popov Yu. P. Difference Methods for Solving Problems of Gas Dinamics. [in Russian]. Moscow : Nauka, 1980.
4. Richtmyer R. D., Morton K. W. Difference methods for initial — value problems. N.Y. : Wiley and Sons. 1967.
5. Mukhin S. I., Popov S. B., Popov Yu. P. U.S.S.R. *Computational Mathematics and Mathematical Physics.* Springer Translation. 1983. Vol. 23. 45 p.
6. Sengupta T., Dipankar A., Sagaut P., *J. Comput. Physics.* 2007. Vol. 226. 1211 p.
7. Book D. L., Boris J. P., Hain K. *J. of Comput. Phys.* 1975. Vol. 18. 248 p.
8. Harten A., Engquist B., Osher S., Chakrabarthy S. *J. Comput. Physics.* 1987. Vol. 71. 231 p.
9. Shokin Yu. I., Yanenko N. N. Differential Approximation Method [in Russian]. Novosibirsk : Nauka, 1985.
10. Moskalkov M. N. Numerical Analysis [in Russian], Kiev : Int. of Cybernetics, 1978. Vol. 75.
11. Brenner P., Thomee V. *Math, Scandinavia.* 1970. Vol. 27.
12. Nikolskii S. M. Approximation of functions of several variables and imbedding theorems [in Russian], M. : Nauka, 1977.
13. Makarenko A. S. *Chisl. Met. Mech. Sploshn. Sredy* [in Russian]. 1982. Vol. 13. 81 p.
14. Makarenko A. S., Moskalkov M. N. *Chisl. Met. Mech. Sploshn. Sredy* [in Russian]. 1981. Vol. 12. 64 p.
15. Bokanowski O., Martin S., Munos R., Zidan H. *Applied Numerical Mathematics.* 2006. Vol. 56. 1147 p.
16. Makarenko A. S., Moskalkov M. N. Vychisl. Prikl. Matem. [in Russian]. Kiev : Kiev Univ, 1980. Issue 41.

17. Makarenko A., Moskalkov M. U.S.S.R. *Computational Mathematics and Mathematical Physics.* Springer Translation. 1983. Vol. 23. 999 p.

NEW APPROACHES FOR REDUCING ARTIFICIAL OSCILLATIONS IN NUMERICAL SOLUTIONS. ANTI-DIFFUSION, ANTI-DISPERSION AND LONGOLIERS

Abstract. Two most known errors is the artificial smoothing of the solution and oscillations in the solutions near the places with high derivatives of the solutions (near the sharp fronts of the solution). Some methods of improving numerical solutions of evolution equations are proposed on the base of theoretical considerations. The artificial viscosity and artificial dispersion for difference schemes of gas dynamics are proposed as the first examples. A new class of tools for improving numerical solutions is proposed — «Langoliers». «Langoliers» are special difference operators which should be applied at each time steps after the running of original difference schemes. The design of «Langoliers» allows reducing the dissipative and dispersive errors of schemes. The examples are anti-diffusion, anti-dispersion and specially constructed difference schemes.

Key words: *numerical schemes; dispersion; dissipation; non-smooth solutions, anti-dispersion; «Langoliers»; non-linear problems.*

Получено 15.02.2019

УДК 517.946
DOI: 10.32626/2308-5878.2019-19.71-77

В. В. Маринець, д-р фіз.-мат. наук,
О. І. Когутич, магістр

ДВНЗ «Ужгородський національний університет», м. Ужгород

ПРО ОДИН ПІДХІД ДОСЛІДЖЕННЯ КРАЙОВОЇ ЗАДАЧІ ДЛЯ КВАЗІЛІНІЙНОГО РІВНЯННЯ ГІПЕРБОЛІЧНОГО ТИПУ З РОЗРИВНОЮ ПРАВОЮ ЧАСТИНОЮ

Будується конструктивний швидкозбіжний двосторонній метод дослідження та наближеного розв'язання крайової задачі для квазілінійного хвильового рівняння на площині з розривною правою частиною в області із складною структурою краю. Встановлюються достатні умови існування функцій порівняння, регулярного або іррегулярного розв'язку розглядуваної крайової задачі, його єдиності та знакосталості.

Ключові слова: *«вільні» криві, іррегулярний розв'язок, функції порівняння, умови узгодження.*

Вступ. Крайові задачі для квазілінійних рівнянь гіперболічного типу з неперервними правими частинами в різних областях із складною структурою краю розглядались в роботах [1–3], в яких встановлено до-

статні умови існування та єдиності їх регулярних або іррегулярних розв'язків. Дана робота є продовженням досліджень, приведених у [4].

Розглянемо в R^2 область $D = D_1 \cup D_2 \cup D_3$ [5, с. 250], де

$$D_1 = \{(x,y) \mid x \in (x_0, x_1],\ y \in (y_0, y_1]\},$$

$$D_2 = \{(x,y) \mid x \in [x_0, x_1],\ y \in (y_1, g_1(x))\},$$

$$D_3 = \{(x,y) \mid x \in (x_1, x_2],\ y \in (g_2(x), y_1]\},$$

$$x_0 < x_1 < x_2, \quad y_0 < y_1 < y_2,$$

а $y = g_r(x) \Leftrightarrow x = \kappa_r(y)$, $r = 1,2$ — «вільні» криві, причому $g_r'(x) > 0$, $g_1(x_{r-1}) = y_r$, $g_2(x_r) = y_{r-1}$.

Позначимо $D^* := D \setminus E_1 \cup E_2$,

де

$$E_1 = \{(x,y) \mid x \in [x_0, x_1],\ y = y_1\},\quad E_2 = \{(x,y) \mid y \in [y_0, y_1],\ x = x_1\}.$$

Досліджується наступна крайова задача: в просторі функцій $C^*(\overline{D}) := C^{(1.1)}(D^*) \cap C(\overline{D})$ знайти розв'язок диференціального рівняння

$$\begin{aligned} U_{xy}(x,y) + a_1(x,y)U_x(x,y) + a_2(x,y)U_y(x,y) = \\ = f(x,y,U(x,y)) := f[U(x,y)], \end{aligned} \tag{1}$$

який задовольняє крайові умови

$$U(x_0,y) = \psi(y),\ y \in [y_0, y_1],\ U(x,y_0) = \varphi(x),\ x \in [x_0, x_1], \tag{2}$$

$$U(x, g_1(x)) = \omega_1(x),\ x \in [x_0, x_1], \tag{3}$$

$$U(x, g_2(x)) = \omega_2(x),\ x \in [x_1, x_2], \tag{4}$$

$$\psi(y_0) = \varphi(x_0),\ \psi(y_1) = \omega_1(x_0),\ \varphi(x_1) = \omega_2(x_1), \tag{5}$$

а функція $f[U(x,y)] = f_s[U_s(x,y)]$, $(x,y) \in \overline{D_s}$, $s = 1,2,3$, $f_s[U_s(x,y)] \in C(\overline{B}_s)$, $f_s : \overline{B}_s \to R$, $\overline{B}_s \subset R^3$, $Пр_{x0y}\overline{B}_s = \overline{D}_s$, причому

$$U_2(x,y_1) = U_1(x,y_1),\ x \in [x_0, x_1],\ U_3(x_1,y) = U_1(x_1,y),\ y \in [y_0, y_1]. \tag{6}$$

Зауважимо, що умови (5) є умовами узгодженості крайових умов (2)–(4), а (6) — умови неперервності розв'язку задачі (1)–(5), якщо він існує. Права частина рівняння (1) $f[U(x,y)]$ всюди неперервна функція в області $\overline{B} := \overline{B}_1 \cup \overline{B}_2 \cup \overline{B}_3$, за виключенням характеристик $x = x_1$, $y = y_1$ рівняння (1), вздовж яких вона може мати скінчені розриви.

Очевидно, якщо існує розв'язок задачі (1)–(5) $U(x,y)$, то $U(x,y)=U_s(x,y)$, $(x,y)\in\overline{D}_s$, $s=1,2,3$, де $U_1(x,y)$ є розв'язком задачі Гурса (1), (2), (5) при $(x,y)\in\overline{D}_1$, $U_2(x,y)$ — задачі Дарбу (1), (3), (5) і першої з умов (6), $(x,y)\in\overline{D}_2$, а $U_3(x,y)$ — задачі Дарбу (1), (4), (5) і другої з умов (6) при $(x,y)\in\overline{D}_3$.

Надалі вважатимемо, що задані функції

$$a_1(x,y)\in C^{(1.0)}(D),\ a_2(x,y)\in C^{(0.1)}(D),\ \psi(y)\in C^1[y_0,y_1],$$
$$\varphi(x)\in C^1[x_0,x_1],\ \omega_r(x,y)\in C^1[x_{r-1},x_r],\ r=1,2,$$

причому

$$a_{1_x}(x,y)=a_{2_y}(x,y). \tag{7}$$

Лема 1. Якщо $f_s[U_s(x,y)]\in C(\overline{B}_s)$ і виконується умова (7), то крайова задача (1)–(5) еквівалентна системі інтегральних рівнянь вигляду

$$U_s(x,y)=\gamma_s(x,y)+\varepsilon_s T_{1,s}F_1[U_1(\xi,\eta)]+T_sF_s[U_s(\xi,\eta)],$$
$$(x,y)\in\overline{D}_s,\quad s=1,2,3,\quad \varepsilon_1=0,\quad \varepsilon_2=\varepsilon_3=1 \tag{8}$$

(тут і надалі використовуватимуться позначення, приведені в [4]).

Лема 2. Нехай виконуються умови леми 1 і крайова задача (1)–(5) має розв'язок $U(x,y)$.

Тоді він належатиме просторові функцій $C^{(1.1)}(D)\cap C\left(\overline{D}\right)$ (буде регулярним), якщо $f[U(x,y)]\in C(\overline{B})$ і $\rho_1=\rho_2=0$. У супротивному випадку він буде іррегулярним.

Надалі будемо вважати, що функції $F_s[U_s(x,y)]\in C_2(\overline{B}_s)$, тобто, що вони задовольняють наступні умови [6, с. 79]:

1) $F_s[U_s(x,y)]\in C(\overline{B}_s)$;

2) в просторі функцій $C(\overline{B}_{s,1})$, $\overline{B}_{s,1}\subset R^4$, $\Pi p_{xOy}\overline{B}_{s,1}=\overline{D}_s$, $s=1,2,3$, існує така функція

$$H_s[x,y,U_s(x,y);V_s(x,y)]:=H_s[U_s(x,y);V_s(x,y)],\ \text{що}$$

(a) $H_s[U_s(x,y);V_s(x,y)]\equiv F_s[U(x,y)]$,

(b) для довільної пари неперервних функцій $U_s(x,y)$, $V_s(x,y)\in\overline{B}_{s,1}$, які задовольняють умови $U_s(x,y)\ge V_s(x,y)$, $(x,y)\in\overline{D}_s$, в області $\overline{B}_{s,1}$ виконується нерівність

$$H_s[U_s(x,y);V_s(x,y)]-H_s[V_s(x,y);U_s(x,y)]\le 0; \tag{9}$$

3) функція $H_s[U_s(x,y);V_s(x,y)]$ в області $\overline{B}_{s,1}$ задовольняє умові Ліпшиця, тобто, для довільних двох пар неперервних в $\overline{D}_s$ функцій $U_{s,r}(x,y)$, $V_{s,r}(x,y) \in \overline{B}_{s,1}$, виконується умова

$$\left|H[U_{s,1}(x,y);U_{s,2}(x,y)]-H[V_{s,1}(x,y);V_{s,2}(x,y)]\right| \le L_s\left(\left|W_{s,1}(x,y)\right|+\left|W_{s,2}(x,y)\right|\right),$$

де $W_{s,r}(x,y) := U_{s,r}(x,y)-V_{s,r}(x,y)$, $r=1,2$, а L_s — стала Ліпшиця, $s=1,2,3$.

Неважко переконатися, що якщо функція $F_s[U_s(x,y)] \in \in C(\overline{B}_s)$ і має обмежену частинну похідну першого порядку по $U_s(x,y)$, то вона завжди належить просторові $C_2(\overline{B}_s)$, $s=1,2,3$. Зворотнє твердження не справедливе.

Нехай $z_{s,p}(x,y), v_{s,p}(x,y) \in C(\overline{D}_s)$ належать області $\overline{B}_{s,1}$ для всіх $s=1,2,3$ та $p \in N_0 := N \cup \{0\}$.

Введемо позначення:

$$f_s^{\,p}(x,y) := H_s[z_{s,p}(x,y);v_{s,p}(x,y)],$$
$$f_{s,p}(x,y) := H_s[v_{s,p}(x,y);z_{s,p}(x,y)],$$
$$R_s^p(x,y) := \gamma_s(x,y)+\varepsilon_s T_{1,s} f_1^{\,p}(\xi,\eta)+T_s f_s^{\,p}(\xi,\eta),$$
$$R_{s,p}(x,y) := \gamma_s(x,y)+\varepsilon_s T_{1,s} f_{1,p}(\xi,\eta)+T_s f_{s,p}(\xi,\eta),$$
$$\alpha_{s,p}(x,y) := z_{s,p}(x,y)-R_s^p(x,y),$$
$$\beta_{s,p}(x,y) := v_{s,p}(x,y)-R_{s,p}(x,y),$$
$$\overline{z}_{s,p}(x,y) := z_{s,p}(x,y)-q_{s,p}(x,y)W_{s,p}(x,y),$$
$$\overline{v}_{s,p}(x,y) := v_{s,p}(x,y)+c_{s,p}(x,y)W_{s,p}(x,y), \quad p \in N,$$
$$F_s^{\,p}(x,y) := H[\overline{z}_{s,p}(x,y);\overline{v}_{s,p}(x,y)],$$
$$F_{s,p}(x,y) := H[\overline{v}_{s,p}(x,y);\overline{z}_{s,p}(x,y)],$$
$$\overline{R}_s^{\,p}(x,y) := \gamma_s(x,y)+\varepsilon_s T_{1,s} F_1^{\,p}(\xi,\eta)+T_s F_s^{\,p}(\xi,\eta),$$
$$\overline{R}_{s,p}(x,y) := \gamma_s(x,y)+\varepsilon_s T_{1,s} F_{1,p}(\xi,\eta)+T_s F_{s,p}(\xi,\eta),$$

де $q_{s,p}(x,y)$, $c_{s,p}(x,y)$ є довільними з простору $C(\overline{D}_s)$ функціями, які задовольняють умови

$$0 \le q_{s,p}(x,y) \le 0,5, \quad 0 \le c_{s,p}(x,y) \le 0,5, \tag{10}$$

для всіх $p \in N$ та $(x,y) \in \overline{D}_s$, $s=1,2,3$ $(q_{s,0}(x,y)=c_{s,0}(x,y)=0)$.

Побудуємо послідовності функцій $\left\{ z_{s,p}(x,y) \right\}$ та $\left\{ v_{s,p}(x,y) \right\}$ згідно формул [7, с. 132]

$$z_{s,p+1}(x,y) = \overline{R}_s^p(x,y),\ v_{s,p+1}(x,y) = \overline{R}_{s,p}(x,y), \qquad (11)$$

де за нульове наближення $z_{s,0}(x,y),\ v_{s,0}(x,y) \in \overline{B}_{s,1}$ вибираємо довільні з простору $C(\overline{D}_s)$ функції, які при $(x,y) \in \overline{D}_s$ задовольняють умови

$$W_{s,0}(x,y) \geq 0,\ \alpha_{s,0}(x,y) \geq 0,\ \beta_{s,0}(x,y) \leq 0, (x,y) \in \overline{D}_s. \qquad (12)$$

Означення. Функції $z_{s,0}(x,y),\ v_{s,0}(x,y) \in C(\overline{D}_s),\ s = 1,2,3,$ які належать області $\overline{B}_{s,1}$ і задовольняють умови (12), називаються функціями порівняння крайової задачі (1)–(5).

Лема 3. Якщо $F_s[U_s(x,y)] \in C_2(\overline{B}_s)$, то множина функцій порівняння задачі (1)–(5) не порожня.

Лема 4. Якщо $F_s[U_s(x,y)] \in C_2(\overline{B}_s)$ та інтегральні рівняння (8) в просторі функцій $C(\overline{D}_s),\ s = 1,2,3$ мають розв'язки $U_s(x,y)$, які задовольняють нерівності

$$v_{s,0}(x,y) \leq U_s(x,y) \leq z_{s,0}(x,y), \quad (x,y) \in \overline{D}_s,\ s = 1,2,3, \qquad (13)$$

то $\alpha_{s,0}(x,y) \geq 0$ і $\beta_{s,0}(x,y) \leq 0, \quad (x,y) \in \overline{D}_s$.

Теорема. Нехай функції $F_s[U_s(x,y)] \in C_2(\overline{B}_s)$, а $a_1(x,y) \in C^{(1.0)}(D),\ a_2(x,y) \in C^{(0.1)}(D)$ і виконуються умови (7).

Тоді послідовності функцій $\{z_{s,p}(x,y)\}$ та $\{v_{s,p}(x,y)\}$, побудовані згідно закону (10), (11), (12), де функції $q_{s,p}(x,y)$, $c_{s,p}(x,y)$ на кожному кроці ітерації вибираються таким чином, щоб виконувалися умови

$$\begin{aligned} &\overline{R}_s^p(x,y) - \overline{R}_{s,p-1}(x,y) - c_{s,p}(x,y)W_{s,p}(x,y) \geq (\leq) 0, \\ &\overline{R}_{s,p}(x,y) - \overline{R}_s^{p-1}(x,y) + q_{s,p}(x,y)W_{s,p}(x,y) \leq (\geq) 0, \end{aligned} \qquad (14)$$

$(x,y) \in \overline{D}_s, \quad s = 1,2,3$ при p — парних (непарних) і

$$v_{s,0}(x,y) \leq z_{s,1}(x,y),\ z_{s,0}(x,y) \geq v_{s,1}(x,y): \qquad (15)$$

1) збігаються рівномірно до єдиного неперервного розв'язку $U_s(x,y)$ відповідного інтегрального рівняння в (8) при $(x,y)\overline{D}_s$, $s = 1,2,3$;
2) в області $\overline{B}_{s,1}$ виконуються нерівності

$$v_{s,2p}(x,y) \le z_{s,2p+1}(x,y) \le v_{s,2p+2}(x,y) \le z_{s,2p+3}(x,y) \le U_s(x,y) \le$$
$$\le v_{s,2p+3}(x,y) \le z_{s,2p+2}(x,y) \le v_{s,2p+1}(x,y) \le z_{s,2p}(x,y)$$

для всіх $(x,y) \in \overline{D}_s$, $s = 1,2,3$;

3) справедливі оцінки

$$\max_s \sup_{\overline{D}_s} \left| W_{s,p}(x,y) \right| \le \frac{1}{p!} \left[k\, q\, \gamma\, l\, (x - x_0 + y - y_{0)} \right]^p \cdot d,$$

де

$$\max_s \sup_{\overline{D}_s} \left| W_{s,0}(x,y) \right| = d, \quad \max_s L_s = l,$$

$$\max_{s,p} \sup_{\overline{D}_s} \left(1 - q_{s,p}(x,y) - c_{s,p}(x,y) \right) \le q, \quad s = 1,2,3,$$

$$\max\{1, \sup_{\overline{D}} \left(y - y_0 + x - x_0 \right)\} = \gamma,$$

$$\max_s \sup_{\overline{D}_s \times \overline{D}_s} \left| K(x,y;\xi,\eta) \right| \le 0,5k;$$

4) збіжність ітераційного методу (10) — (15) не повільніша збіжності методу (11), коли всі $q_{s,p}(x,y) = c_{s,p}(x,y) = 0$ і, отже,

$$F_s^p(x,y) = f_s^p(x,y), \quad F_{s,p}(x,y) = f_{s,p}(x,y) \ .$$

Лема 5. Якщо функції $F_s[U_s(x,y)] \in C_2(\overline{B}_s)$ і виконуються нерівності (15), то множина функцій $q_{s,p}(x,y)$, $c_{s,p}(x,y)$, які задовольняють умови (10), (14) не порожня.

Для доведення леми 5 достатньо покласти

$$q_{s,p}(x,y) = \begin{cases} (\beta_{s,p}(x,y) + w_{s,p}(x,y)) \cdot \tau_{s,p}^{-1}(x,y), & w_{s,p}(x,y) \ne 0, \\ 0, & w_{s,p}(x,y) = 0 \end{cases}$$

$$c_{s,p}(x,y) = \begin{cases} (-\alpha_{s,p}(x,y) + w_{s,p}(x,y)) \cdot \tau_{s,p}^{-1}(x,y), & w_{s,p}(x,y) \ne 0, \\ 0, & w_{s,p}(x,y) = 0 \end{cases}$$

де $\tau_{s,p}(x,y) := \alpha_{s,p}(x,y) - \beta_{s,p}(x,y) + w_{s,p}(x,y)$.

Наслідок 1. Нехай виконуються умови теореми. Тоді в просторі функцій $C^*(\overline{D})$ існує єдиний іррегулярний розв'язок задачі (1)–(4). Якщо ж права частина рівняння (1) $f[U(x,y)] \in C(\overline{B})$ і виконуються умови $\rho_\kappa = 0, \kappa = 1,2$, то розв'язок крайової задачі (1)–(4) буде регулярним (тобто $U(x,y) \in C^{(1.1)}(D) \cap C(\overline{D})$).

Наслідок 2. Якщо виконуються умови теореми, то нерівності (12) є необхідними і достатніми, щоб виконувались умови (13).

Наслідок 3. Нехай виконуються умови теореми і $\psi(y) = \varphi(x) = 0, \quad (x, y) \in \overline{D}_1, \quad \omega_r(x) = 0, \quad x \in [x_{r-1}, x_r], \quad r = 1, 2,$ а $F_s[U_s(x, y)] \equiv H[U_s(x, y); 0].$

Тоді, якщо $F_s[0] \le (\ge) 0$ в області $\overline{B}_s$, то розв'язок крайової задачі (1)–(4) при $(x, y) \in \overline{D}$ задовольняє нерівності $U(x, y) \le (\ge) 0.$

Список використаних джерел:

1. Marynets V. V., Marynets K. V. On Goursat-Darboux boundary-value problem for systems of non-linear differential equations of hyperbolic type. *Miskolc Mathematical Notes*. 2013. Vol. 14, N 3. P. 1009–1020.
2. Маринець В. В., Маринець К. В., Питьовка О. Ю. Про одну крайову задачу теорії ДРЧП гіперболічного типу в області із складною структурою краю. *Наук. Вісник УжНУ*. Сер. матем. і інформ. 2014. 25, № 2. С.110–117.
3. Маринець В. В., Питьовка О. Ю. Про один підхід дослідження крайових задач для нелінійних рівнянь гіперболічного типу в області зі складною структурою краю. *Математичне та комп'ютерне моделювання.* Серія: Фіз.-мат. науки : зб. наук. праць ІК ім. В. М. Глушкова НАНУ. 2017. Вип. 15. С. 113–119.
4. Маринець В. В., Питьовка О. Ю. Дослідження крайової задачі для нелінійного хвильового рівняння з розривною правою частиною. *Наук. Вісник УжНУ*. Сер. матем. і інформ. 2018. Вип. № 1(32). С. 127–134.
5. Collatz L. Funktionalanalysis und numerische mathematic. Berlin ; Göttingen ; Heidelberg: Springeler-Verlag, 1964. P. 440.
6. Marynets V. V., Marynets K. V., Pytovka O. Yu. On one constructive method of the differential equations of the hyperbolic type. *Науковий вісник Ужгородського університету*. 2015. Вип. № 2 (27). P. 76–85.
7. Красносельский М. А., Вайникко Г. М., Забрейко П. П., Рутицкий Я. Б., Стеценко В. Я. Приближенное решение операторных уравнений. М. : Наука, 1969. 456 с.

ON ONE APPROACH OF INVESTIGATION OF THE BOUNDARY VALUE PROBLEM FOR A QUASILINEAR HYPERBOLIC TYPE EQUATION WITH DISCONTINUITIES IN THE RIGHT HAND — SIDE

We build a constructive method for investigation of the boundary value problem for the wave equation in the domain with complex structure of the boundary.

Key words: *«free» curves, irregular solution, comparison functions, consistency conditions.*

Одержано 20.01.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.78-84

М. О. Недашковський, д-р фіз.-мат. наук

Інститут механіки і прикладної інформатики Університету імені Казимира Великого, м. Бидгощ, Республіка Польща

ОБЧИСЛЕННЯ КОРТЕЖІВ РОЗВ'ЯЗКІВ МАТРИЧНИХ ПОЛІНОМІАЛЬНИХ РІВНЯНЬ

Наведені схеми для обчислення кортежів розв'язків матричних поліноміальних рівнянь *n*-го порядку**.**

Ключові слова: *матричні поліноміальні рівняння, кортежі розв'язків.*

Вступ. Розглядаються матричні рівняння загального вигляду

$$A_n X^n + A_{n-1} X^{n-1} + \ldots + A_1 X + A_0 = 0, \tag{1}$$

де $A_i \in R^{m\times m}$ $(i = \overline{0,n})$ – квадратні ненульові матриці порядку m із дійсними елементами, $X \in R^{m\times m}$ — невідома квадратна матриця.

Найпростіші матричні рівняння розв'язувались ще у другій половині XIX століття [1]. Проблема розкладу на множника глибоко проаналізвана в [2]. Більшість відомих обчислювальних схем описані в [1, 3]. Проте багато задач в цій області ще не розв'язано. Тут зупинимося на схемах обчислення кортежів розв'язків.

Матричні рівняння другого порядку. Розглянемо рівняння

$$AX^2 + BX + C = 0, \tag{2}$$

де A, B, C і $X \in R^{m\times m}$.

Після перегрупування його членів можна записати

$$X = -\left(B + AX\right)^{-1} C. \tag{3}$$

На основі (3) можна проводити наступну ітераційну процедуру

$$X_{(i)} = -\left(B + AX_{(i-1)}\right)^{-1} C. \tag{4}$$

З іншого боку композиція (3) дає наступне розвинення розв'язку (2) в матричний ланцюговий дріб

$$X = -\left(B - A\left(B - A\left(B - A\left(B - \ldots\right)^{-1} C\right)^{-1} C\right)^{-1} C\right)^{-1} C. \tag{5}$$

Існує й інша схема побудови ітераційного методу для обчислення X. У припущенні, що існують обернені матриці A^{-1}, X^{-1} можна для (2) записати рекурентну формулу

$$X = -A^{-1} \cdot B - A^{-1} \cdot C \cdot X^{-1}. \quad (6)$$

Рівність (6) може бути використана для ітераційного процесу

$$X_{(i)} = -A^{-1} \cdot B - A^{-1} \cdot C \cdot X_{(i-1)}^{-1}.$$

А композиція (6) дає інше розвинення X в неперервний дріб

$$X = -A^{-1}B - A^{-1}C(-A^{-1}B - A^{-1}C(-A^{-1}B - ...)^{-1})^{-1})^{-1}, \quad (7)$$

яке може бути збіжне до іншого розв'язку.

Симетричне квадратне рівняння 2-го порядку. Нехай тепер

$$AX + XB + XFX + C = 0, \quad (8)$$

де A, B, C, F і X є матрицями розміру $m \times m$.

Перегрупувавши його члени можна записати

$$X = -F^{-1}B + (A + XF)^{-1} \cdot (AF^{-1}B - C). \quad (9)$$

Із (9) отримуємо ітераційну формулу для обчислення X

$$X_{(i)} = -F^{-1}B + (A + X_{(i-1)}F)^{-1} \cdot (AF^{-1}B - C).$$

На основі (9) можна також подати розв'язок у вигляді МЛД

$$X = -F^{-1}B + (A + (-F^{-1}B + ...) \cdot F)^{-1}(AF^{-1}B - C)) \cdot F)^{-1} \cdot (AF^{-1}B - C). \quad (10)$$

З іншого боку, для рівняння (8) маємо

$$X = -AF^{-1} + (AF^{-1}B - C) \cdot (FX + B)^{-1}. \quad (11)$$

Із рівняння (11) можна також записати рекурентну формулу для обчислення X

$$X_{(i)} = -AF^{-1} + (AF^{-1}B - C) \cdot (FX_{(i-1)} + B)^{-1}.$$

На основі (11) отримуємо ще один розв'язок рівняння у вигляді МЛД

$$X = -AF^{-1} + (AF^{-1}B - C) \cdot (F \cdot (-AF^{-1} +$$
$$+(AF^{-1}B - C) \cdot (F \cdot (-AF^{-1} + ...) + B)^{-1}) + B)^{-1}.$$

Дискретне рівняння Ріккаті. Розглянемо тепер відоме в застосуваннях [1] рівняння виду:

$$A^T XA - X - A^T XB(R + B^T XB)^{-1} B^T XA + Q = 0; \quad (12)$$

де A, B, C, R, Q і X — матриці розмірністю $m \times m$.

В результаті перетворень для рівняння (12) можна записати

$$X = Q + A^T \left(A^{-1}BR^{-1}B^T + A^{-1}X^{-1} \right)^{-1}. \quad (13)$$

Цей вираз можна використати для ітераційного обчислення X:

$$X_{(i)} = Q + A^T \left(A^{-1}BR^{-1}B^T + A^{-1}X_{(i-1)}^{-1} \right)^{-1}. \quad (14)$$

На основі (14) отримуємо розв'язок у вигляді двоперіодичного МЛД

$$X = Q + A^T (A^{-1}BR^{-1}B^T + A^{-1}(Q + A^T (A^{-1}BR^{-1}B^T +$$
$$+A^{-1}(Q + A^T (A^{-1}BR^{-1}B^T + ...)^{-1})^{-1})^{-1})^{-1})^{-1}. \quad (15)$$

З іншого боку для рівняння (12) можна одержати:

$$X = Q + [B^T (A^T)^{-1} + RB^{-1} X^{-1} (A^T)^{-1}]^{-1} RB^{-1} A. \qquad (16)$$

На підставі (16) маємо ітераційну формулу для обчислення розв'язку

$$X_{(i)} = Q + [B^T (A^T)^{-1} + RB^{-1} X_{(i-1)}^{-1} (A^T)^{-1}]^{-1} RB^{-1} A.$$

Із (16) отримується інше розвинення розв'язку у двоперіодичній МЛД

$$\begin{aligned} X = Q + (B^T (A^T)^{-1} + RB^{-1}(Q + (B^T (A^T)^{-1} + \\ + RB^{-1}(Q + ...)^{-1} (A^T)^{-1})^{-1} RB^{-1} A)^{-1} (A^T)^{-1})^{-1} RB^{-1} A. \end{aligned} \qquad (17)$$

Таким чином, для кожного з розглянутих рівнянь (2), (8), (12) може бути побудовані свої особливі схеми розвинення розв'язків у МЛД, які дають на практиці кортежі розв'язків.

Матричні рівняння n -го порядку. Виявляється, у цьому випадку можна побудувати деяку загальну схему. Нехай маємо рівняння

$$X^n + A_{n-1} X^{n-1} + A_{n-2} X^{n-2} + ... + A_1 X + A_0 = 0, \qquad (18)$$

де матриці $A_i \in \mathrm{R}^{m \times m}$ $(i = \overline{0, n-1})$, $X \in R^{m \times m}$, а $n \geq 2$ — ціле число.

Виявляється, що розв'язок рівняння (18) може бути у вигляді одноперіодичного гіллястого МЛД із $n-1$ вітками розгалуження

$$X = P_0 + \sum_{k=1}^{n-1} P_k (P_0 + \sum_{k=1}^{n-1} P_k (P_0 + \sum_{k=1}^{n-1} P_k (... - Q_k)^{-1} X - Q_k)^{-1} - Q_k)^{-1}. \qquad (19)$$

Він отримується композицією дробово-лінійних виразів вигляду

$$X = P_0 + \sum_{k=1}^{n-1} P_k (X - Q_k)^{-1}, \qquad (20)$$

де $P_k \in R^{m \times m}$ та $Q_k \in R^{m \times m} (k = \overline{0, n-1})$ — квадратні матриці, елементи яких $p_{i,j,k} (i, j = \overline{1, m})$ та $q_{i,j,k} (i, j = \overline{1, m}; k = \overline{1, n-1})$ визначаються із системи рівнянь

$$(-1)^{n-1} Q_1 Q_2 ... Q_{n-1} + P_0 = A_1,$$

$$(-1)^{n-2} \sum_{k=1}^{n-1} \prod_{l=1}^{k-1} Q_l \prod_{l=k+1}^{n-1} Q_l - \sum_{k=1}^{n-1} P_k + (-1)^{n-1} P_0 Q_1 Q_2 ... Q_{n-1} = A_2;$$

$$(-1)^{n-2} \sum_{k=2}^{n-1} \sum_{l=k+1}^{n-2} (1 - \delta_{kl}) \prod_{r=1}^{k-1} Q_k \prod_{r=k+1}^{l-1} Q_r \prod_{r=l+1}^{n-2} Q_r +$$

$$+ \sum_{k=1}^{n-1} P_k \prod_{r=1}^{k-1} Q_r \prod_{r=k+1}^{n-1} Q_r + (-1)^{n-1} \sum_{k=1}^{n-1} \prod_{r=1}^{k-1} P_0 Q_r = A_3;$$

. . . . . .

$$\sum_{k=1}^{n-1} Q_k + \sum_{k=2}^{n-1}\sum_{l=k+1}^{n-1} P_1 Q_k Q_l + ... + \sum_{k=1}^{n-1}\sum_{l=k+1}^{n-1} (1-\delta_{kr}) P_r Q_k Q_l +$$

$$+ ... + (-1)^{n-1} \sum_{k=1}^{n-1}\prod_{r=1}^{k-1} P_{n-1} Q_k Q_l = A_{n-1};$$

$$\sum_{k=1}^{n-1} P_1 Q_k + ... + \sum_{k=1}^{n-1} (1-\delta_{kr}) P_r Q_r + ... + \sum_{k=1}^{n-2} P_{n-1} Q_k + \sum_{k=1}^{n-1} P_0 Q_k = A_0.$$

Якщо покласти $Q_k = q_k \cdot E$ $(k = \overline{1, n-1})$ і надати усім q_k попарно різні конкретні числові значення, то остання система n матричних рівнянь стане лінійною відносно невідомих P_i для всіх $(i = \overline{0, n-1})$ і буде мати єдиний розв'язок. Для обчислення числового значення розв'язку X на ОС рекурентна формула (20) переписується у вигляді

$$X_{(i)} = P_0 + \sum_{k=1}^{n-1} P_k (X_{(i-1)} - Q_k)^{-1}. \qquad (21)$$

Неканонічне матричне поліноміальне рівняння n-го порядку. Введемо тепер до розгляду

$$X^n + X^{n-1} A_{n-1} + X^{n-2} A_{n-2} + ... + XA_1 + A_0 = 0, \qquad (22)$$

де матриці $A_i \in \mathrm{R}^{m\times m}$ $(i = \overline{0, n-1})$, $X \in R^{m\times m}$, а $n \geq 2$ — ціле число.

Для рівняння (22) розв'язок теж може бути записаний у вигляді одноперіодичного гіллястого МЛД із $n-1$ вітками розгалуження

$$X = P_0 + \sum_{k=1}^{n-1} (P_0 - Q_k + ... + \sum_{k_s=1}^{n-1} (P_0 - Q_{k_s} + ... P_{k_s})^{-1} P_{k_1})^{-1} P_k.$$

Він отримується композицією дробово-лінійних виразів вигляду

$$X = P_0 + \sum_{k=1}^{n-1} (X - Q_k)^{-1} P_k, \qquad (23)$$

де $P_k \in R^{m\times m}$ та $Q_k \in R^{m\times m}$ $(k = \overline{0, n-1})$ — квадратні матриці, елементи яких $p_{i,j,k}$ $(i, j = \overline{1, m})$ та $q_{i,j,k}$ $(i, j = \overline{1, m}; k = \overline{1, n-1})$ визначаються із системи рівнянь:

$$(-1)^{n-1} Q_1 Q_2 ... Q_{n-1} + P_0 = A_1;$$

$$(-1)^{n-2} \sum_{k=1}^{n-1}\prod_{l=1}^{k-1} Q_l \prod_{l=k+1}^{n-1} Q_l - \sum_{k=1}^{n-1} P_k + (-1)^{n-1} Q_1 Q_2 ... Q_{n-1} P_0 = A_2;$$

$$(-1)^{n-2} \sum_{k=2}^{n-1}\sum_{l=k+1}^{n-2} (1-\delta_{kl}) \prod_{r=1}^{k-1} Q_k \prod_{r=k+1}^{l-1} Q_r \prod_{r=l+1}^{n-2} Q_r +$$

$$+\sum_{k=1}^{n-1}\prod_{r=1}^{k-1}Q_r\prod_{r=k+1}^{n-1}Q_rP_k+(-1)^{n-1}\sum_{k=1}^{n-1}\prod_{r=1}^{k-1}Q_rP_0=A_3;$$

. . ..

$$\sum_{k=1}^{n-1}Q_k+\sum_{k=2}^{n-1}\sum_{l=k+1}^{n-1}Q_kQ_lP+...+\sum_{k=1}^{n-1}\sum_{l=k+1}^{n-1}(1-\delta_{kr})Q_kQ_lP_r+$$

$$+....+(-1)^{n-1}\sum_{k=1}^{n-1}\prod_{r=1}^{k-1}Q_kQ_lP_{n-1}=A_{n-1},$$

$$\sum_{k=1}^{n-1}Q_kP_1+....+\sum_{k=1}^{n-1}(1-\delta_{kr})Q_rP_r+...+\sum_{k=1}^{n-2}Q_kP_{n-1}+\sum_{k=1}^{n-1}Q_kP_0=A_0.$$

Якщо покласти $Q_k = q_k \cdot E \quad (k=\overline{1,n-1})$ і надати усім q_k попарно різні конкретні числові значення, то остання система n матричних рівнянь стане лінійною відносно невідомих P_i для всіх $(i=\overline{0,n-1})$ і буде мати єдиний розв'язок. Для обчислення розв'язку X на ОС рекурентну формулу (23) слід переписати у вигляді:

$$X_{(i)} = P_0 + \sum_{k=1}^{n-1}(X_{(i-1)} - Q_k)^{-1}P_k. \qquad (24)$$

Для неформального розв'язання і дослідження усіх розглянутих рівнянь треба додатково проводити дослідження збіжності до розв'язку і стійкості відповідних гіллястих МЛД.

Умови закінчення ітераційного процесу при розв'язанні матричних рівнянь гіллястими ланцюговими дробами. Тепер розглянемо ознаки збіжності ітераційного процесу до розв'язку та обґрунтування критеріїв закінчення обчислення розв'язків матричних рівнянь із застосуванням апарату гіллястих МЛД. Якщо співставити між собою вирази (16), (20) і (23), то неважко зауважити, що всі матричні ланцюгові дроби ними утворені є частковим випадком наступного закону композиції

$$X_{(i)} = P_0 + \sum_{k=1}^{n-1}P_k(X_{(i-1)} - Q_k)^{-1}R_k. \qquad (25)$$

Власне самі ознаки збіжності для гіллястих МЛД вже подані [4], але важливе ще і одержання надійних критеріїв закінчення ітерацій та збіжності процесу саме до розв'язку конкретного рівняння.

Ознаки збіжності матричних гіллястих ланцюгових дробів до розв'язків. Припускається, що розв'язок цього рівняння на деякому інтервалі знаходиться за ітераційною процедурою вигляду

$$X_{(i+1)} = P_0 + \sum_{k=1}^{n-1}P_k(-Q_k + X_{(i)})^{-1}R_k. \qquad (25)$$

Матричні елементи P_0, P_k, Q_k, R_k $(k = \overline{1, n-1})$ визначаються із системи лінійних рівнянь складеної із коефіцієнтів даного рівняння. Тоді розв'язок $X = \lim\limits_{i\to\infty} X_{(i)}$ може бути поданий як розвинення у нескінчений одноперіодичний матричний ланцюговий дріб

$$X = P_0 + \sum_{k=1}^{n-1} P_k (P_0 - Q_k + \sum_{k=1}^{n-1} P_k (P_0 - Q_k + ...R_k)^{-1} R_k)^{-1} R_k . \qquad (26)$$

Звичайно, цей дріб буде збіжним, і тим більше, збіжним до розв'язу лише за певних умов. Для вивчення проблеми збіжності до розв'язку подібних розвинень, розглянемо далі неканонічний МГЛД

$$D = b_0 + \underset{s=1}{\overset{\infty}{D}} \sum_{k_{(s)}=1}^{n} a_{k_{(s)}} b_{k_{(s)}}^{-1} c_{k_{(s)}} \qquad (27)$$

та сформулюємо критерії для прийняття рішення про закінчення ітерацій та збіжності процесу саме до розв'язку конкретних рівнянь.

Теорема 1. Якщо в інтервалі $[-n, n]$ існує, причому лише один розв'язок поліноміального матричного рівняння то його розвинення (26) за ітераційною процедурою (25) у МГЛД з елементами, що задовольняють умовам [4]

$$\left\| b_{k(s)}^{-1} \right\| \le \frac{1}{\left\| a_{k(s)} \right\| \left\| c_{k(s)} \right\| + n} (1 \le k_s \le n; s = 1, 2, 3, ...) \qquad (28)$$

збігається до цього розв'язку.

Теорема 2. Якщо в інтервалі $\left[-\sum\limits_{k_1=1}^{n} \left\| a_{k_1} \right\| \left\| c_{k_1} \right\|, \sum\limits_{k_1=1}^{n} \left\| a_{k_1} \right\| \left\| c_{k_1} \right\| \right]$ існує, причому лише один розв'язок поліноміального матричного рівняння, то його розвинення (26) за ітераційною процедурою (25) у МГЛД з елементами, що задовольняють умовам [4]

$$\left\| b_{k_{(s)}}^{-1} \right\| \le \frac{1}{1 + \sum\limits_{k_{s+1}=1}^{n} \left\| a_{k_{(s+1)}} \right\| \left\| c_{k_{(s+1)}} \right\|}, \ (s = 1, 2, 3...) \qquad (29)$$

збігається до цього розв'язку.

Слід зауважити, що на практиці ітераційні процеси обчисленя $X_{(k)}$ нерідко є збіжними при значно менш жорстких умовах. Так в одному з тестових рівнянь 3-го порядку виду (22) $\left\| Q_0^{-1} \right\| = 0.09671109$ в той же час $1/(1 + \left\| P_1 \right\| + \left\| P_2 \right\|) = 0,003842229$. Тобто, умова збіжності

не виконується навіть наближено. Але процес обчислення розв'язку все одно достатньо швидко збігається і для різних наборів Q_i $(i = \overline{1,n})$ вдається обчислити кортеж із 4-х розв'язків.

Висновки. Отже, наведені підходи для обчислення кортежів розв'язків матричних поліноміальних рівнянь, сформульовано достатні критерії закінчення ітераційних процесів та збіжності їх до розв'язку.

Список використаних джерел:

1. Икрамов Х. Д. Численное решение матричных уравнений. М. : Наука, 1984. 192 с.
2. Казимірський П. С. Розклад матричних многочленів на множники. Київ : Наук. думка, 1983. 247 с.
3. Кублановская В. Н. К спектральной задаче для полиномиальных пучков матриц. *Зап. науч. семинаров Ленигр. отд. Мат. ин-та АН СССР*, 1978. Т. 80. С. 83–97.
4. Недашковський М. О. Ознаки збіжності матричних гіллястих ланцюгових дробів. *Математичні методи та фізико-механічні поля.* Львів, 2003. Т. 46, № 4. С. 50–56.

CALCULATION OF CORTEGES OF SOLUTIONS OF MATRIX POLYNOMIAL EQUATIONS

The schemes are presented for calculating the corteges of solutions of matrix polynomial equations.

Key words: *matrix polynomial equations, corteges of solutions.*

Одержано 21.01.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.85-91

А. Н. Нестеренко,
О. В. Попов, канд. фіз.-мат. наук,
О. В. Рудич
Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

РОЗВ'ЯЗУВАННЯ СИСТЕМ НЕЛІНІЙНИХ РІВНЯНЬ НА КОМП'ЮТЕРАХ З ПАРАЛЕЛЬНОЮ ОРГАНІЗАЦІЄЮ ОБЧИСЛЕНЬ

Запропоновано методологію розв'язування систем нелінійних рівнянь з розрідженими матрицями Якобі на комп'ютерах з паралельною організацією обчислень, яка використовує багаторівневу модель паралельних обчислень, структурну регуляризацію та декомпозицію розріджених даних для приведення матриці системи до блочно-розрідженого вигляду, високопродуктивні блочні та блочно-циклічні алгоритми розв'язування систем лінійних рівнянь.

Ключові слова: *системи нелінійних рівнянь, розріджені матриці Якобі, комп'ютери MIMD-архітектури, комп'ютери гібридної архітектури, структурна регуляризація розріджених даних, декомпозиція розріджених даних.*

Вступ. При чисельному моделюванні природних явищ, поведінки об'єктів під впливом дії навколишнього середовища, проектуванні будівель та механізмів часто виникають, наприклад, при використанні тривимірних моделей, розрахункові (дискретні) задачі з надвеликою кількістю (яка може перевищувати 10^7) рівнянь, у тому числі нелінійних. Причому дані (матриці Якобі) таких нелінійних систем (СНР) мають розріджену структуру, наприклад, блочно-тридіагональну або блочно-п'ятидіагональну. Тобто кількість ненульових елементів значно менша (приблизно дорівнює kn, де n — порядок матриці, а $k << n$) загальної кількості елементів матриці.

Зростання параметрів задач, що розв'язуються, розрахунок на комп'ютерах більш повних моделей об'єктів, процесів, явищ вимагає відповідного зростання продуктивності комп'ютерів. В даний час зростання продуктивності обчислень досягається за рахунок розпаралелювання, яке базується на використанні комп'ютерів з багатьма процесорними пристроями, зокрема з багатоядерними процесорами. В цих комп'ютерах, як правило, реалізується MIMD-архітектура (архітектура з множинним потоком команд і даних). В останні роки також набули поширення гібридні обчислювальні системи, в яких використовуються співпроцесори, наприклад, графічні процесори (GPU), для прискорення

обчислень при виконанні великих обсягів однорідних арифметичних операцій. На таких співпроцесорах-прискорювачах, як правило, реалізується SIMD-архітектура паралельних обчислень. Такі комп'ютери гібридної архітектури вже зайняли провідні позиції у світовому рейтингу найпродуктивніших комп'ютерів TOP500 [1].

У цій роботі розглядається використання багаторівневої моделі паралельних обчислень для розв'язування систем нелінійних рівнянь на комп'ютерах гібридної архітектури та комп'ютерах з багатоядерними процесорами Intel Xeon Phi серії x200.

Багаторівнева модель паралельних обчислень. Архітектура сучасних високопродуктивних комп'ютерів надає можливість використовувати багаторівневу модель паралельних обчислень — багаторівневий паралелізм:

- верхній рівень (MIMD-модель) — паралелізм процесів (process level parallelism, PLP) — процеси паралельно виконують макрооперації (підзадачі), наприклад, множення матричних блоків, використовуючи як розподілену між ними, так і спільну пам'ять і синхронізуючи обчислення та обміни даними;
- другий рівень (SIMD-модель) — паралелізм потоків (thread level parallelism, TLP) — розпаралелення виконання кожної з макрооперацій, використовуючи декілька потоків і спільну пам'ять;
- третій рівень (векторизація) паралелізм обробки даних векторними процесорними пристроями (data level parallelism, DLP) — паралельно виконуються операції з векторами, наприклад, додавання векторів.

На верхньому рівні використовуються (як правило) засоби MPI, на другому — засоби (директиви) OpenMP (Open Multi-Processing) або програмні модулі багатопотокової бібліотеки Intel MKL. Третій рівень — автоматичне включення паралелізму при компіляції програми.

Постановка задачі. В області $D = \{a_i \le x_i \le b_i,\ i = 1, 2, \ldots, n\}$ знайти n-вимірний вектор $x = (x_1, x_2, \ldots, x_n)^T \in D$, який задовольняє системі n нелінійних рівнянь

$$F(x) = 0, \qquad (1)$$

де $F(x) = (F_1(x), F_2(x), \ldots, F_n(x))^T$ — n-вимірна вектор-функція, причому $F(x)$ є наближенням до точної вектор-функції $\Phi(x)$ і для цих функцій виконується нерівність $\|F(x) - \Phi(x)\| \le \delta$ для будь-якого $x \in D$. Для розв'язування задачі (1) задаються початкове наближення $x^{(0)} \in D$ і необхідна точність ε отримання наближення до розв'язку системи.

У прикладних галузях, зокрема в розрахунку міцності конструкцій (див., напр., [2]) використовують наступну постановку нелінійної

задачі. Математично статична нелінійна задача розрахунку міцності конструкцій, використовуючи принцип можливих переміщень, може бути поставлена у нескінченновимірному функціональному просторі можливих переміщень U_0 у вигляді варіаційної задачі: знайти вектор-функцію $u \in U_0$, яка для будь-якої вектор-функції $v \in U_0$ задовольняє відповідній інтегральній тотожності

$$a(u,v) = l(f,v); \qquad (2)$$

де нелінійний по u і лінійний по v функціонал $a(u,v)$ пропорційний потенційній енергії деформації, а лінійний по v функціонал $l(f,v)$ пропорційний роботі прикладених зусиль f при навантаженні.

Розв'язки нелінійних задач (2) знаходяться одним з проекційно-варіаційних методів, переважно методом скінченних елементів (МСЕ). Наближені розв'язки МСЕ шукаються у скінченновимірному підпросторі $U_0^h \subset U_0$. Вектор-функції з підпростору U_0^h є кусково-поліноміальними і можуть бути представлені у вигляді лінійної комбінації базисних вектор-функцій $u_h(\chi) = \sum_{j=1}^{n} x_j \varphi_j(\chi)$, де φ_j ($j = 1, 2, \ldots, n$) — заданий вище кусково-поліноміальний базис U_0^h. Підставивши в (2) вектор-функції з підпростору U_0^h отримуємо систему нелінійних (відносно x_j) рівнянь

$$a(u_h, \varphi_i) = l(f, \varphi_i), \quad i = 1, 2, \ldots, n. \qquad (3)$$

Методи розв'язування систем нелінійних рівнянь. У багатьох прикладних застосуваннях реалізуються ітераційні методи розв'язування СНР (1) або (3), які базуються (див., напр., [2, 3]) тією чи іншою мірою на класичному методі Ньютона, що має квадратичну швидкість збіжності. Ітераційний процес збігається, якщо виконується оцінка

$$\| x^{(k)} - x \| \le c \| x^{(k-1)} - x \|^{\alpha},$$

де c — деяка величина, обмежена зверху; α — порядок збіжності методу. Якщо $\alpha = 2$, то досягається квадратична швидкість збіжності ітераційного процесу, якщо $1 < \alpha < 2$, то ітераційний процес збігається надлінійно.

Позначимо $H(x) = \left\{ \frac{\partial F_i}{\partial x_j} \right\}_{i,j=1}^{n}$ — матриця Якобі системи (1) або (3), $B(x)$ — деяке наближення до $H(x)$. В прикладних застосуваннях для обчислення наближеної матриці Якобі системи (3) часто використовують похідну функціонала $a(u,v)$ у наступному вигляді [2]:

$$a'(u,v,w) = \frac{d}{d\tau} a(u + \tau w, v)\Big|_{\tau=0}, \quad w \in U_0. \quad \text{Тоді} \quad a'(u_h, v_h, w_h) = H(x)y$$

(якщо $y = (y_1, y_2, \ldots, y_n)^T$, $u_h = \sum_{j=1}^{n} x_j \varphi_j$, $w_h = \sum_{j=1}^{n} y_j \varphi_j$, а $v_h = \sum_{j=1}^{n} \varphi_j$), а елементи матриці Якобі системи (3) можна обчислити за формулою

$$H(x) \equiv \{h_{ij}\}_{i.j=1}^{n} = \{a'(u_h, \varphi_i, \varphi_j)\}_{i.j=1}^{n}.$$

Ітераційний процес методу Ньютона при заданому початковому наближенні записується так ($k = 1, 2, \ldots$ — номер ітерації, $F^{(k)} \equiv F(x^{(k)})$ і $B^{(k)} \equiv H^{(k)} \equiv H(x^{(k)})$)

$$B^{(k-1)} w^{(k)} = -F^{(k-1)}, \tag{4}$$

$$x^{(k)} = x^{(k-1)} + w^{(k)}. \tag{5}$$

У багатьох випадках для розв'язування СНР використовують модифікації методу Ньютона, які називають квазіньютонівськими (ці методи мають надлінійну швидкість збіжності, див. [3]), в тому числі:

- методи Бройдена і Пауелла (для симетричних матриць Якобі), у яких у ході ітераційного процесу (4), (5) обчислені на основі початкового наближення $B^{(0)} \equiv H^{(0)}$ матриці уточнюється з використанням матрично-векторних операцій за формулами

$$B^{(k)} = B^{(k-1)} + \frac{y^{(k)}(w^{(k)})^T + w^{(k)}(y^{(k)})^T}{(w^{(k)})^T w^{(k)}} - \frac{((y^{(k)})^T w^{(k)}) w^{(k)} (w^{(k)})^T}{((w^{(k)})^T w^{(k)})^2}$$

(в методі Пауелла) та $B^{(k)} = B^{(k-1)} + \frac{y^{(k)}(w^{(k)})^T}{(w^{(k)})^T w^{(k)}}$ (в методі Бройдена);

тут $y^{(k)} = F^{(k)} - F^{(k-1)} - B^{(k-1)} w^{(k)}$;

- метод Бурдакова, який при спеціальному виборі ітераційного параметра забезпечує глобальну збіжність до одного з розв'язків системи на основі заданого початкового наближення; ітераційний процес реалізується за формулами (4) та $x^{(k)} = x^{(k-1)} + \alpha_k w^{(k)}$, де $B^{(k)}$ — скінченно-різницева апроксимація матриці Якобі $H^{(k)}$, а ітераційний параметр α_k обчислюється спеціальним чином (див. [3]).

Отже, розв'язування СНР з розрідженою структурою даних, як правило, базується на лінеаризації нелінійних рівнянь — пошук розв'язків реалізується через розв'язування послідовності систем лінійних рівнянь. Ці СЛАР (4) мають декілька особливостей, які впливають на вибір методів та засобів для їх розв'язування на комп'ютерах з паралельною організацією обчислень, а саме:

- високий порядок — від 100 000 до десятків мільйонів;
- розріджена структура матриць СЛАР — стрічкова, профільна, блочно-розріджена тощо;
- симетричність або несиметричність матриць СЛАР;

- додатно визначеність або напіввизначеність матриць СЛАР.

Паралельні алгоритми розв'язування СНР з розрідженою структурою даних. Ітераційні процеси викладених вище методів розв'язування СНР (1) або з розрідженою структурою даних в загальному випадку можна записати у вигляді (k = 1, 2, ...) (4) та (5) або $x^{(k)} = x^{(k-1)} + \alpha_k w^{(k)}$ в методі Бурдакова.

Структурна регуляризація розріджених матриць. Отже основною операцією на кожній ітерації є розв'язування СЛАР (4) з розрідженою матрицею $B^{(k-1)}$. Структура розріджених матриць визначається нумерацією невідомих і може бути регулярною (наприклад, стрічковою) або нерегулярною.

З метою зменшення кількості арифметичних операцій для розв'язування СЛАР (4) з розрідженою матрицею шляхом структурної регуляризації — перестановки рядків і стовпчиків (тобто перенумерації невідомих) таку матрицю приводять до одного із стандартних виглядів: стрічкового, профільного, блочно-діагонального з обрамленням, «хмарочосного» тощо. Існує декілька алгоритмів [4] оптимізації структури розрідженої матриці (фактор-дерев, Катхілл–Маккі, паралельних перерізів, мінімальної степені тощо).

Багаторівневий паралелізм передбачає використання блочних та блочно-циклічних алгоритмів на основі блочного представлення матриць. Тому доцільно структурно регуляризувати розріджену матрицю — оптимізувати блочно-розріджену структуру матриці, визначивши в блочному розбитті нульові блоки та якомога більше заповнені ненульові блоки і використавши один з названих вище алгоритмів (причому замість елементів матриці в алгоритмах використовуються блоки) [5].

Декомпозиція та розподіл між процесорними пристроями розріджених даних. Розподіл між процесорними пристроями даних СНР визначається розподілом даних для розв'язування СЛАР (4). Достатньо добру збалансованість завантаження процесів забезпечують паралельні версії алгоритмів прямих методів розв'язування СЛАР, в яких використовуються так звані циклічні схеми розподілу і обробки матриць (див. напр. [3]). У випадках стрічкових, профільних та хмарочосних матриць паралельні алгоритми, в яких використовується одновимірні блочно-циклічні схеми розподілу елементів матриць, дозволяють досягти приблизно рівного обсягу обчислень і обмінів, що виконуються кожним паралельним процесом в кожний момент часу. У випадку блочно-діагональної матриці з обрамленням використовуються блочні алгоритми та відповідний блочний розподіл елементів матриці [5].

Методологія розв'язування СНР. Отже, пропонується наступна послідовність дій для розв'язування СНР з розрідженими даними на сучасних високопродуктивних комп'ютерах, в тому числі гібридної архітектури:

- використовуючи один з алгоритмів структурної регуляризації, формування блочно-розрідженої структури матриць СЛАР (4) на основі вихідної структури її ненульових елементів;
- декомпозиція розріджених матриць та розподіл отриманих рядків або стовпчиків ненульових блоків між процесорними пристроями;
- ітераційний процес розв'язування СНР — на k-й ітерації ($k = 1, 2, \dots$) алгоритму методу Ньютона або квазіньютонівського виконуються наступні макрооперації:
  1) обчислення розподілених між MPI-процесами компонент вектор-функції $F^{(k-1)}$ та елементів ненульових блоків матриці $B^{(k-1)}$;
  2) розв'язування отриманої СЛАР (4), використовуючи відповідний (до структури матриці $B^{(k-1)}$) паралельний алгоритм [3, 5–7], та обчислення розподілених між MPI-процесами компонент наступного наближення до розв'язку СНР $x^{(k)}$;
  3) перевірка умов закінчення ітераційного процесу за формулами [7]: з початку $\| F^{(k)} \| \le \varepsilon$, а далі — $\| (H^{(k)})^{-1} \| \| F^{(k)} \| \le \varepsilon$.

Ці макрооперації виконуються на верхньому рівні паралелізму з використанням нижніх рівнів для виконання великих обсягів однорідних обчислень, у тому числі матрично-векторних та матрично-матричних операцій.

Висновки. Використовуючи багаторівневу модель паралельних обчислень з урахуванням особливостей архітектури комп'ютера розроблено ефективні алгоритми та програми розв'язування СНР на паралельних комп'ютерах гібридної архітектури та з процесорами Intel Xeon Phi серії x200. Це алгоритмічно-програмне забезпечення використано для розв'язування низки задач прогнозування ресурсу відповідальних зварних конструкцій [5]. При цьому час розв'язування задач суттєво скорочується, що дає можливість розв'язувати задачі високих порядків у реальному часі, які висуває сучасне життя перед наукою.

Список використаних джерел:

1. URL: http://www.top500.org
2. Городецкий А. С., Евзеров И. Д. Компьютерные модели конструкций. Киев : ФАКТ. 2007. 394 с.
3. Химич А. Н., Молчанов И. Н., Попов А. В. и др. Параллельные алгоритмы решения задач вычислительной математики. Киев : Наук. думка, 2008. 248 с.
4. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. М. : Мир, 1984. 334 с.
5. Velikoivanenko E. A., Milenin A. S., Popov A. V. at other. Methods of Numerical Forecasting of Serviceability of Welded Structures on Computers of Hybrid Architecture. *Cybernetics and Systems Analysis.* 2019. Vol. 53, N 1, January. P. 117–127.
6. Химич А. Н., Попов А. В., Чистяков А. В. Гибридные алгоритмы решения алгебраической проблемы собственных значений с разреженными матрицами. *Кибернетика и системный анализ.* 2017. Т. 53, № 6. С. 132–146.

7. Нестеренко А. Н., Химич А. Н., Яковлев М. Ф. Некоторые вопросы решения систем нелинейных уравнений на многопроцессорных вычислительных системах с распределенной памятью. *Вестник компьютерных и информационных технологий.* М., 2006. № 10. С. 54–56.

SOLVING OF THE SYSTEMS OF NON-LINEAR EQUATIONS ON COMPUTERS WITH PARALLEL ORGANIZATION OF CALCULATIONS

The methodology for the solving of non-linear systems with sparse Jacobi matrices on parallel computers is proposed, which uses a multilevel parallel computing model, structural regularization and decomposition of sparse data for reducing system's matrix to the block-sparse form, high-performance block and block-cyclic algorithms for solving systems of linear equations.

Key words: *systems of non-linear equations, sparse Jacobi matrices, MIMD-architecture computers, hybrid-architecture computers, structural regularization of the sparse data, decomposition of the sparse data.*

Одержано 15.02.2019

УДК 519.9
DOI: 10.32626/2308-5878.2019-19.91-97

О. П. Нечуйвітер, д-р фіз.-мат. наук,
Г. В. Каргапольцева, здобувач,
К. В. Дараган, аспірантка

Українська інженерно-педагогічна академія, м. Харків

ОПТИМАЛЬНА ЗА ПОРЯДКОМ ТОЧНОСТІ КУБАТУРНА ФОРМУЛА НАБЛИЖЕНОГО ОБЧИСЛЕННЯ ПОДВІЙНОГО ІНТЕГРАЛУ ВІД ШВИДКООСЦИЛЮЮЧИХ ФУНКЦІЙ ЗАГАЛЬНОГО ВИДУ

Розглядається оптимальна за порядком точності кубатурна формула наближеного обчислення подвійного інтегралу від швидкоосцилюючих функцій загального виду на класі диференційовних функцій у випадку, коли інформація про функції задана їх слідами на відповідних лініях.

Ключові слова: *кубатурна формули, інтеграли від швидкоосцилюючих функцій, клас диференційовних функцій.*

Вступ. Задача наближеного обчислення інтегралів від швидкоосцилюючих функцій двох змінних загального виду

$$I(f,g,\omega)=\int_0^1\int_0^1 f(x,y)\,\mathrm{e}^{i\omega g(x,y)}\,dxdy, \qquad (1)$$

має як класичне розв'язання [1], так і у випадку різних інформаційних операторів [2, 3]. Однак, не дослідженим залишилось питання побудови оптимальних за порядком точності кубатурних формул, у випадку, коли інформація про $f(x,y)$ та $g(x,y)$ задана відповідними їх слідами на лініях. Дана стаття присвячена актуальному питанню: побудові оптимальної за порядком точності кубатурної формули обчислення інтегралу виду (1) на класі диференційовних функцій.

Оптимальна за порядком точності кубатурна формула. Припустимо, що $f(x,y)\in F$, $g(x,y)\in G$, F, G — множини функцій, визначених в області $[a,b]\times[a,b]$. Позначимо L_N множину всіх квадратурних формул $l_N(f,g)$, що використовують інформацію про значення функцій $f(x,y)$ та $g(x,y)$ не більше ніж на N лініях. Введемо величини $R_N(f,g,\omega,l_N)=\left|I(f,g,\omega)-l_N(f,g)\right|$,

$$R_N(F,G,\omega,l_N)=\sup_{f\in F,\, g\in G} R_N(f,g,\omega,l_N),$$

$$R_N(F,G,\omega)=\inf_{l_N\in L_N} R_N(F,G,\omega,l_N).$$

Кубатурну формулу $l_N^*(f,g)$, на якій досягається $R_N(F,G,\omega)$, будемо називати оптимальною за точністю кубатурною формулою. Якщо $R_N\left(F,G,\omega,\bar{l}_N\right)\le R_N(F,G,\omega)+\eta$, $\eta>0$, то $\bar{l}_N$ називається оптимальною за точністю формулою обчислення $I(f,g,\omega)$ з точністю до η. Якщо $\eta=o(R_N)$ або $\eta=O(R_N)$, то $\bar{l}_N$ називається асимптотично оптимальною або оптимальною за порядком точності.

Розглянемо $H^{2,r}(M,M)$ — клас дійсних функцій, визначених на $G=[0,1]^2$ і таких, що частинні похідні порядку r по змінній x та y обмежені, тобто

$$\left|f^{(r,0)}(x,y)\right|\le M,\ \left|f^{(0,r)}(x,y)\right|\le M,\ r\ne 0,\ \left|f^{(r,r)}(x,y)\right|\le M,\quad r\ge 0.$$

Теорема 1 [3]. Нехай $f(x,y),\ g(x,y),\in H^{2,r}\left(M,M\right)$, функції $f(x,y)$, $g(\mathrm{x},\mathrm{y})$ задані слідами на відповідних системах взаємно перпендикулярних прямих в області $G=[0,1]^2$, тоді

$$R_N\left(H^{2,r}\left(\mathrm{M},M\right),H^{2,r}\left(M,M\right),\omega\right)\ge K\max\left\{\frac{1}{\ell^{2r}},\min\left\{1,\frac{|\omega|}{\ell^{2r}}\right\}\right\}.$$

Під слідом функції $f(x,y)$ на лініях $x_k = k\Delta_1 - \Delta_1/2$, $y_j = j\Delta_1 - \Delta_1/2$, $k, j = \overline{1,\ell_1}$, $\Delta_1 = 1/\ell_1$ розуміємо відповідно функції однієї змінної $f(x_k, y)$, $0 \le y \le 1$, $f(x, y_j)$, $0 \le x \le 1$. Під слідом функції $g(x,y)$ на лініях $x_p = p\Delta_2 - \Delta_2/2$, $y_s = s\Delta_2 - \Delta_2/2$, $p,s = \overline{1,\ell_2}$, $\Delta_2 = 1/\ell_2$ розуміємо відповідно функції однієї змінної $g(x_p, y)$, $0 \le y \le 1$, $g(x, y_s)$, $0 \le x \le 1$.

Нехай

$$h1_{0k}(x) = \begin{cases} 1, x \in X1_k, \\ 0, x \notin X1_k, \end{cases} \quad k = \overline{1,\ell_1}, \qquad H1_{0j}(y) = \begin{cases} 1, y \in Y1_j, \\ 0, y \notin Y1_j, \end{cases} \quad j = \overline{1,\ell_1},$$

$$X1_k = \left[x_{k-1/2}, x_{k+1/2}\right], \; Y1_j = \left[y_{j-1/2}, y_{j+1/2}\right],$$

$$x_k = k\Delta_1 - \Delta_1/2, \qquad y_j = j\Delta_1 - \Delta_1/2, \quad k, j = \overline{1,\ell_1}, \; \Delta_1 = 1/\ell_1,$$

$$h2_{0p}(x) = \begin{cases} 1, x \in X2_p, \\ 0, x \notin X2_p, \end{cases} \quad p = \overline{1,\ell_1}, \qquad H2_{0j}(y) = \begin{cases} 1, y \in Y2_s, \\ 0, y \notin Y2_s, \end{cases} \quad s = \overline{1,\ell_1},$$

$$X1_p = \left[x_{p-1/2}, x_{p+1/2}\right], \; Y1_s = \left[y_{s-1/2}, y_{s+1/2}\right],$$

$$x_p = p\Delta_2 - \Delta_2/2, \quad y_s = s\Delta_2 - \Delta_2/2, \quad \mathrm{p,s} = \overline{1,\ell_2}, \; \Delta_2 = 1/\ell_2.$$

Розглянемо оператори

$$J_{\ell_1}(x,y) = \sum_{k=1}^{\ell_1} f(x_k, y) h1_{0k}(x) + \sum_{j=1}^{\ell_1} f(x, y_j) H1_{0j}(y) -$$

$$- \sum_{k=1}^{\ell_1}\sum_{j=1}^{\ell_1} f(x_k, y_j) h1_{0k}(x) H1_{0j}(y),$$

$$O_{\ell_2}(x,y) = \sum_{p=1}^{\ell_2} g(x_p, y) h2_{0p}(x) + \sum_{s=1}^{\ell_2} g(x, y_s) H2_{0s}(y) -$$

$$- \sum_{p=1}^{\ell_2}\sum_{s=1}^{\ell_2} g(x_p, y_s) h2_{0p}(x) H2_{0s}(y).$$

Кубатурна формула

$$\Phi^2(\omega) = \int_0^1 \int_0^1 J_{\ell_1}(x,y) \mathrm{e}^{i\omega O_{\ell_2}(x,y)} \, dxdy$$

пропонується для наближеного обчислення інтегралу

$$I^2(\omega) = I(f, g, \omega).$$

Теорема 2. Нехай $f(x,y),\ g(x,y) \in H^{2,1}\left(M,M\right)$, функції $f(x,y),\ g(x,y)$ задані слідами $f\left(x_k,y\right),\ k=\overline{1,\ell_1}$, $f\left(x,y_j\right),\ j=\overline{1,\ell_1}$; $g(x_p,y),\ p=\overline{1,\ell_2}$, $g(x,y_s),\ s=\overline{1,\ell_2}$ на $N=2\ell_1+2\ell_2$ взаємно перпендикулярних прямих в області $G=\left[0,1\right]^2$. Тоді кубатурна формула $\Phi^2(\omega)$ є оптимальною за порядком точності, для якої справедливі наступні оцінки:

$$\rho\left(I^2(\omega),\Phi^2(\omega)\right)=$$

$$=\left|\int_0^1\int_0^1 f(x,y)\,\mathrm{e}^{i\omega g(x,y)}\,dxdy-\int_0^1\int_0^1 J_{\ell_1}\left(x,y\right)\mathrm{e}^{i\omega O_{\ell_2}(x,y)}\,dxdy\right|\le$$

$$\le\frac{M}{16}\frac{1}{\ell_1^{\,2}}+M\,\min\left(2;\frac{M\omega}{16}\frac{1}{\ell_2^{\,2}}\right),$$

та

$$R_N\left(H^{2,1}\left(\mathrm{M},M\right),H^{2,1}\left(M,M\right),\omega\right)\ge\ K\max\left\{\frac{1}{\ell^2},\min\left\{1,\frac{|\omega|}{\ell^2}\right\}\right\}$$

при $\ell_1=\ell_2=\ell$.

Доведення. Оцінка

$$R_N\left(H^{2,1}\left(\mathrm{M},M\right),H^{2,1}\left(M,M\right),\omega\right)\ge\ K\max\left\{\frac{1}{\ell^2},\min\left\{1,\frac{|\omega|}{\ell^2}\right\}\right\}$$

отримується з теореми 1 при $r=1$.

В роботі [2] показано, що

$$\rho\left(I^2(\omega),\Phi^2(\omega)\right)=$$

$$=\left|\int_0^1\int_0^1 f(x,y)\,\mathrm{e}^{i\omega g(x,y)}\,dxdy-\int_0^1\int_0^1 J_{\ell_1}\left(x,y\right)\mathrm{e}^{i\omega O_{\ell_2}(x,y)}\,dxdy\right|\le$$

$$\le\int_0^1\int_0^1\left|f(x,y)-J_{\ell_1}\left(x,y\right)\right|dxdy+\int_0^1\int_0^1\left|f(x,y)\right|\left|\mathrm{e}^{i\omega g(x,y)}-\mathrm{e}^{i\omega O_{\ell_2}(x,y)}\right|dxdy\le$$

$$\le\sum_{k=1}^{\ell_1}\sum_{j=1}^{\ell_1}\int_{x_{k-\frac12}}^{x_{k+\frac12}}\int_{y_{j-\frac12}}^{y_{j+\frac12}}\left|\int_{x_k}^{x}\int_{y_j}^{y} f^{(1,1)}\left(\xi,\eta\right)d\xi d\eta\right|dxdy+$$

$$+2M\sum_{p=1}^{\ell_2}\sum_{s=1}^{\ell_2}\int_{x_{p-\frac{1}{2}}}^{x_{p+\frac{1}{2}}}\int_{y_{s-\frac{1}{2}}}^{y_{s+\frac{1}{2}}}\min\left(1;\frac{\omega}{2}\left|\int_{x_p}^{x}\int_{y_s}^{y}g^{(1,1)}(\xi,\eta)d\xi d\eta\right|\right)dxdy\le$$

$$\le\frac{M}{16}{\Delta_1}^2+2M\min\left({\ell_2}^2{\Delta_2}^2,\frac{M\omega}{2}{\ell_2}^2\frac{{\Delta_2}^2}{4}\frac{{\Delta_2}^2}{4}\right)=$$

$$=\frac{M}{16}{\Delta_1}^2+M\min\left(2;\frac{M\omega}{16}{\Delta_2}^2\right)=\frac{M}{16}\frac{1}{{\ell_1}^2}+M\min\left(2;\frac{M\omega}{16}\frac{1}{{\ell_2}^2}\right).$$

При $\ell_1=\ell_2=\ell$, маємо

$$\rho\left(I^2(\omega),\Phi^2(\omega)\right)\le\frac{M}{16}\frac{1}{\ell^2}+M\min\left(2;\frac{M\omega}{16}\frac{1}{\ell^2}\right)\le C\max\left(\frac{1}{\ell^2},\min\left(1;\frac{|\omega|}{\ell^2}\right)\right).$$

Порівнюючи оцінки зверху та знизу, робимо висновок про оптимальність за порядком точності кубатурної формули. **Теорема доведена.**

Теорема 3. Нехай для $f(x,y),\ g(x,y)$ виконуються умови теореми 2. Тоді для кубатурної формули

$$\Phi_s^2(\omega)=\int_0^1\int_0^1 J_{\ell_1}(x,y)\sin\left(\omega O_{\ell_2}(x,y)\right)dxdy$$

наближеного обчислення

$$I_s^2(\omega)=\int_0^1\int_0^1 f(x,y)\sin\left(\omega g(\mathrm{x},\mathrm{y})\right)dxdy$$

справедлива наступна оцінка:

$$\rho\left(I_s^2(\omega),\Phi_s^2(\omega)\right)\le\frac{M}{16}\frac{1}{{\ell_1}^2}+M\min\left(2;\frac{M\omega}{16}\frac{1}{{\ell_2}^2}\right).$$

Доведення. Аналогічно доведенню теореми 2,

$$\rho\left(I_s^2(\omega),\Phi_s^2(\omega)\right)\le\int_0^1\int_0^1\left|f(x,y)-J_{\ell_1}(x,y)\right|dxdy+$$

$$+\int_0^1\int_0^1|f(x,y)|\left|\sin\left(\omega g(x,y)\right)-\sin\left(\omega O_{\ell_2}(x,y)\right)\right|dxdy\le$$

$$\le\int_0^1\int_0^1\left|f(x,y)-J_{\ell_1}(x,y)\right|dxdy+$$

$$+\int_0^1\int_0^1|f(x,y)|\left|2\sin\frac{\omega\,\mathrm{g}(\mathrm{x},\mathrm{y})-\omega O_{\ell_2}(x,y)}{2}\cos\frac{\omega\,\mathrm{g}(\mathrm{x},\mathrm{y})+\omega O_{\ell_2}(x,y)}{2}\right|dxdy\le$$

$$\le \sum_{k=1}^{\ell_1}\sum_{j=1}^{\ell_1}\int_{x_{k-\frac12}}^{x_{k+\frac12}}\int_{y_{j-\frac12}}^{y_{j+\frac12}}\left|\int_{x_k}^{x}\int_{y_j}^{y} f^{(1,1)}(\xi,\eta)\,d\xi\,d\eta\right|dxdy +$$

$$+2M\sum_{p=1}^{\ell_2}\sum_{s=1}^{\ell_2}\int_{x_{p-\frac12}}^{x_{p+\frac12}}\int_{y_{s-\frac12}}^{y_{s+\frac12}}\min\left(1;\frac{\omega}{2}\left|\int_{x_p}^{x}\int_{y_s}^{y} g^{(1,1)}(\xi,\eta)d\xi\,d\eta\right|\right)dxdy \le$$

$$\le \frac{M}{16}\Delta_1{}^2 + M\min\left(2;\frac{M\omega}{16}\Delta_2{}^2\right) = \frac{M}{16}\frac{1}{\ell_1{}^2} + M\min\left(2;\frac{M\omega}{16}\frac{1}{\ell_2{}^2}\right).$$

Чисельні результати. Обчислимо $I_s^2(\omega)$ за формулою $\Phi_s^2(\omega)$ (таблиця) у випадку, коли $f(x,y)=\sin(x+y)$, $g(x,y)=\cos(x+y)$ в MathCad 15.0. Точні значення інтегралів:

$$I_s^2(2\pi) = 0.062699216073162,\ I_s^2(5\pi) = 0.022780463640219.$$

Нехай $\varepsilon_{ex} = \left|I_s^2(\omega) - \Phi_s^2(\omega)\right|$. Покажемо, що

$$\varepsilon_{ex} \le \varepsilon_{th},\ \ \varepsilon_{th} = \frac{M}{16}\frac{1}{\ell_1{}^2} + M\min\left(2;\frac{M\omega}{16}\frac{1}{\ell_2{}^2}\right).$$

Таблиця

Обчислення $I_s^2(\omega)$ *за формулою* $\Phi_s^2(\omega)$

ω	ℓ_1	ℓ_2	$\Phi_s^2(\omega)$	ε_{ex}	ε_{th}
2π	4	4	0.062432583948326	$2.6\cdot10^{-4}$	$2.8\cdot10^{-2}$
2π	7	7	0.062683978467995	$1.5\cdot10^{-5}$	$9.2\cdot10^{-3}$
5π	6	4	0.022786668787906	$6.2\cdot10^{-6}$	$2.9\cdot10^{-2}$
5π	10	4	0.022808425368659	$2.7\cdot10^{-5}$	$6.1\cdot10^{-2}$
5π	10	10	0.02277048162594	$9.9\cdot10^{-6}$	$1.04\cdot10^{-2}$

Висновки. Розглядається кубатурна формула наближеного обчислення подвійних інтегралів від швидкоосцилюючих функцій загального виду у випадку, коли інформація про функції задана їх слідами на відповідних лініях. Доведена оптимальність за порядком точності запропонованої кубатурної формули на класі диференційовних функцій. Чисельний експеримент підтвердив теоретичні твердження.

Список використаних джерел:

1. Сергієнко І. В., Задірака В. К., Литвин О. М., Мельникова С. С., Нечуйвітер О. П. Оптимальні алгоритми обчислення інтегралів від швидкоосци-

люючих функцій та їх застосування : у 2 т. Т. 1. Алгоритми : [монографія]. Київ : Наук. думка, 2011. 447 с.
2. Lytvyn O. M., Nechuiviter O., Pershyna Yu., Mezhuyev V. Input Information in the Approximate Calculation of Two-Dimensional Integral from Highly Oscillating Functions (Irregular Case). Recent Developments in Data Science and Intelligent Analysis of Information. *Proceedings of the XVIII International Conference on Data Science and Intelligent Analysis of Information*, June 4–7, 2018. Kyiv, Ukraine. P. 365–373.
3. Mezhuyev V., Lytvyn O. M., Nechuiviter O., Pershyna Yu., Lytvyn O. O., Keita K. Cubature formula for approximate calculation of integrals of two-dimensional irregular highly oscillating functions. *U.P.B. Sci. Bull.*, Series A. Vol. 80, Iss. 3. 2018. P. 169–182.
4. Нечуйвітер О. П., Кейта К. В. Оптимальне інтегрування двовимірних швидкоосцилюючих функцій загального виду. *Математичне та комп'ютерне моделювання.* Сер. Фіз.-мат. науки : зб. наук. пр. Кам'янець-Подільський : Кам'янець-Подільський нац. ун-т ім. Івана Огієнка, 2017. Вип. 15. С. 139–144.

THE OPTIMAL BY THE ORDER OF EXACTNESS CUBATURE FORMULA FOR CALCULATION OF TWO-DIMENSIONAL INTEGRAL FROM HIGHLY OSCILLATING FUNCTIONS OF GENERAL VIEW

The paper is devoted to the optimal by the order of exactness cubature formula for calculation of two-dimensional integral from highly oscillating functions of general view in case when the information about functions is a set of lines.

Key words: *cubature formula, integral from highly oscillating function, class of differentiable functions.*

Одержано 24.01.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.98-104

Ю. І. Першина*, д-р фіз.-мат. наук,
В. О. Пасічник**, канд. техн. наук

*Українська інженерно-педагогічна академія, м. Харків,
**Харківська державна академія дизайну і мистецтв, м. Харків

РОЗВ'ЯЗАННЯ ЗАДАЧІ ВІДНОВЛЕННЯ РОЗРИВНИХ ФУНКЦІЙ МЕТОДОМ МІНІМАКСА

Запропоновано метод, за допомогою якого можна наблизити функції однієї та двох змінних з розривами першого роду в точках чи на лініях розривними апроксимаційними сплайнами. У двовимірному випадку область визначення досліджуваної функції розбивається на прямокутні елементи. Для розв'язування цієї задачі в даній роботі будується розривний білінійний апроксимаційний сплайн, невідомі параметри якого знаходяться методом мінімакса, тобто будується такий сплайн, який на кожному інтервалі чи прямокутному елементі має найменше максимальне відхилення від наближуваної функції. Як експериментальні дані виступають односторонні границі досліджуваної функції у заданих вузлах. Запропонований метод дозволяє уникати явище Гіббса, яке виникає при наближенні розривних функцій класичними неперервними конструкціями. В роботі детально описаний чисельний експеримент, який підтверджує ефективність запропонованого методу. Автори вважають перспективним розвиток теорії наближення розривних функцій багатьох змінних розривними сплайнами та побудову математичних моделей розривних процесів на основі розробленої теорії, оскільки існує багато практично важливих наукових та технічних галузей, в яких об'єкти дослідження математично описуються розривними. В подальшому планується узагальнити цей метод на випадок, коли вузли розривного сплайну не співпадають з точками розриву досліджуваної функції. Запропонований метод можна буде використати для відновлення внутрішньої структури об'єктів, що мають різну щільність, в медичних, геологічних, космічних та інших дослідженнях; в методах цифрової радіографії, обчислювальної томографії для визначення місця розташування і геометричних розмірів прихованих дефектів у контрольованому виробі, а також для контролю виробів у реальному масштабі часу його технологічного виготовлення.

Ключові слова: *розривні функції, розривний сплайн, апроксимація, інтерполяція, мінімакс.*

Вступ. Існує багато практично важливих наукових та технічних галузей, в яких об'єкти дослідження математично описуються величинами, що зазнають розрив. Такі об'єкти часто виникають у задачах, які

використовують дистанційні методи і, зокрема, в задачах томографії. В багатьох задачах геофізики встановлення місця розташування границь, що розділяють блоки з різними фізичними властивостями, є першим етапом у подальших дослідженнях, направлених на визначення фізичних величин, що характеризують внутрішню будову Землі. В комп'ютерній томографії при дослідженні внутрішньої структури тіла корисно враховувати його неоднорідність, тобто різну щільність у різних частинах тіла.

Той факт, що на сьогоднішній день не існує загальної теорії описів вказаних явищ та процесів, говорить про актуальність створення теорії наближення розривних функцій розривними конструкціями та розробки методів виявлення точок або ліній розриву функції.

Аналіз останніх досліджень. Задача наближення неперервних функцій неперервними сплайнами з достатньою повнотою описана у багатьох роботах. Існують багато технічних задач, в яких наближуюча функція не обов'язково є гладкою, іноді допустима її розривність — лише б похибка наближення була достатньо мала. Наближення такого типу раніше детально не розглядалося, існують тільки підходи до розв'язання такого типу задач, які працюють для частинних випадків. Існують методи розв'язання крайових задач з розривними розв'язками, в розвиток яких внесли значний вклад такі вчені, як І. В. Сергієнко, В. С. Дейнека, В. В. Скопецький, О. М. Литвин та інші [1]. У роботі А. Л. Агеева, Т. В. Антонової [2] запропонований метод визначення числа точок розриву та їх положення на основі використання явища Гіббса. Але для цього потрібна додаткова інформація: найменша та найбільша величини стрибків наближуючої функції. Крім того припускається, що інтервали, в яких знаходяться явища Гіббса, не перетинаються, тобто неможливо відділити точки розриву, що знаходяться близько один від одного. У роботі [3] розроблені методи відновлення ліній розриву за допомогою вейвлетів. Ці методи відновлення використовують полігармонійні вейвлети, які мають нескінченний носій. Такого типу конструкції, узагалі кажучи, можуть привести до згладжування сигналу, який досліджується, і вимагати додаткового аналізу отриманих результатів. У роботі автори пропонують загальну теорію побудови розривних сплайнів, множина яких, як частинний випадок, включають множину неперервних та неперервно-диференційовних до заданого порядку сплайнів, які можуть мати розриви першого роду в заданих точках або на заданій множині ліній ε -границь елементів.

У роботі [4] авторами запропонований метод відновлення розривної лінійної функції однієї змінної та алгоритм виявлення точок ε -розриву. В статті [5] запропонований метод наближення розривної функції однієї змінної розривним сплайном, використовуючи метод мінімакса. Дана робота присвячена узагальненню статті [5] на випадок наближення розривної функції двох змінних.

Метод наближення розривної функції однієї змінної. Нехай задано функцію однієї змінної $f(x)$ на інтервалі $[a,b]$ з можливими розривами першого роду в точках x_k, $k=\overline{1,n-1}$. Припускаємо, що хоча б в одному вузлі x_k функція має розрив. Задані вузли розбивають інтервал $[a,b]$ на $n-1$ частин.

Визначення 1. Розривним інтерполяційним лінійним сплайном на відрізку $[x_k, x_{k+1}], k=\overline{1,n-1}$ називається функція:

$$S(x)=Sp_k(x,C)=C_k^+\frac{x-x_{k+1}}{x_k-x_{k+1}}+C_{k+1}^-\frac{x-x_k}{x_{k+1}-x_k},\ k=\overline{1,n-1}, \qquad (1)$$

де $C_k^+, C_{k+1}^-,\ k=\overline{1,n-1}$ — параметри сплайну $S(x)$, що визначаються у вигляді односторонніх границь

$$C_k^+=\lim_{x\to x_k+0} f(x),\quad C_{k+1}^-=\lim_{x\to x_{k+1}-0} f(x).$$

Треба знайти такі параметри $C_k^+, C_{k+1}^-,\ k=\overline{1,n-1}$ у сплайні (1), щоб наближення було найкращим у тому чи іншому сенсі. Для розв'язування цієї задачі використовуємо методом мінімакса.

Сплайном найкращого наближення будемо вважати сплайн, який на кожному з інтервалів $[x_k, x_{k+1}],\ k=\overline{1,n-1}$ має найменше максимальне відхилення від наближуваної функції $f(x)$.

Теорема 1. Якщо на кожному з інтервалів $[x_k, x_{k+1}],\ k=\overline{1,n-1}$ невідомі параметри $C_k^+, C_{k+1}^-,\ k=\overline{1,n-1}$ знаходити з умови

$$\max_{1\le k\le n-1}\left|f(x)-Sp_k(x)\right|\to\min_C, \qquad (2)$$

то отримаємо розривний сплайн найкращого наближення.

Теорема 2. Якщо наближувана функція $f(x)$ є розривною лінійною функцією з точками розриву $x=x_k, k=\overline{1,n}$ і наближуємо її лінійним розривним сплайном $S(x)$, що визначається формулами (1), і невідомі параметри-елементи $C_k^+, C_{k+1}^-,\ k=\overline{1,n-1}$ знаходимо з умови (2), то отримаємо точно наближувану функцію, тобто $S(x)=f(x)$.

Точки розриву функції збігаються з точками розриву наближувального сплайна і найкраще наближення сплайна до функції виконуємо аналітично. На кожному з інтервалів $[x_k, x_{k+1}],\ k=\overline{1,n-1}$ знаходимо максимальне значення відхилення сплайна від функції, яке буде дорівнювати одному із значень:

$$J_{[x_k,x_{k+1}]}(C)=\max\limits_{[x_k,x_{k+1}]}\{\left|f_k(x_k)-Sp_k(x_k,C)\right|,\left|f_k(x_{k+1})-Sp_k(x_{k+1},C)\right|,\\ \left|f_k(a_2)-Sp_k(a_1,C)\right|,...,\left|f_k(a_m)-Sp_k(a_m,C)\right|\},\tag{3}$$

де a_l, $l=\overline{1,m}$ — стаціонарні точки функції $J_k(x,C)=f_k(x)-Sp_k(x,C)$ на інтервалі $[x_k,x_{k+1}]$, $k=\overline{1,n-1}$.

Потім знаходимо мінімум від отриманого максимуму по всіх інтервалах:

$$W=\min\limits_{1\le k\le n-1}(J_{[x_k,x_{k+1}]}(C))=\min\limits_{1\le k\le n-1}(\max\limits_{a\le x\le b}\left|f(x)-Sp_k(x,C)\right|).$$

Отримуємо матрицю W, яка і представляє собою шукану матрицю параметрів C_k^+, C_{k+1}^-, $k=\overline{1,n-1}$.

Метод наближення розривної функції двох змінних. Нехай в області $D=[0,1]^2$ задано розривну функцію $f(x,y)$ та деяке розбиття на елементи (прямокутники) $\Pi_{i,j}=[x_i,x_{i+1}]\times[y_j,y_{j+1}]$, $0=x_1<$ $<x_2<...<x_m=1$, $0=y_1<y_2<...<y_n=1$. Вважаємо, що в кожній точці (x_i,y_j) може бути задано чотири різних значення досліджуваної функції:

$$C_{i,j}^{++}=\lim\limits_{\substack{x\to x_i+0\\y\to y_j+0}}f(x,y),\ C_{i,j}^{-+}=\lim\limits_{\substack{x\to x_i-0\\y\to y_j+0}}f(x,y),$$

$$C_{i,j}^{+-}=\lim\limits_{\substack{x\to x_i+0\\y\to y_j-0}}f(x,y),\ C_{i,j}^{--}=\lim\limits_{\substack{x\to x_i-0\\y\to y_j-0}}f(x,y).$$

Визначення 2. Будемо називати розривним білінійним інтерполяційним сплайном на прямокутній сітці сплайн вигляду

$$S(x,y)=p_{ij}(x,y,C)=C_{i,j}^{++}\frac{x-x_{i+1}}{x_i-x_{i+1}}\frac{y-y_{j+1}}{y_j-y_{j+1}}+C_{i+1,j}^{-+}\frac{x-x_i}{x_{i+1}-x_i}\frac{y-y_{j+1}}{y_j-y_{j+1}}+\\ +C_{i,j+1}^{+-}\frac{x-x_{i+1}}{x_i-x_{i+1}}\frac{y-y_j}{y_{j+1}-y_j}+C_{i+1,j+1}^{--}\frac{x-x_i}{x_{i+1}-x_i}\frac{y-y_j}{y_{j+1}-y_j},\tag{4}$$
$$(x,y)\in\Pi_{i,j},i=\overline{1,m-1},\ \ j=\overline{1,n-1}.$$

Треба знайти такі параметри: $C_{i,j}^{++}, C_{i,j}^{-+}, C_{i,j+1}^{+-}, C_{i,j+1}^{--}$ у сплайні (1), щоб наближення було найкращим у тому чи іншому сенсі. Для розв'язування цієї задачі використовуємо методом мінімакса [6].

Теорема 3. Якщо на кожному з прямокутних елементів $\Pi_{i,j}$, $i=\overline{1,m-1}$, $j=\overline{1,n-1}$ невідомі параметри $C_{i,j}^{++}, C_{i,j}^{-+}$, $C_{i,j+1}^{+-}, C_{i,j+1}^{--}$ знаходити з умови

$$\max_{\substack{1\le i\le m-1\\ 1\le j\le n-1}} \left| f(x,y) - p_{ij}(x,y) \right| \to \min_{C}, \qquad (5)$$

то отримаємо розривний сплайн найкращого наближення.

Теорема 4. Якщо наближувана функція $f(x,y)$ є розривною лінійною функцією з лініями розриву $x = x_i, i = \overline{1,m}$, $y = y_j$, $j = \overline{1,n}$ і наближуємо її лінійним розривним сплайном $S(x,y)$, що визначається формулами (4), і невідомі параметри-елементи $C_{i,j}^{++}, C_{i,j}^{-+}, C_{i,j+1}^{+-}, C_{i,j+1}^{--}$ знаходимо з умови (5), то отримаємо точно наближувану функцію, тобто $S(x,y) = f(x,y)$.

Точки розриву функції збігаються з точками розриву наближуваного сплайна. Знайдемо найкраще наближення сплайна до функції. На кожному з прямокутних елементів $\Pi_{i,j}$, $i = \overline{1,m-1}$, $j = \overline{1,n-1}$ знаходимо максимальне значення відхилення сплайна від функції, яке буде дорівнювати одному із значень:

$$J_{\Pi_{i,j}}(C) = \max_{\substack{[x_i, x_{i+1}]\\ [y_j, y_{j+1}]}} \{ \left| f(x_i, y_j) - p_{ij}(x_k, y_j, C) \right|, \left| f(x_i, y_{j+1}) - p_{ij}(x_i, y_{j+1}, C) \right|,$$
$$\left| f(x_{i+1}, y_j) - p_{ij}(x_{i+1}, y_j, C) \right|, \left| f(x_{i+1}, y_{j+1}) - p_{ij}(x_{i+1}, y_{j+1}, C) \right|$$
$$\left| f(D_1) - p_{ij}(D_1, C) \right|, ..., \left| f(D_k) - p_{ij}(D_k, C) \right|, \left| f(B_1) - p_{ij}(B_1, C) \right|, ..., \qquad (6)$$
$$\left| f(B_\ell) - p_{ij}(B_\ell, C) \right| \},$$

де D_k, $k = \overline{1,(m-1)\cdot(n-1)}$ — стаціонарні точки функції $J_{\Pi_{ij}}(x,y,C) = f(x,y) - p_{ij}(x,y,C)$ всередині прямокутного елемента $\Pi_{i,j}$, $i = \overline{1,m-1}$, $j = \overline{1,n-1}$, B_ℓ, $\ell = \overline{1,L}$ — критичні точки функції $J_{\Pi_{ij}}(x,y,C)$ на сторонах прямокутного елемента $\Pi_{i,j}$, $i = \overline{1,m-1}$, $j = \overline{1,n-1}$, L — кількість критичних точок.

Потім знаходимо мінімум від отриманого максимуму по всіх прямокутних елементах:

$$W = \min_{\substack{1\le i\le m-1\\ 1\le j\le n-1}} (J_{\Pi_{ij}}(C)) = \min_{\substack{1\le i\le m-1\\ 1\le j\le n-1}} (\max_{x\in D} \left| f(x,y) - p_{ij}(x,y,C) \right|).$$

Отримуємо матрицю W, яка і представляє собою шукану матрицю параметрів $C_{i,j}^{++}, C_{i,j}^{-+}, C_{i,j+1}^{+-}, C_{i,j+1}^{--}$.

Приклад. Нехай задано функцію $f(x,y)$, яка є нелінійною, на області $[0;\pi]\times[0;1]$ з однією лінією розриву:

$$f(x,y)=\begin{cases} x+y, & 0\le x\le \dfrac{\pi}{2},\ 0\le y\le 1,\\ \sin(x+y), & \dfrac{\pi}{2}< x\le \pi,\ 0\le y\le 1. \end{cases}$$

Обираємо сітку вузлів: $x_1=0,\ x_2=-\pi/2, x_3=\pi,\ \ y_1=0,\ y_2=1$. Наближуємо сплайном вигляду (4). У цьому випадку маємо дві точки, в яких побудована функція має розриви першого роду:

$$f^{-+}\left(\frac{\pi}{2};0\right)=1{,}57;\ \ f^{++}\left(\frac{\pi}{2};0\right)=0;\ \ f^{--}\left(\frac{\pi}{2};1\right)=2{,}57;\ \ f^{+-}\left(\frac{\pi}{2};1\right)=-0.9.$$

За допомогою системи комп'ютерної математики MathCad була отримана наступна матриця коефіцієнтів:

$$C=\begin{pmatrix} 0 & 1.57 & 1 & 2.57\\ 0 & 0 & -0.9 & 0.9 \end{pmatrix}.$$

Тобто найкраще наближення заданої функції $f(x,y)$ має вигляд:

$$S(x,y)=\begin{cases} x+y, & (x,y)\in \Pi_{11},\\ 1{,}16xy-2{,}73y, & (x,y)\in \Pi_{21}. \end{cases}$$

Висновки. Таким чином, в роботі запропонований метод, за допомогою якого можна наблизити функцію однієї та двох змінних з розривами першого роду розривним сплайном, використовуючи метод мінімакса. В подальшому планується узагальнити цей метод на випадок, коли вузли сплайна не співпадають з точками розриву функції.

Як вже зазначалося, цей метод можна буде використати для відновлення внутрішньої структури об'єктів, що мають різну щільність, у медичних, геологічних, космічних та інших дослідженнях.

Список використаних джерел:

1. Дейнека В. С., Сергиенко И. В. Анализ многокомпонентных распределенных систем и оптимальное управление : монография. Киев : Наук. думка, 2007. 703 с.
2. Агеев А. Л., Антонова Т. В. Аппроксимация линий разрыва зашумленной функции двух переменных. *Сибирский журнал индустриальной математики*. Новосибирск, 2012. Т.15, № 1(49). С. 3–13.
3. Rossini M. Detecting discontinuities in two-dimensional signals sampled on a grid. *Journal of Numerical Analysis, Industrial and Apply Mathematics*. 2007. Vol. 1, № 1. P. 1–13.
4. Першина Ю. І., Пасічник В. О. Чисельна реалізація методу виявлення точок розриву першого роду функції однієї змінної. *Вісник ХНТУ*. Херсон, 2017. № 3 (62), Т. 1. С. 80–84.
5. Першина Ю. І, Пасічник В. О. Наближення розривних функцій розривними сплайнами методом мінімакса. *Вісник ХНТУ*. Херсон, 2018. № 3(66), Т. 2. С. 82–87.

6. Демьянов В. Ф., Малоземов В. Н. Введение в минимакс. Москва : Наука, 1972. 368 с.

SOLUTION OF THE PROBLEM OF RESTORING DISCONTINUOUS FUNCTIONS BY THE MINIMAX

The article suggests a method for approximating a function of one and two variables with discontinuities of the first kind by a discontinuous approximation spline. The experimental data are the one-sided boundaries of the given nodes. To solve this problem in this paper, we use the minimax method.

Key words: *discontinuous functions, discontinuous spline, approximation, interpolation, minimax.*

Одержано 14.02.2019

УДК 519.1
DOI: 10.32626/2308-5878.2019-19.104-111

В. І. Петренюк, канд. фіз.-мат. наук, доцент

Центральноукраїнський національний технічний університет,
м. Кропивницький

СТРУКТУРА 20-ТИ 9-ТИ ВЕРШИННИХ ГРАФІВ-ОБСТРУКЦІЇ ТОРА

Досліджено структуру решти 9-ти вершинних графів-обструкцій для тору.

Ключові слова: *граф-обструкція, тор, φ-перетворення графів.*

Вступ. Основні визначення та позначення взято з [1]. У роботі [2] запропоновано спосіб побудови графів-обструкцій обмеженого орієнтованого роду як φ-образу двох графів, один з яких має бути квазізіркою, з'єднаних шляхом ототожнення пар вершин, для випадку несуттєвості порядку ототожнення зазначених пар точок; тобто один із підграфів породжених підмножинами точок допускатиме перестановку довільної пари тачок з'єднання, наприклад, є повним. Цей підхід може видавати такі графи, які набуватимуть статус обструкцій після стискання в точку усіх лишніх ребер-променів квазізірки, саме так побудовані зазначені графи. Однак не всі графи-обструкції для тору можливо отримати цим способом. Однією з причин відсутності лишніх ребер є наявність двостороннього доступу до деяких точок із тих пар точок, що підлягають ототожненню в точку-вершину графа.

Задача полягатиме у завершенні розпочатої в [4] роботи по вивченню структури 9-ти вершинних графів-обструкцій для тору, наве-

дених у [3] для використання при побудові n-вершинних, $n > 9$, графів-обструкцій для тору.

Лема 1. Виконуються наступні твердження:

1) D_{32} є φ-образом графів $K_6 \setminus K_2^1$ та квазізірки H з центром $K_{2,3}$, де $K_6^0 \setminus K_2^1 = \{i'\}_1^6$, $K_6^1 \setminus K_2^1 = K_6^1 \setminus \{(1',3'),(2',5')\}$, $H^0 = \{i''\}_1^6 \cup \{a,b,v\}$, $H^1 = K_{2,3}^1 \cup \{(2'',a),(b,3''),(b,1''),(b,5'')\}$, при перетворенні заданому формулою: $\varphi(K_6 \setminus K_2^1 + H, \sum_{i=1}^{6} (i' \dot{+} i'')) \to (D_{32}, \{i\}_{i=1}^6)$ та виконаному шляхом ототожнення усіх пар (i', i'') вершин з множин $M' = \{i'\}_1^6$ та $M'' = \{i''\}_1^6$;

2) D_{33} — граф-обструкція для тору є φ-образом графів K_5 та $H \cup St_3(b)$, де H -квазізірка з центром $K_{2,3}$, $K_5^0 = \{i'\}_1^5$, $H^1 = K_{2,3}^1 \cup \{(1'',a),(v,3''),(c,5'')\}$, $H^0 = \{i''\}_1^5 \cup \{a,c,v\}$, $St_3(b)$ — проста зірка з центром b, $St_3^0(b) = \{i''\}_1^3$, при перетворенні заданому наступною формулою: $\varphi(K_5 + H \cup St_3(b), \sum_{i=1}^{5} (i' \dot{+} i'')) \to (D_{33}, \{i\}_{i=1}^5)$ та виконаному шляхом ототожнення усіх пар (i', i'') вершин з множин $M' = \{i'\}_1^5$ та $M'' = \{i''\}_1^5$;

3) D_{34} є φ-образом графів $K_6 \setminus K_3^1$ та H, де $K_6^0 = \{i'\}_1^6$, $K_6(\{i''\}_{i=4}^6) = \overline{K_3}$, $H^0 = \{i''\}_1^6 \cup \{a,b,c\}$, $K_6(\{i''\}_{i=}^3) = K_3$, $H^1 = K_5^1 \setminus \setminus \{(c,1''),(c,4'')\} \cup \{(c,6''),(1'',6''),(d,1''),(2'',b),(3'',a),(5'',c)\}$, де вершина 6'' розділяє ребро $(c,1'')$ графа $K_5 \setminus (c,4'')$, при перетворенні заданому: $\varphi(K_6 \setminus K_3^1 + H, \sum_{i=1}^{6} (i' \dot{+} i'')) \to (D_{34}, \{i\}_{i=1}^6)$ шляхом ототожнення усіх пар (i', i'') з множин приєднання $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$;

4) D_{35} φ-образ графів $K_6 \setminus (K_3^1 + K_2^1)$ та H, де $K_6^0 = \{i'\}_1^6$, $K_6(\{i''\}_{i=4}^6) = \overline{K_3}$, $K_6(\{i''\}_{i=}^3) = K_3 \setminus (4',5')$, $H^0 = \{i''\}_1^6 \cup \{a,b,v\}$, $H^1 = K_{2,3}^1 \cup \{(b,2''),(4'',v),(3'',a),(5'',v)\}$, при перетворенні заданому: $\varphi(K_6 \setminus (K_3^1 + K_2^1) + H, \sum_{i=1}^{6} (i' \dot{+} i'')) \to (D_{35}, \{i\}_{i=1}^6)$ ототожненням усіх пар (i', i'') з множин приєднання $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$.

Лема 2.

1) D_{36} — обструкція для тору є φ-образом графів $K_6 \setminus K_3^1$ та H, де $K_6^0 = \{i'\}_1^6$, $K_6(\{i"\}_{i=4}^6) = \overline{K_3}$, $K_6(\{i"\}_{i=}^3) = K_3$, $H^0 = \{i"\}_1^6 \cup \{a,b,v\}$, $H^1 = K_4^1 \setminus \{(a,v)\} \cup \{(1",a),(v,1"),(4",b),(5",b),(2",a),(3",v)\}$, де вершина $1"$ розділяє ребро (a,v) графа K_4, при перетворенні заданому наступною формулою: $\varphi(K_6 \setminus K_3^1 + H, \sum_{i=1,}^{6} (i' \dot{+} i")) \to (D_{36}, \{i\}_{i=1}^6)$ шляхом ототожнення усіх пар $(i', i")$ вершин з множин приєднання $M' = \{i'\}_1^6$, $M" = \{i"\}_1^6$;

2) D_{37} — обструкція для тору є φ-образом графів $K_{3,3}$ та H, де $K_{3,3}^0 = \{i'\}_1^6$, $K_{3,3}(\{i"\}_{i=4}^6) = \overline{K_3}$, $K_{3,3}(\{i"\}_{i=1}^3) = K_3$, $H^1 = K_4^1 \cup \cup\{(5",a), \cup\{(5",a),(a,4"),(2",b),(5",b),(2",c),(4",c),(a,1"),(a,3")\}$, $H^0 = \{i"\}_1^6 \cup \cup\{a,b,c\}$, $H\{a,b,c,6"\} = K_4$, де вершини $5",4",2"$ розділяють ребра $(a,b),(c,b),(a,c)$ графа K_4 (із трьома кратними ребрами), при перетворенні заданому: $\varphi(K_{3,3} + +H, \sum_{i=1,}^{6} (i' \dot{+} i")) \to (D_{37}, \{i\}_{i=1}^6)$ шляхом ототожнення усіх пар $(i', i")$ вершин з множин приєднання $M' = \{i'\}_1^6$, $M" = \{i"\}_1^6$;

3) D_{38} — граф-обструкція для тору є φ-образом графів K_5, $K_5^0 = \{i'\}_1^7$, де вершини $6',7'$ розділяють ребро $(4',5')$, та H, де H — квазізірка з центром $K_5 \setminus e$, $H^0 = \{i"\}_2^5 \cup \{a,7",6",v\}$, $H(\{a,5",6",7",4",v\}) \cong K_4$, $H^1 = K_5^1 \setminus \{(v,6"),(a,7")\} \cup \{(3",a),(v,4"),(6",4"),(a,5"),(7",5")\}$, при перетворенні заданому: $\varphi(K_5 + +H, \sum_{i=2}^{7} (i' \dot{+} i")) \to (D_{38}, \{i\}_{i=2}^7)$ та виконаному шляхом ототожнення усіх пар $(i', i")$ вершин з $M' = \{i'\}_2^7$ та $M" = \{i"\}_2^7$;

4) D_{39} є φ-образом графів K_5, $K_5^0 = \{i'\}_1^5$, та H, де H — квазізірка із центром C_4 — простим циклом довжини 4, де $C_5^1 = \{(a,b),(v,b),(a,c),(c,v)\}$, $H^1 = C_5^1 \cup \{(5",a),(5",b),(4",a),(4",b),(v,1"),(v,3"),(c,1"),(c,3"),(v,2"),(b,2")\}$, $H^0 = \{i"\}_1^5 \cup C_4^0$ при перетворенні за-

даному формулою: $\varphi(K_5 + H, \sum_{i=1}^{5} (i' \dot{+} i")) \to (D_{39}, \{i\}_{i=1}^{5})$ та виконаному шляхом ототожнення усіх пар $(i', i")$ з множин $M' = \{i'\}_1^5$ та $M" = \{i"\}_1^5$.

Лема 3.

1) D_{40} — обструкція для тору є φ-образом графів $K_{3,3}$ та H, де $K_{3,3}^0 = \{i'\}_1^6$, $H^1 = 3K_4^1$, $H^0 = \{i"\}_1^6 \cup \{a,b,c\}$, причому кожна пара графів K_4 матиме тільки одну спільну вершину з множини $\{a,b,c\}$, де $H\{a,b,c\} = K_3$, при перетворенні заданому наступною формулою: $\varphi(K_{3,3} + H, \sum_{i=1,}^{6} (i' \dot{+} i")) \to (D_{40}, \{i\}_{i=1}^{6})$ шляхом ототожнення усіх пар $(i', i")$ вершин з множин приєднання $M' = \{i'\}_1^6$, $M" = \{i"\}_1^6$;

2) D_{41} є φ-образ графів K_5 та H, де $H^0 = \{i"\}_1^5 \cup \{a,b,c,v\}$, $K_5^0 = \{i'\}_1^5$, $H^1 = K_5^1 \setminus (v,2") \cup \{(a,5"),(c,5"),(a,4"),(c,4"),(b,3"),(c,3"),(a,1"),(b,1")\}$, H — квазізірка з центром $K_{1,3}$, який на множині вершин $\{i"\}_1^4 \cup \{a,b,c,v\}$ породжує підграф гомеоморфний $K_5 \setminus (v,2")$, при перетворенні заданому наступною формулою: $\varphi(K_5 + H, \sum_{i=1,}^{5} (i' \dot{+} i")) \to (D_{41}, \{i\}_{i=1}^{5})$ шляхом ототожнення усіх пар $(i', i")$ вершин з множин приєднання $M' = \{i'\}_1^5$, $M" = \{i"\}_1^5$;

3) D_{42} — граф-обструкція для тору є φ-образом графів K_5, $K_5^0 = \{i'\}_1^5$, та H, де H — квазізірка з центром C_4 — простим циклом довжини 4, який на вершинах $a,b,c,v,2"$ породжує підграф гомеоморфний K_4, де $C_4^1 = \{(a,b),(c,b),(a,v),(c,v)\}$, $H^0 = \{i"\}_1^5 \cup C_4^0$, $H^1 = K_4^1 \cup \{(1",a),(1",b),(3",v),(3",b),(v,4"),(a,4"),(c,5")\}$, при перетворенні заданому формулою: $\varphi(K_5 + H, \sum_{i=1}^{5} (i' \dot{+} i")) \to$ $\to (D_{42}, \{i\}_{i=1}^{5})$ та виконаному шляхом ототожнення усіх пар $(i', i")$ з множин $M' = \{i'\}_1^5$ та $M" = \{i"\}_1^5$;

4) D_{43} — обструкція для тору є φ-образом графів K_5 та H, де $H^0 = \{i"\}_1^5 \cup \{a,b,c,v\}$, $H^1 = K_5^1 \setminus (1",2") \cup \{(a,4"),(c,4"),(b,3"),(v,5")\}$,

$K_5^0 = \{i'\}_1^5$, H — квазізірка з центром C_4, який на множині вершин $\{i''\}_1^2 \cup \{a,b,c,v\}$ породжує підграф гомеоморфний $K_5 \setminus (1'',2'')$, при перетворенні заданому формулою: $\varphi(K_5 + H, \sum_{i=1,}^{5} (i' \dot{+} i'')) \to (D_{43}, \{i\}_{i=1}^5)$ шляхом ототожнення усіх пар (i', i'') вершин з множин приєднання $M' = \{i'\}_1^5$, $M'' = \{i''\}_1^5$.

Лема 4. Виконуються наступні твердження:

1) D_{44} — обструкція для тору є φ-образом графів $K_{3,3}$ та H, де $H^0 = \{i''\}_1^6 \cup \{a,b,c\}$, $H^1 = K_4^1 \setminus (a,c) \cup \{(a,1''),(c,1''),(a,5''),(b,5''),(a,3''),(b,3''),(c,2''),(b,2''),(c,4''),(b,4'')\}$, $K_{3,3}^0 = \{i'\}_1^6$, де $H\{a,b,c,6'',1''\} \cong K_4$, причому вершина $1''$ розділяє ребро (a,c), при перетворенні заданому наступною формулою: $\varphi(K_{3,3} + H, \sum_{i=1,}^{6} (i' \dot{+} i'')) \to (D_{44}, \{i\}_{i=1}^6)$ шляхом ототожнення усіх пар (i', i'') вершин з множин приєднання $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$;

2) D_{45} — обструкція для тору є φ-образом графів $K_6 \setminus K_3^1$ та H, де $H^0 = \{i''\}_1^5 \cup \{a,v,c\}$ $H = St_5(a) + St_5(c) + St_3(v)$, $H^1 = St_5^1(a) \cup \cup St_5^1(c) \cup St_3^1(v)$, $K_6^0 = \{i'\}_1^6$, при перетворенні заданому наступною формулою: $\varphi(K_6 \setminus K_3^1 + H, \sum_{i=1,}^{6} (i' \dot{+} i'')) \to (D_{45}, \{i\}_{i=1}^6)$ шляхом ототожнення усіх пар (i', i'') вершин з множин приєднання $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$;

3) D_{46} — обструкція для тору є φ-образом графів K_5 та H, де $H^0 = \{i''\}_1^5 \cup \{a,b,c'',v\}$, $K_5^0 = \{i'\}_1^5 \cup \{c'\}$, вершина c' розділяє ребро $(1',3')$, $H^1 = K_5^1 \setminus \{(1'',3'')\} \cup K_4^1 \cup \{(a,4''),(a,5'')\}$, $H\{2'',3'',c'',v\} = = K_4$ $H\{1'',3'',v,a,b,c''\} = K_5 \setminus (1'',3'')$, $H(\{4'',5'',a,b,c''\}) = K_{2,3}$, H — квазізірка з центром на множині вершин $\{a,b,v\}$ породжує підграф гомеоморфний K_3, при перетворенні заданому формулою: $\varphi(K_5 + H, \sum_{i=1}^{5} (i' \dot{+} i''), (c' + c'')) \to (D_{47}, \{i\}_{i=1}^5, c)$ шляхом отото-

жнення усіх пар (i',i'') вершин з множин приєднання $M' = \{i'\}_1^5$, $M'' = \{i''\}_1^5$, та пари $(c'c'')$;

4) D_{47} — обструкція для тору є φ-образом графів K_5 та H, де $H^0 = \{i''\}_1^5 \cup \{a,b,c,v\}$, $K_5^0 = \{i'\}_1^5$, $H^1 = K_{3,3}^1 \cup K_{2,3}^1$, $K_{3,3}^0 = \{1'',3'',v,\} \cup \cup \{a,b,c\}$, $H\{1'',3'',v,\} = H\{a,b,c\} = \overrightarrow{K_3}$, $H(\{4'',5'',a,b,c\}) = K_{2,3}$ — квазізірка з центром на множині вершин $\{a,b,c,v\}$ породжує підграф гомеоморфний $K_{1,3}$, при перетворенні заданому наступною формулою: $\varphi(K_5 + H, \sum_{i=1,}^{5} (i' \dot{+} i'')) \to (D_{47}, \{i\}_{i=1}^5)$ шляхом ототожнення усіх пар (i',i'') вершин з множин приєднання $M' = \{i'\}_1^5$, $M'' = \{i''\}_1^5$.

Лема 5. Виконуються наступні твердження:

1) D_{48} є φ-образом $K_6 \setminus K_{1,2}^1$ та H, де $K_{3,3}^0 \setminus K_{1,2}^1 = \{i'\}_1^6$, $H^0 = \{i''\}_{1,i\neq 2}^6 \cup \{a,b,b''\}$, вершина b' розділяє $(2',6')$, $H^1 = K_5^1 \setminus \setminus (5'',6'') \cup \{(a,1''),(b,1''),(a,5''),(a,4''),(c'',3'')\}$, де $H\{a,b,c'',6'',5''\} \cong \cong K_5 \setminus (5'',6'')$, при перетворенні заданому формулою: $\varphi((K_6 \setminus K_{1,2}^1) + H, (\sum_{i=1,i\neq 2}^{6} (i' \dot{+} i''), (b' + b''))) \to (D_{48}, (\{i\}_{i=1,i\neq 2}^6, b)))$ та виконаному шляхом ототожнення усіх пар (i',i'') вершин з множин $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$ та (b',b'') в b, причому вершина $2'$ стане вершиною 2;

2) D_{49} є φ-образом графів $K_6 \setminus 2K_2^1$ та H, де $K_6 \setminus 2K_2^1 = \{i'\}_1^6$, $K_6^1 \setminus 2K_2^1 = K_6^1 \setminus \{(4',6'),(2',5')\}$, $H^0 = \{i''\}_1^6 \cup \{a,b,c\}$, $H(\{i''\}_1^2 \cup \{a,b\}) = = K_4$, $H(\{1'',2'',4'',6''c\}) = St_4(c)$, $H^1 = K_4^1 \cup St_4^1(c) \cup \{(a,6''),(b,3''),$ $(a,4''),(a,5'')\}$, при перетворенні заданому формулою: $\varphi((K_6 \setminus 2K_2^1) + H, \sum_{i=1}^{6} (i' \dot{+} i'')) \to (D_{49}, \{i\}_{i=1}^6))$ та виконаному шляхом ототожнення усіх пар (i',i'') вершин з множин $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$;

3) D_{50} — φ-образ графів K_6 та H, де $K_6^0 = \{i'\}_1^6$, $H^0 = \{i''\}_1^6 \cup \{a,b,c\}$, $H(\{i''\}_4^5 \cup \{c,b\}) = K_4$, $H(\{1'',6'',4'',5''a\}) = St_4(a)$, $H^1 = K_4^1 \cup St_4^1(a) \cup \cup\{(c,2''),(b,3'')\}$, при перетворенні заданому формулою: $\varphi(K_6 + H, \sum_{i=1}^{6}(i' + i'')) \to (D_{50}, \{i\}_{i=1}^6))$ та виконаному шляхом ототожнення усіх пар (i', i'') вершин з множин $M' = \{i'\}_1^6$, $M'' = \{i''\}_1^6$;

4) D_{51} — обструкція для тору є φ-образом графів K_5 та H, де $H^0 = \{i''\}_1^5 \cup \{a,b,c,v\}$, $K_5^0 = \{i'\}_1^5$, $H^1 = H_1^1 \cup H_2^1 \cup \{(b,1''),(v,3'')\}$, $H_1^0 = \{4'',5'',c,b,v\}$, $H_2^1 = \{2'',4'',5'',a,c\}$, $H_1 = H(\{4'',5'',c,b,v\}) = = K_5 \setminus (4'',5'')$, $H_2 = H(\{2'',4'',5'',a,c\}) = K_5 \setminus (4'',5'')$, причому H_2 має з K_5 спільний ланцюг довжини 2 на вершинах $\{4'',5'',c\}$, де H — квазізірка з центром на множині вершин $\{a,b,c,v\}$ на якій породжує підграф $K_4 \setminus K_{1,2}$, при перетворенні заданому наступною формулою: $\varphi(K_5 + H, \sum_{i=1}^{5}(i' + i'')) \to (D_{51}, \{i\}_{i=1}^5)$ шляхом ототожнення усіх пар (i', i'') вершин з множин приєднання $M' = \{i'\}_1^5$, $M'' = \{i''\}_1^5$;

Із вищенаведених лем 1–5 випливатиме ***основний результат.***

Теорема. Кожна граф-обструкція роду 2 D_{32},..., D_{51} [5] на 9-ти вершинах є результатом φ-перетворення трьох зв'язних графів X, Y, Z, які задовольняють одному з наступних випадків:

1) граф Y гомеоморфний K_5 чи $K_{3,3}$ (можливо із кількома додатковими ребрами) вкладений в тор σ, граф Z відсутній, а інший граф X є або площинним 2-мінімальним відносно множини точок приєднання до графа Y на недвоклітці $\sigma \setminus Y$ із нульовими характеристиками θ та $\partial\theta$ для множини точок приєднання до графа Y, або площинним 3-мінімальним на s недвоклітці тора, $s \in \sigma_1 \setminus Y$, із характеристиками θ, $\partial\theta$, де $\theta = 1$ чи $\partial\theta = 1$, для множини точок приєднання графа X до графа Y;
2) граф Y один з графів K_5 чи $K_{3,3}$, можливо без ребра, вкладений в тор σ, а інший граф X роду 1 є 2-мінімальним відносно множини точок приєднання на недвоклітці $\sigma \setminus Y$ із нульовими характеристиками θ, $\partial\theta$ множини точок приєднання графа X до графа Y, граф Z відсутній;

3) граф Y містить частину гомеоморфну K_5 чи $K_{3,3}$ (можливо із кількома додатковими ребрами) вкладений в тор σ, граф Z — проста зірка, граф X є площинною квазізіркою із центральним графом M на двох вершинах, яка не є 2-мінімальним графом на недвоклітці s, $s \in \sigma \setminus Y$, причому існує, принаймні одна, пара вершин простої зірки Z, сформована із елементів множини приєднання графа X до графа Y, що розділяє на ∂s пару кінцевих вершин з множини приєднання графа X до графа Y.

Список використаних джерел:

1. Хоменко М. П. φ-перетворення графів. Київ, 1971. 378 с.
2. Петренюк В. І. Построение графов-обструкций ограниченного ориентируемого рода. *XVI Международная конференция «Проблемы теоретической кибернетики».* Нижний Новгород, 2011. С. 363–368.
3. Hur Suhjin. The Kuratowski covering conjecture for graphs of order less than 10. PhD dissertation, Ohio State University, 2008.
4. Петренюк В. І. Структура 28-ми 9-ти вершинних графів-обструкцій тора. *Математичне та комп'ютерне моделювання.* Серія фізико-математичних наук. Кам'янець-Подільський національний університет імені Івана Огієнка. 2017. Т. 16. С. 145–151.

STRUCTURE OF 20 9-VERTECES GRAPHS OBSTRUCTIONS FOR TORUS

Structure all 9-verteces graphs obstructions for torus was found.

Key words: *structure, 9-verteces graph obstructions, torus, φ-transformation.*

Одержано 14.01.2019

УДК 517.9
DOI: 10.32626/2308-5878.2019-19.112-118

О. О. Покутний, д-р фіз.-мат. наук
Інститут математики НАН України, м. Київ

ГОМОКЛІНІЧНИЙ ХАОС ТА РІВНЯННЯ НАВ'Є–СТОКСА

У роботі розглядається збурена система рівнянь Нав'є–Стокса, яка переписується у вигляді операторно-диференціального рівняння. З допомогою отриманих апріорних оцінок для відповідного оператора встановлено властивість експоненціальної дихотомії для породжуючого однорідного рівняння. Отримано необхідні та достатні умови існування обмежених на всій осі розв'язків породжуючого лінійного однорідного рівняння. Відповідна множина розв'язків представляється з допомогою побудованого оператора Гріна. Для нелінійної системи рівнянь Нав'є–Стокса введено операторне рівняння для породжуючих елементів. З допомогою операторного рівняння для породжуючих елементів отримано необхідну умову біфуркації розв'язків рівняння Нав'є–Стокса. Необхідно знайти такий обмежений на всій осі розв'язок системи рівнянь, який перетворюється у породжуючий обмежений розв'язок відповідного однорідного рівняння, коли $\varepsilon = 0$. У роботі отримано достатню умову існування обмеженого на всій осі розв'язку системи рівнянь Нав'є–Стокса. Побудовано ітеративні алгоритмі типу Ньютона–Канторовіча для його знаходження. З допомогою представлення встановлено оцінки відповідних розв'язків у просторах інтегровних функцій.

Ключові слова: *рівняння Нав'є–Стокса, гомоклінічний хаос, псевдообернений за Муром–Пенроузом оператор.*

Вступ. Добре відомо, що поняття експоненціальної дихотомії відіграє важливу роль у якісній теорії диференціальних рівнянь.

Слід відзначити деякі роботи присвячені отриманню умов існування обмежених розв'язків для різного класу диференціальних рівнянь у скінченновимірному та нескінченновимірному просторах [1–4].

Поняття експоненціальної дихотомії на додатній та від'ємній півосях дає можливість отримати умови за виконання яких відповідний оператор є Фредгольмовим. Це добре відома лема Палмера [4]. У даній роботі наведено ідея методу який застосовується для дослідження питання існування обмежених на всій осі розв'язків збуреної системи типу Нав'є–Стокса.

Постановка задачі. Розглянемо питання стосовно існування обмежених розв'язків для систем рівнянь Нав'є–Стокса. Розглянемо однорідне рівняння вигляду

$$\frac{\partial u(x,t)}{\partial t} - \mu\Delta u(x,t) + \varepsilon\sum_{i=1}^{3}(u_i(x,t)\frac{\partial u(x,t)}{\partial x_i} + \frac{\partial u_i(x,t)}{\partial x_i}u(x,t)) = 0, \quad (1)$$

або у розширеному вигляді

$$\frac{\partial u_j(x,t)}{\partial t} - \mu\Delta u_j(x,t) + \varepsilon\sum_{i=1}^{3}(u_i(x,t)\frac{\partial u_j(x,t)}{\partial x_i} + \frac{\partial u_i(x,t)}{\partial x_i}u_j(x,t)) = 0,\ j = 1,2,3$$

зі стандартними крайовими умовами:

$$divu = 0,\ u|_{\partial\Omega} = 0.$$

Розглянемо множину гладких скінченновимірних соленоїдальних векторів. Замкненість цієї множини за нормою $L_2(\Omega)$ позначимо H, а за нормою $H^1(\Omega)$ E. Припустимо, що P — проектор з $L_2(\Omega)$ на H [5].

Позначимо

$$F(u(x,t)) = -\mu\Delta u(x,t) + \varepsilon\sum_{i=1}^{3}(u_i(x,t)\frac{\partial u(x,t)}{\partial x_i} + \frac{\partial u_i(x,t)}{\partial x_i}u(x,t)) = 0.$$

Тоді

$$A(t)h(x,t) = F'(u(x,t))h(x,t) = -\mu\Delta h(x,t) + \varepsilon\sum_{i=1}^{3}[u_i(x,t)\frac{\partial h(x,t)}{\partial x_i} +$$

$$+h_i(x,t)\frac{\partial u(x,t)}{\partial x_i}] + \varepsilon\sum_{i=1}^{3}[\frac{\partial h_i(x,t)}{\partial x_i}u(x,t) + \frac{\partial u_i(x,t)}{\partial x_i}h(x,t)].$$

Розглянемо таку форму

$$(A(t)h(\cdot,t),\bar{h}(\cdot,t))_{L_2(\Omega)} = \mu(\nabla h(\cdot,t),\nabla\bar{h}(\cdot,t))_{L_2(\Omega)} +$$

$$+\varepsilon\sum_{i=1}^{3}[(u_i(\cdot,t)\frac{\partial h(\cdot,t)}{\partial x_i},\bar{h}(\cdot,t))_{L_2(\Omega)} + (h_i(\cdot,t)\frac{\partial u(\cdot,t)}{\partial x_i},\bar{h}(\cdot,t))_{L_2(\Omega)} +$$

$$+(\frac{\partial h_i(\cdot,t)}{\partial x_i}u(\cdot,t),\bar{h}(\cdot,t))_{L_2(\Omega)} + (\frac{\partial u_i(\cdot,t)}{\partial x_i}h(\cdot,t),\bar{h}(\cdot,t))_{L_2(\Omega)}].$$

Інтегруючи частинами та враховуючи крайові умови можемо отримати таку оцінку знизу (взявши супремум по t)

$$\sup_{t\in R_+}\frac{\mu}{2}\|\nabla h(\cdot,t)\|^2_{L_2(\Omega)} + \frac{\mu k}{2}\sup_{t\in R_+}\|h(\cdot,t)\|^2_{L_2(\Omega)} -$$

$$-3c_1\varepsilon\sup_{t\in R_+}\|h(\cdot,t)\|^2_{L_2(\Omega)} - c_2\varepsilon\sup_{t\in R_+}\|h(\cdot,t)\|^2_{L_2(\Omega)} -$$

$$-c_3\varepsilon\sup_{t\in R_+}\|h(\cdot,t)\|^2_{L_2(\Omega)} - c_4\varepsilon\sup_{t\in R_+}\|h(\cdot,t)\|^2_{L_2(\Omega)},$$

де

$$c_1 = \sup_{(x,t)\in\Omega\times(0,+\infty)}\max_{i=1,2,3}|u_i(x,t)|,$$

$$c_2 = \sqrt{3} \sup_{(x,t)\in\Omega\times(0,+\infty)} \max_{i,\,j=1,2,3} |\frac{\partial u_j(x,t)}{\partial x_i}|,$$

$$c_3 = \sup_{(x,t)\in\Omega\times(0,+\infty)} \max_{j=1,2,3} |u_j(x,t)|,$$

$$c_4 = \sum_{i=1}^{3} \sup_{(x,t)\in\Omega\times(0,+\infty)} |\frac{\partial u_i(x,t)}{\partial x_i}|.$$

Таким чином для достатньо малого ε справедлива така нерівність

$$\sup_{t\in R_+} (A(t)h(\cdot,t), \overline{h}(\cdot,t))_{L_2(\Omega)} \geq L \sup_{t\in R_+} \| h(\cdot,t) \|^2_{W_2^1(\Omega)},$$

де L — додатна стала. Аналогічні оцінки можна отримати для від'ємної напівосі. Звідси випливає, що спектр оператора A не перетинається з уявною віссю. Таким чином неоднорідну крайову задачу можна записати у такому вигляді:

$$\begin{aligned} &u_t(t,\varepsilon) + A(t)u(t,\varepsilon) + \varepsilon F_1(u(t,\varepsilon)) = \\ &= u_t(t,\varepsilon) + F'(u(t,\varepsilon))u(t,\varepsilon) + \varepsilon F_1(u(t,\varepsilon)) = f(t). \end{aligned} \tag{2}$$

Тут

$$F_1(u(t,\varepsilon)) = -\sum_{i=1}^{3} (u_i(x,t)\frac{\partial u(x,t)}{\partial x_i} + u(x,t)\frac{\partial u_i(x,t)}{\partial x_i}).$$

Таким чином, коли $\varepsilon = 0$, розглянута крайова задача допускає експоненціальну дихотомію та можна застосувати добре розвинену теорію напівгруп до дослідження питання існування обмежених на всій осі розв'язків. Дослідимо більш детально розглянуту задачу.

Лінійний випадок. Породжуюче рівняння. Розглянемо лінійну задачу у тому випадку, коли $\varepsilon = 0$:

$$\frac{du_0(t)}{dt} = A(t)u_0(t) + f(t), \tag{3}$$

де вектор-функція $f(t)$ діє з R у простір Гільберта

$$W_2^1(\Omega),\ f(t) \in BC(R, W_2^1(\Omega)),$$

$$BC(R, W_2^1(\Omega)) := \{ f(\cdot) : R \to W_2^1(\Omega),\ f(\cdot) \in C(R, W_2^1(\Omega)),$$

$|||\ f\ ||| = \sup_{t\in R} \| f(t) \|_{W_2^1(\Omega)} < \infty\}$ — банахів простір функцій, неперервних та обмежених на R із значенням у соболєвському просторі $W_2^1(\Omega)$. Нехай $T(t, s)$ — еволюційний оператор [6, 7], асоційований з однорідним рівнянням, що є експоненціально дихотомічним [7, 8] на півосях з проекторнозначними функціями $P(t)$, $Q(t)$. Основний результат для лінійної задачі наступний.

Теорема 1. Нехай $T(t, s)$ сильно неперервний еволюційний оператор однорідного рівняння. Припустимо, що виконано такі умови:

1) $T(t, s)$ є експоненціально дихотомічним на півосях з проекторнозначними оператор-функціями $P(t)$, $Q(t)$, відповідно;
2) оператор $D = P(0) – (I – Q(0))$ є узагальнено-оборотним [9–11].

Тоді:

1) для того, щоб існували обмежені на всій осі розв'язки рівняння, необхідно та достатньо, щоб $f \in BC(R, B)$ задовольняла умові

$$\int_{-\infty}^{+\infty} H(t) f(t) dt = 0, \tag{4}$$

де $H(t) = P_{N(D^*)} P(0) T(0, t)$;

2) за виконання умови (4), розв'язки рівняння (3) мають такий вигляд:

$$u_0(t, c) = T(t, 0) P(0) P_{N(D)} c + (G[f])(t, 0),\ c \in B, \tag{5}$$

де

$$(G[f])(t, s) =$$

$$= \begin{cases} \int_s^t T(t,\tau)P(\tau)f(\tau)d\tau - \int_t^{+\infty} T(t,\tau)(I-P(\tau))f(\tau)d\tau + \\ +T(t,s)P(s)D^-[\int_s^{+\infty} T(s,\tau)(I-P(\tau))f(\tau)d\tau + \int_{-\infty}^{s} T(t,\tau)Q(\tau)f(\tau)d\tau], t \geq s, \\ \int_{-\infty}^t T(t,\tau)Q(\tau)f(\tau)d\tau - \int_t^s T(t,\tau)(I-Q(\tau))f(\tau)d\tau + \\ +T(t,s)(I-Q(s))D^-[\int_s^{+\infty} T(s,\tau)(I-P(\tau))f(\tau)d\tau + \int_{-\infty}^{s} T(t,\tau)Q(\tau)f(\tau)d\tau], s \geq t. \end{cases}$$

узагальнений оператор Гріна задачі про обмежені на всій осі розв'язки, $P_{N(D)} = I - D^- D, P_{N(D^*)} = I - DD^-$ — проектори на ядро та коядро оператора D [11].

Нелінійний випадок. Будемо шукати обмежений розв'язок рівняння (2), який при $\varepsilon = 0$ перетворюється у розв'язок породжуючого рівняння (3) $u(t, 0) = u_0(t)$. Ця задача може бути розв'язана за допомогою операторного рівняння:

$$F(c) = \int_{-\infty}^{+\infty} H(t) F_1(u_0(t, c)) dt = 0. \tag{6}$$

Теорема 2. (необхідна умова). Припустимо, що однорідне рівняння є експоненціально-дихотомічним на півосях з проекторнозначними оператор-функціями $P(t)$, $Q(t)$ відповідно, а нелінійне рів-

няння (2) має обмежений розв'язок який при $\varepsilon = 0$ перетворюється у один з розв'язків породжуючого рівняння (3) з елементом $c = c^0$: $u(t,0) = u_0(t, c^0)$. Тоді елемент c^0 повинен задовольняти рівняння для породжуючих елементів (6).

Достатню умову можна отримати з допомогою оператора

$$B_0 = \int_{-\infty}^{+\infty} H(t)A_1(t)T(t,0)P(0)P_{N(D)}dt : B \to B,\ A_1(t) = F_1^{(1)}(v)\,|_{v=u_0,\varepsilon=0}$$

(похідна Фреше).

Теорема 3. (достатня умова). Припустимо, що однорідне рівняння є експоненціально-дихотомічним на півосях з проекторнозначними оператор — функціями $P(t)$, $Q(t)$ відповідно. Нехай для оператора B_0 виконано такі умови:

1) оператор B_0 узагальнено-оборотний;

2) $P_{N(B_0^*)}P_{N(D^*)}P(0) = 0$.

Тоді, для довільного елемента $c = c^0$, який задовольняє рівняння для породжуючих елементів (6), існує принаймні один обмежений розв'язок нелінійного рівняння (2). Його можна знайти за допомогою такого ітераційного процесу

$$\overline{y}_{k+1}(t,\varepsilon) = \varepsilon G[Z(u_0(\tau, c^0) + y_k, \tau, \varepsilon)](t,0),$$

$$c_k = -B_0^- \int_{-\infty}^{+\infty} H(t)\{A_1(t)\overline{y}_k(t,\varepsilon) + R(y_k(t,\varepsilon),t,\varepsilon)\}dt,$$

$$y_{k+1}(t,\varepsilon) = T(t,0)PP_{N(D)^c k} + \overline{y}_{k+1}(t,\varepsilon),$$

$$u_k(t,\varepsilon) = u_0(t,c^0) + y_k(t,\varepsilon), k = 0,1,2,..., y_0(t,\varepsilon) = 0,$$

$$u(t,\varepsilon) = \lim_{k\to\infty} u_k(t,\varepsilon).$$

Доведення проводиться аналогічним чином як у роботі [8].

Зауваження. Слід зазначити, що з теорем 2, 3 випливає наявність складної поведінки у крайової задачі для збуреної системи Нав'є–Стокса. А саме, за умов дихотомії (які випливають з отриманих оцінок), у системі спостерігається гомоклінічний хаос.

Оцінки розв'язків. Згідно представлення

$$y(t,\varepsilon) = T(t,0)P(0)P_{N(D)^c} + \overline{y}(t,\varepsilon),$$

справедливі такі оцінки

$$\| y(t,\varepsilon)\|_{L_2(\Omega)} \le \| T(t,0)P(0)P_{N(D)^c}\|_{L_2(\Omega)} + \| \overline{y}(t,\varepsilon)\|_{L_2(\Omega)} \le$$

$$\le Me^{-\alpha t} \| P_{N(D)^c}\|_{L_2(\Omega)} + \varepsilon \| G[F_1(u_0 + y)](t,0)\|_{L_2(\Omega)}.$$

$$\| G[F_1(u_0 + y)(t,0)] \|_{L_2(\Omega)} \leq \| G \|_{L_2(\Omega)} \| F_1 \|_{L_2(\Omega)} \| u_0(t, c^0) + y(t,\varepsilon) \|_{L_2(\Omega)} \leq$$

$$\leq \| G \|_{L_2(\Omega)} \| F_1 \|_{L_2(\Omega)} \| u_0(t, c^0) \|_{L_2(\Omega)} + \| G \|_{L_2(\Omega)} \| F_1 \|_{L_2(\Omega)} \| y(t,\varepsilon) \|_{L_2(\Omega)}.$$

Таким чином справедлива така оцінка

$$\| y(t,\varepsilon) \|_{L_2(\Omega)} \leq Me^{-\alpha t} \frac{\| P_{N(D)^c} \|_{L_2(\Omega)}}{1 - \varepsilon \| G \|_{L_2(\Omega)} \| F_1 \|_{L_2(\Omega)}} +$$

$$+ \varepsilon \frac{\| G \|_{L_2(\Omega)} \| F_1 \|_{L_2(\Omega)} \| u_0(t, c^0) \|_{L_2(\Omega)}}{1 - \varepsilon \| G \|_{L_2(\Omega)} \| F_1 \|_{L_2(\Omega)}}.$$

Висновки. Розглянуто умови біфуркації розв'язків збуреної системи рівнянь Нав'є–Стокса на всій осі. За допомогою побудованого узагальненого оператора Гріна у лінійному породжуючому випадку отримано оцінки норми розв'язку у просторі інтегровних функцій.

Список використаних джерел:

1. Baskakov A. G. Invertibility and the fredholm property of difference operators. *Mathematical notes*. 2000. 6 (67). P. 690–698.
2. Baskakov A. G. On differential and difference Fredholm operators. *Reports of Mathematics.* 2007. 2 (76). P. 669–672.
3. Boichuk A. A. Solutions of weakly nonlinear differential equations bounded on the whole line. *Nonlinear Oscillations.* 1999. 1 (2). P. 3–10.
4. Palmer K. J. Exponential dichotomies and transversal homoclinic points. *Journ. of Diff. Eq*. 1984. 55. P. 225–256.
5. Levitan B. M., Gikov V. V. Almost periodic functions and differential equations. M.: MGU, 1978. 205 p.
6. Krein S. G. Linear differential equations in the Banach space. M. : Science, 1967. 464 p.
7. Henry D. Geometric theory of semilinear parabolic equations. M. : World, 1985. 376 p.
8. Pokutnyi A. A. Bounded solutions of linear and weakly nonlinear differential equations in Banach space with unbounded linear part. *Diff. Eq.* 2012. 6(48). P. 803–813.
9. Moore E. H. On the Reciprocal of the General Algebraic Matrix (Abstract). *Bull. Amer. Math. Soc*. 1920. 26. P. 394–395.
10. Penrose R. A. Generalized Inverse for Matrices. *Proc. Cambridge Philos. Soc.* 1955. 51. P. 406–413.
11. Boichuk A. A., Samoilenko A. M. Generalized Inverse Operators and Fredholm Boundary-Value Problems 2nd ed. Berlin/Boston : Walter De Gruyter GmbH, 2016. 296 p.

HOMOCLINIC CHAOS AND NAVIER STOKES EQUATIONS

In this paper we consider a perturbed system of Navier–Stokes equations, which is rewritten in the form of an operator-differential equation. Using the obtained a priori estimates for the corresponding operator, the property of the exponential dichotomy for a generating homogeneous equation is established.

The necessary and sufficient conditions for the existence of solutions of a generating linear homogeneous equation bounded on the entire axis are obtained. The corresponding set of solutions is represented by the constructed Green operator. For a nonlinear Navier–Stokes equation, we introduce the operator equation for generating elements. Using the operator equation for the generating elements, we obtain the necessary condition for the bifurcation of the solutions of the Navier–Stokes equation. It is necessary to find such a solution of the system of equations that is bounded on the entire axis, which transforms into a generating bounded solution of the corresponding homogeneous equation when $\varepsilon = 0$. In this paper we obtain a sufficient condition for the existence of a solution of the Navier–Stokes equation bounded on the entire axis. An iterative Newton-Kantorovich type algorithm for its finding was constructed. With the help of representation, estimates of the corresponding solutions in the spaces of integrable functions are established.

Key words: *Navier–Stokes equation, homoclinic chaos, Moore–Penrose pseudoinverse operator.*

Одержано 24.01.2019

УДК 519.7
DOI: 10.32626/2308-5878.2019-19.118-124

О. Д. Поліщук, канд. фіз.-мат. наук

Інститут прикладних проблем механіки і математики
імені Я. С. Підстригача НАН України, м. Львів

ЦЕНТРАЛЬНІСТЬ У СКЛАДНИХ МЕРЕЖАХ ТА ПОСЕРЕДНИЦТВО У МЕРЕЖЕВИХ СИСТЕМАХ

Аналізуються концепції центральності та впливу вузлів складних мереж для визначення їх важливості у структурі системи. Вводяться поняття міри, області та потужності посередництва вузлів та ребер мережі для ідентифікації їх важливості у процесі функціонування мережевих систем. Ці показники кількісно виражають ступінь сприяння відповідного елемента рухові потоків у системі та визначають втрати, які її очікують у разі блокування цього вузла або ребра чи цілеспрямованої атаки на нього. Аналогічні поняття посередництва вводяться для визначення функціональної важливості окремих підсистем мережевих систем. Наводяться приклади практичного застосування отриманих результатів.

Ключові слова: *складна мережа, мережева система, центральність, посередництво.*

Вступ. Однією з основних концепцій теорії складних мереж є так звана центральність (centrality) вузла, яка дозволяє визначати його важливість у мережі: найбільш впливові особи у соціальних мережах,

ключові вузли у Інтернеті та транспортних мережах тощо [1, 2]. Однак, поняття «важливість» може мати різний зміст, що призвело до появи багатьох визначень терміну «центральність». Перерахуємо найбільш вживані показники центральності вузла в складній мережі (СМ):

- центральність за ступенем (degree centrality) визначається кількістю ребер, які безпосередньо поєднують даний вузол з іншими вузлами мережі (був першим та локальним показником центральності) [3];
- центральність близькості (closeness centrality) обчислюється, як середня довжина найкоротшого шляху між вузлом та всіма іншими вузлами мережі [4];
- центральність посередництва (betweenness centrality) обчислюється, як відношення кількості найкоротших шляхів між всіма вузлами, які проходять через даний вузол, до загальної кількості найкоротших шляхів мережі [5];
- власна центральність або центральність за власним вектором (eigenvector centrality) визначається через власні значення матриці суміжності бінарної мережі, яка описує структуру СМ [1];
- перколяційна центральність (percolation centrality) визначається важливістю вузла для сприяння перколяції мережі [6];
- крос-клік (cross-clique) центральність обчислюється кількістю клік, з якими пов'язаний вузол (кліка — підмережа у якій кожний вузол пов'язаний зі всіма іншими вузлами цієї підмережі) [7].

Можна назвати ще цілу низку визначень центральності вузла в мережі — подібні власній центральності Катца (Katz) [8] та PageRank [9], подібна центральності близькості гармонійна (harmonic) центральність [10], центральність Фрімана (Freeman) та альфа (alpha) центральність [3] тощо. При цьому одне значення центральності може суперечити іншому та центральність, важлива для однієї задачі, може бути несуттєвою для іншої. Цей феномен був яскраво підтверджений Д. Кракхардом [11], який навів приклад простої мережі, до складу якої входить 10 вузлів, для яких центральність за ступенем, посередництвом та близькістю приймали абсолютно різні значення, тобто дали три різні вибори найбільш важливих вузлів у її структурі. Звідси слідує, що перераховані вище визначення центральності вузла мають достатньо відносне значення. Це стало причиною введення поряд із поняттями центральності пов’язаних з ними показників впливу вузлів на мережеву структуру. Основними показниками впливу вузла є його доступність (accessibility) та очікувана сила (expected force) [12]. Доступність вузла визначається кількістю вузлів до яких з нього можна перейти за визначений проміжок часу. Очікувана сила впливу вузла визначається кількістю вузлів, до яких можна з нього перейти за два або більше кроків руху (крок — перехід одним ребром мережі). Очевидно, що показники

центральності та впливу вузла визначаються виключно властивостями структури та є характеристиками цієї структури, а не системи загалом.

У праці [13] як показники функціональної важливості вузла мережевої системи (МС) були введені параметри вхідного та вихідного впливу вузла на систему, які визначаються обсягами потоків, які приймаються вузлом або генеруються у ньому, областей та потужностей вхідного та вихідного впливу, які визначаються множинами вузлів-приймачів та генераторів потоків з даного вузла СМ та кількістю елементів цих областей. Ці поняття дозволяють кількісно оцінити участь окремого вузла як приймача або генератора потоків в процесі функціонування системи та його участь у цьому процесі. Інший показник важливості взаємодії вузла з МС, який вводиться та досліджується у цій статті, є мірою його сприяння транзиту потоків мережею.

Посередництво елементів мережевих систем. Однією з найбільш вживаних поряд із центральністю за ступенем (або ступенем вузла) у теорії складних мереж є центральність посередництва. Можливо термін «посередництво» є найбільш вдалим для визначення участі елемента МС у процесі спільного функціонування та взаємодії усіх вузлів мережі або певної її частини. Тому для визначення функціональної важливості вузла або ребра СМ в системі вживатимемо саме термін «посередництво».

Нехай $\mathbf{V} = \{v_{ij}\}_{i,j=1}^{N}$ — потокова матриця суміжності МС [13], значення елементів якої є рівними обсягами потоків, які проходять ребром мережі, яке пов'язує вузли n_i та n_j за період $[0,T]$, N — кількість вузлів мережі. Позначимо $P_{ij}^{K_{ij}} = \{p_{ij}^{k}\}_{k=1}^{K_{ij}}$ — сукупність шляхів, які поєднують вузли-генератори та вузли-приймачі потоків МС, та містять, як елемент, ребро (n_i, n_j), $i, j = \overline{1,N}$. Нехай v_{ij}^{k} — обсяг потоків, які пройшли шляхом p_{ij}^{k} від вузла-генератора до вузла-приймача, а отже і ребром (n_i, n_j), за період $[0,T]$. Тоді величина

$$V_{ij}^{K_{ij}} = \sum_{k=1}^{K_{ij}} v_{ij}^{k}$$

визначає сумарний обсяг потоків, які пройшли сукупністю шляхів $P_{ij}^{K_{ij}}$, а отже і ребром (n_i, n_j), за цей же проміжок часу. Величину

$$\Phi_{ij} = V_{ij}^{K_{ij}} / s(\mathbf{V}),\ s(\mathbf{V}) = \sum_{i,j=1}^{N} v_{ij},$$

яка визначає питому вагу потоків, що проходять ребром (n_i, n_j) за період $[0,T]$, називатимемо мірою посередництва цього ребра в процесі функціонування МС.

Множину L_{ij} усіх вузлів СМ, які лежать на шляхах із сукупності $P_{ij}^{K_{ij}}$, називатимемо областю посередництва, а кількість η_{ij} цих вузлів — потужністю посередництва ребра (n_i, n_j) (рис. 1).

Параметри міри, області та потужності посередництва ребра (n_i, n_j) є глобальними характеристиками його важливості у процесі функціонування МС, $i, j = \overline{1, N}$. Вони, зокрема, визначають, яким чином блокування цього ребра вплине на роботу області його посередництва, величину цієї області і, внаслідок цього, — всієї системи [14].

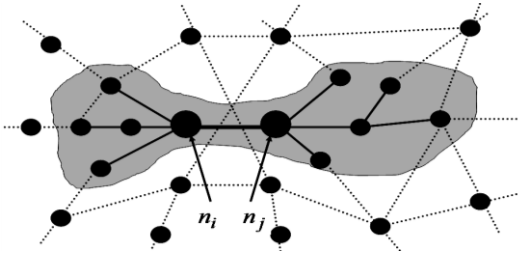

Рис. 1. *Область посередництва ребра* (n_i, n_j) *у процесі функціонування МС*

Позначимо $P_i^{K_i} = \{p_i^k\}_{k=1}^{K_i}$ — сукупність шляхів, які поєднують вузли-генератори та вузли-приймачі потоків МС, та проходять через вузол n_i, $i = \overline{1, N}$. Нехай v_i^k — обсяг потоків, які пройшли шляхом p_i^k від вузла-генератора до вузла-приймача, а отже і через вузол n_i, за період $[0, T]$. Тоді величина

$$V_i^{K_i} = \sum_{k=1}^{K_i} v_i^k$$

визначає сумарний обсяг потоків, які пройшли сукупністю шляхів $P_i^{K_i}$, а отже і через вузол n_i, за цей же проміжок часу. Величину

$$\Phi_i = V_i^{K_i} / s(\mathbf{V}),$$

яка визначає питому вагу потоків, що проходять через вузол n_i за період $[0, T]$, називатимемо мірою посередництва цього вузла в процесі функціонування МС. Множину M_i усіх вузлів СМ, які лежать на шляхах із сукупності $P_i^{K_i}$, називатимемо областю посередництва, а кількість η_i цих вузлів — потужністю посередництва вузла n_i.

Параметри міри, області та потужності посередництва вузла n_i є глобальними характеристиками його важливості у процесі функціо-

нування МС, $i = \overline{1, N}$. Вони, зокрема, визначають, яким чином блокування цього вузла вплине на роботу області його посередництва, величину цієї області і, внаслідок цього, — всієї системи.

Посередництво підсистем мережевих систем. Не менш важливими для аналізу процесу функціонування МС є параметри посередництва окремих її підсистем, які визначимо наступним чином. Позначимо $P_S^{K_s} = \{p_S^k\}_{k=1}^{K_s}$ — сукупність шляхів, які поєднують вузли-генератори та вузли-приймачі потоків МС, та проходять через елементи підсистеми S. Нехай v_S^k — обсяг потоків, які пройшли шляхом p_S^k від вузла-генератора до вузла-приймача, а отже і через елементи підсистеми S, за період $[0, T]$. Тоді величина

$$V_S^{K_s}(t) = \sum_{k=1}^{K_s} v_S^k(t)$$

визначає сумарний обсяг потоків, які пройшли сукупністю шляхів $P_S^{K_s}$, а отже і через елементи підсистеми S, за цей же проміжок часу. Величину

$$\Psi_S = V_S^{K_i}(t) / s(\mathbf{V}(t)),$$

яка визначає питому вагу потоків, що проходять через елементи підсистеми S за період $[0, T]$, називатимемо мірою посередництва цієї підсистеми в процесі функціонування МС.

Множину M_S усіх вузлів СМ, які лежать на шляхах із сукупності $P_S^{K_s}$, називатимемо областю посередництва (рис. 2), а кількість η_S цих вузлів — потужністю посередництва підсистеми S.

Параметри міри, області та потужності посередництва підсистеми S є глобальними характеристиками її важливості у процесі функціонування МС. Вони зокрема визначають, яким чином блокування цієї підсистеми вплине на роботу області її посередництва, величину цієї області і, внаслідок цього, — всієї системи. Окрім того, невеликі значення параметрів посередництва підсистеми S також можуть свідчити про те, що вона утворює спільноту в межах МС.

Поведінка похідних параметрів посередництва елементів та підсистем МС дозволяє визначати тенденції та швидкість зростання або падіння їх важливості у процесі функціонування МС. Для глибшого дослідження параметрів посередництва складових системи доцільно використовувати методи коротко- та довгострокового прогнозування їх поведінки [15].

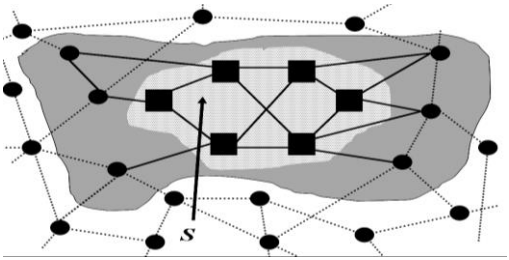

Рис. 2. *Область посередництва підсистеми S в процесі функціонування МС*

Висновки. Визначені у статті параметри посередництва складових мережевих систем дозволяють ідентифікувати найважливіші для роботи МС елементи і підсистеми та сприяють кращому розумінню процесів, які перебігають у них. Ці параметри дають можливість складати значно реалістичніші сценарії потенційних атак на систему та будувати надійніші засоби її захисту. Отримані результати можуть бути використані для зменшення уразливості МС від негативних зовнішніх та внутрішніх впливів, розробки новітніх методів захисту інформаційних та безпекових систем, підвищення ефективності функціонування транспортних і промислових мереж різного типу та призначення і т. ін.

Список використаних джерел:

1. Bonacich P. Power and Centrality: A Family of Measures. *American Journal of Sociology*. 1987. Vol. 92 (5). P. 1170–1182.
2. Borgatti S. P. Centrality and network flow. *Social Networks*. 2005. Vol. 27 (1). P. 55–71.
3. Freeman L. C. Centrality in social networks conceptual clarification. *Social networks.* 1979. Vol. 1 (3). P. 215–239.
4. Bavelas A. Communication patterns in task-oriented groups. *Journal of American Acoustic Society*. 1950. Vol. 22 (6). P. 725–730.
5. Freeman L. C. A set of measures of centrality based upon betweenness. *Sociometry*. 1977. Vol. 40. P. 35–41.
6. Piraveenan M. Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks. *PLOS ONE.* 2013. Vol. 8 (1). e53095.
7. Faghani M., Nguyen U. T. A Study of XSS Worm Propagation and Detection Mechanisms in Online Social Networks. *IEEE Trans. Inf. Forensics and Security*. 2013. Vol. 8 (11). P. 1815–1826.
8. Katz L. A New Status Index Derived from Sociometric Index. *Psychometrika*. 1953. Vol. 18 (1). P. 39–43.
9. Bonacich P., Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks.* 2001. Vol. 23 (3). P. 191–201.
10. Marchiori M., Latora V. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications.* 2000. Vol. 285 (3–4). P. 539–546.
11. Krackhardt D. Assessing the Political Landscape: Structure, Cognition, and Power in Organizations. *Administrative Science Quarterly*. 1990. Vol. 35 (2). P. 342–369.

12. Glenn L. Understanding the influence of all nodes in a network. *Scientific Reports.* 2015. Vol. 5. 8665.
13. Polishchuk O. Flow Models of Complex Network Systems. *Intern. Scientific-Practical Conf. on Problems of Infocommunications. Science and Technology*. 2018. P. 317–322.
14. Polishchuk O., Polishchuk D. Monitoring of flow in transport networks with partially ordered motion. *XXIII Conf. Carpenko physics and mechanics institute*, NASU. 2013. P. 326–329.
15. Polishchuk D., Polishchuk O., Yadzhak M. About complex evaluation of hierarchically-network systems. *IVth Conf. «Knowledge — Ontology — Theory»*. 2013. P.68–79.

CENTRALITY IN COMPLEX NETWORKS AND BETWEENNESS IN NETWORK SYSTEMS

The concepts of centrality and influence of complex network nodes are analyzed for the purpose of determining their importance in the systems structure. The notions of measure, domain and power of betweenness of network nodes and edges are introduced to identify their importance in the operation process of network systems. These indicators quantitatively express the degree of assistance of the corresponding element for the motion of flows in the system and determine the losses that are expected in the case of blocking this node or edge or targeted attack on it. Similar notions of betweenness are introduced to determine the functional importance of separate subsystems of network systems. Examples of practical use of the obtained results are given.

Key words: *complex network, network system, centrality, betweenness.*

Одержано 21.01.2019

УДК 517.5
DOI: 10.32626/2308-5878.2019-19.125-131

М. Ю. Савкіна, канд. фіз.-мат. наук,
Інститут математики НАН України, м. Київ

РІВНІСТЬ ОЦІНОК МНК ТА ЕЙТКЕНА МОДЕЛІ ЛІНІЙНОЇ РЕГРЕСІЇ У ВИПАДКУ ГЕТЕРОСКЕДАСТИЧНИХ ВІДХИЛЕНЬ

В роботі у випадку гетероскедастичних незалежних відхилень досліджується модель лінійної регресії, функція якої має вигляд $f(t) = at + b$, де a та b — невідомі параметри. Наближені значення (спостереження) функцій $f(t)$ реєструються в рівновіддалених точках відрізка $[0,1]$. Сформульовано теорему 1, яка дає умови на дисперсії відхилень, при яких оцінка Ейткена параметра a збігається з його оцінкою МНК. При цих умовах оцінки Ейткена та МНК параметра $M_2 = \max\left\{\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}\right\}$. не будуть збігатися. Також сформульовано теорему 2, яка дає умови для збігу оцінки Ейткена та оцінки МНК параметра b. На підставі теорем 1 та 2 в даній роботі досліджено властивості дисперсій відхилень, які надають рівність цим оцінкам окремо для параметра a та для параметра b. Показано, для рівності оцінок Ейткена та МНК параметра a відхилення будуть мати найбільшу та найменшу дисперсію в двох сусідніх точках спостереження, розташованих в середині відрізка $[0,1]$, для рівності оцінок параметра b — в околі точки 2/3. Знайдено асимптотичні значення дисперсій всіх відхилень, якщо відношення найбільшої до найменшої дисперсії прямує до нескінченності. Доведено, що в цьому випадку дисперсії всіх відхилень будуть не більше найменшої дисперсії ніж у 3 рази для параметра a та не більше ніж у 5 разів для параметра b.

Ключові слова: *метод найменших квадратів, регресійна модель, оцінка Ейткена.*

Вступ. У класичній регресії передбачається, що відхилення в регресійній моделі гомоскедастичні та не корелюють одне з іншим. Це доволі жорстка умова, яка досить часто не виконується. В зв'язку з цим дослідження моделі, в якій відхилення корельовані, або принаймні, гетероскедастичні, має великий інтерес. Проте в тому випадку, коли коваріаційна матриця відхилень не є одиничною, оцінка звичайного методу найменших квадратів (МНК) невідомих параметрів моделі хоч і буде незміщеною та спроможною, вже не буде ефективною. Так виник зва-

жений МНК [1, с. 78]; оцінка невідомих параметрів моделі, яка отримується за допомогою зваженого МНК, називається оцінкою Ейткена.

Однак використання ефективної оцінки Ейткена передбачає знання коваріаційної матриці відхилень, яка на практиці як правило невідома. Тому доводиться користуватися оцінкою МНК. Отже знаходження випадків, коли ці оцінки збігаються, має велике значення.

У [2, с. 610] доведено теорему, яка дає необхідну і достатню умову на матрицю плану [3, с. 49] для збігу оцінки МНК та оцінки Ейткена. Зауважимо, що в цій теоремі оцінки збігаються для всіх параметрів одночасно.

Розглянемо таку модель лінійної регресії

$$y_i = at_i + b + \varepsilon_i, i = 0,1,\ldots,n, \tag{1}$$

де $t_i = \frac{i}{n}, i = 0,1,\ldots,n,$ — точки спостереження, $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n$ — незалежні у сукупності випадкові величини з $E\varepsilon_i = 0$ та $D\varepsilon_i = \gamma_i \sigma^2, i = 0,1,\ldots,n,$ $m = \gamma_{\frac{2n+2}{3}}$, та b — невідомі параметри, які підлягають оцінюванню.

Поставимо задачу знайти умови на дисперсії відхилень, при яких збігаються оцінка МНК та оцінка Ейткена кожного параметру моделі (1) окремо (одночасно для обох параметрів ці оцінки будуть збігатися тільки у випадку всіх рівних дисперсій). Виходячи з загальних формул для оцінки МНК та оцінки Ейткена параметрів лінійної регресійної моделі, в роботі [4] доведено наступні теореми для відповідних оцінок параметрів a та $(n+1)$ моделі (1).

Теорема 1. Якщо в моделі (1) умова i не виконується, то оцінка МНК та оцінка Ейткена параметра a збігаються тоді і тільки тоді, коли:

1) у випадку парного n

$$(n+1)$$

$\gamma_{\frac{n}{2}}, \gamma_{\frac{n}{2}+1}$ — будь-які, але $\gamma_{\frac{n}{2}} \neq \gamma_{\frac{n}{2}+1}$;

2) у випадку непарного n

$$\gamma_i = \frac{(2i-n)\gamma_{\frac{n-1}{2}}\gamma_{\frac{n+1}{2}}}{\left(i - \frac{n-1}{2}\right)\gamma_{\frac{n-1}{2}} + \left(i - \frac{n+1}{2}\right)\gamma_{\frac{n+1}{2}}}, \tag{2}$$

$$i = 0,1,\ldots,\frac{n-3}{2},\frac{n+3}{2},\ldots,n,$$

$\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}$ — будь-які, але $\gamma_{\frac{n-1}{2}} \neq \gamma_{\frac{n+1}{2}}$.

Використовуючи формулу (2), дослідимо властивості γ_i, $i = 0, 1, \ldots, \frac{n-3}{2}, \frac{n+3}{2}, \ldots, n.$

Позначимо $m = \min\left\{\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}\right\}$, $M = \max\left\{\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}\right\}$. Маємо

$$M = \alpha m,\ \alpha \geq 1.$$

Нехай $m = \gamma_{\frac{n-1}{2}}$. Представимо γ_i у вигляді

$$\gamma_i = \frac{2i-n}{\left(i-\frac{n-1}{2}\right)\alpha^{-1} + \left(i-\frac{n+1}{2}\right)} \cdot m. \tag{3}$$

Оскільки $0 < \alpha^{-1} < 1$, з (3) маємо

$$m < \gamma_i < \min\left\{2\left(1+\frac{1}{2i-n-1}\right)m, M\right\}, i = 0,1,\ldots,\frac{n-3}{2}, \frac{n+3}{2}, \ldots, n.$$

Далі,

$$\lim_{\alpha\to\infty} \gamma_i = 2\left(1+\frac{1}{2i-n-1}\right)m < 2m,\ i = 0,1,\ldots,\frac{n-3}{2},$$

$$\lim_{\alpha\to\infty} \gamma_i = 2\left(1+\frac{1}{2i-n-1}\right)m < 3m,\ i = \frac{n+3}{2},\ldots,n,$$

тобто якщо $m \ll M$, то

$$m < \gamma_i < 3m,\ i = 0,1,\ldots,\frac{n-3}{2}, \frac{n+3}{2}, \ldots, n. \tag{4}$$

Якщо $m = \gamma_{\frac{n+1}{2}}$, маємо

$$\gamma_i = \frac{2i-n}{\left(i-\frac{n-1}{2}\right) + \left(i-\frac{n+1}{2}\right)\alpha^{-1}} \cdot m.$$

У випадку $m \ll M$ також отримуємо (4), тобто жодне з γ_i крім $\gamma_{\frac{n-1}{2}}$, $\gamma_{\frac{n+1}{2}}$ не може бути більше найменшого з $\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}$ більше ніж у 3 рази і не може бути менше його.

Крім того,

$$\gamma_i > \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{n-1}{2}},$$

$$\gamma_i < \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{n+1}{2}},$$

$$i = 0,1,\ldots,\frac{n-3}{2},\frac{n+1}{2},\ldots,n\ .$$

Теорема 2. Якщо в моделі (1) умова $\gamma_i = \gamma, i = 0,1,\ldots,n,$ не виконується, то оцінка МНК та оцінка Ейткена параметра b збігаються тоді і тільки тоді, коли:

1) у випадку $n = 3k+1, k = 1,2,\ldots$

$$\gamma_i = \gamma_{\frac{2n-2}{3}}, i = 0,1,\ldots,\frac{2n-5}{3},\frac{2n+4}{3},\ldots,n,$$

$\gamma_{\frac{2n+1}{3}}, \gamma_{\frac{2n-2}{3}}$ — будь-які, але $\gamma_{\frac{2n+1}{3}} \neq \gamma_{\frac{2n-2}{3}}$;

2) у випадку $n = 3k,\ \ k = 1,2,\ldots$

$$\gamma_{\frac{2n+3l}{3}} = \frac{(3l-1)\gamma_{\frac{2n+3}{3}}\gamma_{\frac{2n}{3}}}{(l-1)\gamma_{\frac{2n+3}{3}} + 2l\gamma_{\frac{2n}{3}}}, l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}, \qquad (5)$$

$\gamma_{\frac{2n}{3}}, \gamma_{\frac{2n+3}{3}}$ — будь-які, але $\gamma_{\frac{2n-3}{3}} \neq \gamma_{\frac{2n}{3}}$;

3) у випадку $n = 3k+2,\ \ k = 1,2,\ldots$

$$\gamma_{\frac{2n+2+3l}{3}} = \frac{(3l+1)\gamma_{\frac{2n-1}{3}}\gamma_{\frac{2n+2}{3}}}{(l+1)\gamma_{\frac{2n-1}{3}} + 2l\gamma_{\frac{2n+2}{3}}}, l = -\frac{2n+2}{3},\ldots,-2,1,2,\ldots,\frac{n-2}{3}, \qquad (6)$$

$\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}$ — будь-які, але $\gamma_{\frac{2n-1}{3}} \neq \gamma_{\frac{2n+2}{3}}$.

У випадку $n = 3k,\ \ k = 1,2,\ldots,$ використовуючи формулу (5), дослідимо властивості $\gamma_i, i = 0,1,\ldots,\frac{2n-3}{3},\ \frac{2n+6}{3},\ldots,n.$

Позначимо $m_1 = \min\left\{\gamma_{\frac{2n}{3}}, \gamma_{\frac{2n+3}{3}}\right\},\ M_1 = \max\left\{\gamma_{\frac{2n}{3}}, \gamma_{\frac{2n+3}{3}}\right\}.$ Маємо

$$M_1 = \beta m_1,\ \beta \geq 1.$$

Нехай $m_1 = \gamma_{\frac{2n}{3}}$. Представимо $\gamma_{\frac{2n+3l}{3}}$ у вигляді

$$\gamma_{\frac{2n+3l}{3}} = \frac{3l-1}{(l-1)+2l\beta^{-1}} \cdot m_1. \qquad (7)$$

Оскільки $0 < \beta^{-1} < 1$, з (7) маємо

$$m_1 < \gamma_i < \min\left\{\left(3+\frac{2}{l-1}\right)m_1\ , M_1\right\}, l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}\ .$$

Далі,

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}} = \left(3+\frac{2}{l-1}\right)m_1 < 3m_1,\ l = -\frac{2n}{3},\ldots,-2,-1\ ,$$

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}} = \left(3+\frac{2}{l-1}\right)m_1 < 5m_1,\ l = 2,\ldots,\frac{n}{3},$$

тобто, якщо $m_1 \ll M_1$, то $m_1 < \gamma_{\gamma_{\frac{2n+3l}{3}}} < 5m_1,\ l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$

Якщо $m_1 = \gamma_{\frac{2n+3}{3}}$, маємо

$$\gamma_{\frac{2n+3l}{3}} = \frac{3l-1}{(l-1)\beta^{-1}+2l}\cdot m_1. \qquad (8)$$

Оскільки $0 < \beta^{-1} < 1$, з (8) маємо

$$m_1 < \gamma_{\frac{2n+3l}{3}} < \min\left\{\left(\frac{3}{2}-\frac{1}{2l}\right)m_1, M_1\right\}, l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$$

Далі,

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}} = \left(\frac{3}{2}-\frac{1}{2l}\right)m_1 < 2m_1,\ l = -\frac{2n}{3},\ldots,-2,-1\ ,$$

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}} = \left(\frac{3}{2}-\frac{1}{2l}\right)m_1 < \frac{3}{2}m_1,\ l = 2,\ldots,\frac{n}{3},$$

тобто, якщо $m_1 \ll M_1$, то $m_1 < \gamma_{\gamma_{\frac{2n+3l}{3}}} < 2m_1,\ l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$

Крім того,

$$\gamma_i > \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{2n}{3}},$$

$$\gamma_i < \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{2n+3}{3}},$$

$$i = 0,1,\ldots,\frac{2n-3}{3},\ \frac{2n+6}{3},\ldots,n.$$

У випадку $n = 3k+2,\ k = 1,2,\ldots,$ використовуючи формулу (6), дослідимо властивості $\gamma_i, i = 0,1,\ldots,\frac{2n-4}{3},\ \frac{2n+5}{3},\ldots,n.$

Далі, позначимо

$$m_2 = \min\left\{\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}\right\},\ M_2 = \max\left\{\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}\right\}.$$

Маємо

$$M_2 = \delta m_2,\ \delta \geq 1.$$

Нехай $m_2 = \gamma_{\frac{2n-1}{3}}$. Представимо $\gamma_{\frac{2n+3l}{3}}$ у вигляді

$$\gamma_{\frac{2n+2+3l}{3}} = \frac{3l+1}{(l+1)\delta^{-1}+2l}\cdot m_2.$$

Аналогічно попередньому випадку отримуємо, що $m_2 \ll M_2$, то

$$m_2 < \gamma_{\gamma_{\frac{2n+3l}{3}}} < 2m_2,\ l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$$

У випадку $m_2 = \gamma_{\frac{2n+2}{3}}$, маємо

$$\gamma_{\frac{2n+2+3l}{3}} = \frac{3l+1}{(l+1)+2l\delta^{-1}}\cdot m_2.$$

Якщо $m_2 \ll M_2$, то $m_2 < \gamma_{\gamma_{\frac{2n+3l}{3}}} < 5m_2$, $l = -\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}$.

Крім того,

$$\gamma_i > \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{2n-1}{3}},$$

$$\gamma_i < \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{2n+2}{3}},$$

$$i = 0,1,\ldots,\frac{2n-4}{3},\ \frac{2n+5}{3},\ldots,n.$$

Висновки. Доведено, що рівність оцінок МНК та Ейткена параметра a забезпечує родина векторів $(n+1)$-го порядку, i-й елемент вектора з цієї родини — дисперсія відхилення в точці t_i, $i = 0,1,\ldots,n$. Знайдено два сусідні елементи цього вектору, якіможуть приймати будь-які значення, а значення всіх інших елементів знаходяться між ними. Рівність оцінок МНК та Ейткена параметра b забезпечує інша родина векторів $(n+1)$-го порядку, у яких i-й елемент — також дисперсія відхилення в точці t_i, $i = 0,1,\ldots,n$.

Список використаних джерел:

1. Демиденко Е. З. Линейная и нелинейная регрессии. М. : Финансы и статистика, 1981. 302 с.

2. Андерсон Т. Статистический анализ временных рядов. М. : Мир, 1976. 756 с.
3. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. 336 с.
4. Савкіна М. Ю. Умови збігу оцінок МНК та Ейткена параметрів моделі лінійної регресії. *Журнал обчислювальної та прикладної математики.* 1980. № 3 (129). С. 34–42.

EQUALITY OF MNC AND AITKEN ESTIMATIONS OF LINEAR REGRESSION MODEL PAPER IN THE CASE OF HETEROSCEDASTIC DEVIATIONS

At the paper in the case of heteroscedastic independent deviations a linear regression model whose function has the form $f(t) = at + b$, where a and b unknown parameters, is studied. Approximate values (observations) of functions $f(t)$ are registered at equidistant points of the segment $[0,1]$. We formulate Theorem 1, which gives conditions on the variances of deviations, in which the Aitken estimation of parameter a coincides with its estimation of MNCs. Under these conditions, the Aitken and MNC estimations of the parameter b will not coincide. We also formulate Theorem 2, which gives the conditions for the coincidence of the Aitken estimation and the MNC estimation of parameter b. Based on Theorems 1 and 2, in this paper the properties of variances of deviations that give equality with these estimations separately for parameter a and for parameter b are investigated. It is shown that for equality estimations of Aitken and MNC of the parameter a the deviations will have the largest and smallest variance in two adjacent observation points located in the middle of the segment [0, 1], for the equality estimations of the parameter b — in the neighborhood of the point 2/3. The asymptotic values of the variances of all deviations are found, if the ratio of the largest to the smallest variance goes to infinity. It is proved that in this case, the variances of all deviations will be no more the smallest variance than 3 times for parameter a and not more than 5 times for parameter b.

Key words: *least square method, regression model, Aitken estimation.*

Одержано 15.02.2019

УДК 517.988
DOI: 10.32626/2308-5878.2019-19.132-137

В. В. Семёнов, д-р физ.-мат. наук

Киевский национальный университет
имени Тараса Шевченко, г. Киев

МОДИФИЦИРОВАННЫЙ ЭКСТРАГРАДИЕНТНЫЙ МЕТОД С ДИВЕРГЕНЦИЕЙ БРЭГМАНА ДЛЯ ВАРИАЦИОННЫХ НЕРАВЕНСТВ

Предложен новый метод экстраградиентного типа для решения вариационных неравенств с псевдомонотонными и липшицевыми операторами, действующими в конечномерном линейном нормированном пространстве. Доказана теорема сходимости метода и для случая монотонного оператора получены неасимптотические оценки эффективности метода.

Ключевые слова: *вариационное неравенство, монотонность, псевдомонотонность, условие Липшица, экстраградиентный метод, дивергенция Брэгмана.*

Введение. Наиболее известным обобщением метода проекции градиента для вариационных неравенств является экстраградиентный метод Г. М. Корпелевич [1]. Исследованию этого алгоритма посвящено большое количество публикаций. В частности, предлагались модификации алгоритма Г. М. Корпелевич с одним метрическим проектированием на допустимое множество [2–5]. Для вариационных неравенств одним из современных вариантов экстраградиентного метода является проксимальный зеркальный метод А. С. Немировского [6].

Настоящее сообщение посвящено изучению нового метода экстраградиентного типа для приближенного решения вариационных неравенств с псевдомонотонными и липшицевыми операторами, действующими в конечномерном линейном нормированном пространстве. Данный метод является модификацией субградиентного экстраградиентного алгоритма [3–5] с использованием дивергеции Брэгмана вместо евклидового расстояния. К предлагаемой схеме можно прийти и путем замены допустимого множества на специальные опорные для него полупространства во втором этапе проксимального зеркального метода А. С. Немировского [6]. Доказана теорема сходимости метода. А для случая монотонного оператора и компактного допустимого множества получены неасимптотические оценки эффективности.

Модифицированный экстраградиентный метод. Всюду далее работаем в конечномерном действительном линейном пространстве, обозначаемом буквой E. Это пространство снабдим нормой $\|\cdot\|$ (не

обязательно евклидовой). Двойственное пространство обозначим E^*. Для $a \in E^*$ и $b \in E$ будем обозначать через (a,b) значение линейной функции a в точке b. Двойственную норму на E^* обозначим $\|\cdot\|_*$.

Пусть C — непустое подмножество пространства E, A — оператор, действующий из E в E^*. Рассмотрим вариационное неравенство:

$$\text{найти } x \in C: \ (Ax, y-x) \geq 0 \quad \forall y \in C, \tag{1}$$

множество решений которого обозначим S.

Предположим, что выполнены следующие условия:

- множество $C \subseteq E$ — выпуклое и замкнутое;
- оператор $A: E \to E^*$ — псевдомонотонный и липшицевый с константой $L > 0$;
- множество S не пусто.

Заметим, что при данных условиях множество S выпуклое и замкнутое.

Введем необходимые для формулировки алгоритма конструкции. Пусть функция $\varphi: E \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ удовлетворяет условия:

- $\operatorname{int}\operatorname{dom}\varphi \subseteq E$ непустое выпуклое множество;
- φ непрерывно дифференцируема на $\operatorname{int}\operatorname{dom}\varphi$;
- если $\operatorname{int}\operatorname{dom}\varphi \ni x_n \to x \in \operatorname{bd}\operatorname{dom}\varphi$, то $\|\nabla\varphi(x_n)\|_* \to +\infty$;
- φ сильно выпукла относительно нормы $\|\cdot\|$ с константой сильной выпуклости $\sigma > 0$:

$$\varphi(a) \geq \varphi(b) - (\nabla\varphi(b), a-b) + \tfrac{\sigma}{2}\|a-b\|^2 \quad \forall a \in \operatorname{dom}\varphi, \ b \in \operatorname{int}\operatorname{dom}\varphi.$$

Дивергенция Брэгмана задается формулой

$$V(a,b) = \varphi(a) - \varphi(b) - (\nabla\varphi(b), a-b) \quad \forall a \in \operatorname{dom}\varphi, \ b \in \operatorname{int}\operatorname{dom}\varphi.$$

Имеет место полезное 3-точечное тождество

$$V(a,c) = V(a,b) + V(b,c) + (\nabla\varphi(b) - \nabla\varphi(c), a-b).$$

Из сильной выпуклости функции φ следует оценка

$$V(a,b) \geq \tfrac{\sigma}{2}\|a-b\|^2 \quad \forall a \in \operatorname{dom}\varphi, \ b \in \operatorname{int}\operatorname{dom}\varphi.$$

Пусть $K \subseteq \operatorname{dom}\varphi$ непустое замкнутое выпуклое множество, причем $K \cap \operatorname{int}\operatorname{dom}\varphi \neq \varnothing$. Рассмотрим сильно выпуклые задачи минимизации вида

$$P_x^K(a) = \arg\min_{y \in K} \{-(a, y-x) + V(y,x)\}, \ a \in E^*, \ x \in \operatorname{int}\operatorname{dom}\varphi. \tag{2}$$

Известно, что задача (2) имеет единственное решение $z \in K \cap \operatorname{int}\operatorname{dom}\varphi$, причем

$$-(a, y-z)+(\nabla\varphi(z)-\nabla\varphi(x), y-z)\geq 0 \quad \forall y \in K .$$

Точка $P_x^K(a)$ в евклидовом случае совпадает с евклидовой метрической проекцией

$$P_K(x+a)=\arg\min_{y\in K}\left\|y-(x+a)\right\|_2 .$$

Опишем предлагаемый алгоритм для решения вариационного неравенства (1).

Алгоритм 1. Выбираем элемент $x_1 \in E$ и последовательность положительных чисел (λ_n). Полагаем $n=1$.

Шаг 1. Вычислить

$$y_n = P_{x_n}^C(-\lambda_n A x_n).$$

Шаг 2. Если $y_n = x_n$, то СТОП, иначе вычислить

$$x_{n+1} = P_{x_n}^{T_n}(-\lambda_n A y_n),$$

где

$$T_n = \left\{z \in E:\ (\nabla\varphi(x_n)-\lambda_n A x_n - \nabla\varphi(y_n), z-y_n)\leq 0\right\}.$$

Положить $n := n+1$ и перейти на шаг 1.

Замечание 1. Имеем $C \subseteq T_n$. Действительно, если предположить существование точки $w \in C \setminus T_n$, то неравенство

$$(\nabla\varphi(x_n)-\lambda_n A x_n - \nabla\varphi(y_n), w-y_n)>0$$

противоречит равенству $y_n = P_{x_n}^C(-\lambda_n A x_n)$.

Замечание 2. Если $\varphi(\cdot)=\frac{1}{2}\|\cdot\|_2^2$, то алгоритм 1 принимает вид субградиентного экстраградиентного метода [3, 4]:

$$\begin{cases} y_n = P_C(x_n-\lambda_n A x_n), \\ T_n = \left\{z \in H:(x_n-\lambda_n A x_n - y_n, z-y_n)\leq 0\right\}, \\ x_{n+1} = P_{T_n}(x_n-\lambda_n A y_n). \end{cases}$$

Имеет место.

Лемма 1. Если для некоторого $n \in \mathbb{N}$ в алгоритме 1 имеем $y_n = x_n$, то $x_n \in S$.

Далее будем предполагать, что для всех номеров $n \in \mathbb{N}$ условие $y_n = x_n$ не имеет места и перейдем к обоснованию сходимости алгоритма 1.

Лемма 2. Для последовательностей (x_n), (y_n), порожденных алгоритмом 1, имеет место неравенство

$$V(z, x_{n+1}) \le V(z, x_n) - \left(1 - \lambda_n \frac{L}{\sigma}\right) \cdot V(y_n, x_n) - \left(1 - \lambda_n \frac{L}{\sigma}\right) \cdot V(x_{n+1}, y_n),$$

где $z \in S$.

Сходимость и оценки эффективности. Сформулируем один из основных результатов работы.

Теорема 1. Пусть множество $C \subseteq E$ — выпуклое и замкнутое, оператор $A : E \to E^*$ — псевдомонотонный и липшицевый с константой $L > 0$, $S \neq \varnothing$ и $\lambda_n \in [a, b]$, где a, $b \in \left(0, \sigma L^{-1}\right)$. Тогда последовательности (x_n) и (y_n), порожденные алгоритмом 1, сходятся к некоторой точке $\overline{z} \in S$.

Рассмотрим вариационное неравенство (1) с монотонным липшицевым оператором A и выпуклым компактным множеством C. Получим для этого случая неасимптотические оценки эффективности алгоритма 1.

Функцией разрыва называют функцию вида

$$G(x) = \max_{y \in C} (Ay, x - y), \ x \in C.$$

Функция разрыва выпукла, неотрицательна и принимает нулевое значение в точке $x \in C$ тогда и только тогда, когда эта точка принадлежит множеству S. Она применяется для оценки качества приближенного решения вариационых неравенств.

Справедлива следующая теорема.

Теорема 2. Пусть $\lambda_n \in \left(0, \sigma L^{-1}\right]$. Тогда имеет место неравенство

$$G(z_N) = \max_{y \in C} (Ay, z_N - y) \le R_C(x_1) \left(\sum_{n=1}^{N} \lambda_n \right)^{-1},$$

где $R_C(x_1) = \max_{y \in C} V(y, x_1)$, $z_N = \left(\sum_{n=1}^{N} \lambda_n y_n \right) \left(\sum_{n=1}^{N} \lambda_n \right)^{-1}$.

Следствие 1. Пусть $\lambda_n = \lambda = \dfrac{\sigma}{\alpha L}$, где $\alpha \ge 1$. Тогда имеет место неравенство

$$G(z_N) = \max_{y \in C} (Ay, z_N - y) \le \alpha \cdot L \cdot R_C(x_1) \cdot \sigma^{-1} \cdot \frac{1}{N},$$

где $z_N = \dfrac{\sum_{n=1}^{N} y_n}{N}$.

Следствие 2. Пусть необходимо решить задачу (1) при помощи алгоритма 1 в условиях следствия 1 и $\varepsilon > 0$. Тогда после

$$N = \left\lceil \frac{R_C(x_1)}{\lambda\varepsilon} \right\rceil = \left\lceil \frac{\alpha L R_C(x_1)}{\sigma\varepsilon} \right\rceil$$

итераций имеет место оценка

$$G(z_N) = \max_{y \in C} (Ay, z_N - y) \le \varepsilon,$$

где $z_N = \dfrac{\sum_{n=1}^{N} y_n}{N}$ — усредненный выход работы алгоритма 1 за N итераций.

Замечание 3. В ближайшей работе планируется для алгоритма [7] изучить аналог с брэгмановскими проекциями на специально подобранные опорные к допустимому множеству полупространства. А именно, вместо итераций вида

$$\begin{cases} x_{n+1} = P_{x_n}^{C}(-\lambda A y_n), \\ y_{n+1} = P_{x_{n+1}}^{C}(-\lambda A y_n), \end{cases}$$

предлагается рассмотреть процесс

$$\begin{cases} x_{n+1} = P_{x_n}^{H_n}(-\lambda A y_n), \\ y_{n+1} = P_{x_{n+1}}^{C}(-\lambda A y_n), \end{cases}$$

где $H_n = \left\{ z \in E : \left(\nabla\varphi(x_n) - \lambda A y_{n-1} - \nabla\varphi(y_n), z - y_n \right) \le 0 \right\}$.

Список использованных источников:

1. Korpelevich G. M. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*. 1976. Vol. 12. № 4. P. 747–756.
2. Tseng P. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*. 2000. Vol. 38. P. 431–446.
3. Censor Y., Gibali A., Reich S. The subgradient extragradient method for solving variational inequalities in Hilbert space. *Journal of Optimization Theory and Applications*. 2011. Vol. 148. P. 318–335.
4. Lyashko S. I., Semenov V. V., Voitova T. A. Low-cost modification of Korpelevich's methods for monotone equilibrium problems. *Cybernetics and Systems Analysis*. 2011. Vol. 47. P. 631–639.
5. Denisov S. V., Semenov V. V., Chabak L. M. Convergence of the Modified Extragradient Method for Variational Inequalities with Non-Lipschitz Operators. *Cybernetics and Systems Analysis*. 2015. Vol. 51. P. 757–765.

6. Nemirovski A. Prox-method with rate of convergence O(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*. 2004. Vol. 15. P. 229–251.
7. Semenov V.V. A Version of the Mirror descent Method to Solve Variational Inequalities. *Cybernetics and Systems Analysis*. 2017. Vol. 53. P. 234–243.

A MODIFIED EXTRA-GRADIENT METHOD WITH BREGMAN DIVERGENCE FOR VARITIONAL INEQUALITIES

A new method of extra-gradient type for the approximate solution of variational inequalities with pseudo-monotone and Lipschitz-continuous operators acting in a finite-dimensional linear normed space is proposed. A theorem on the convergence of the method is proved and, in the case of a monotone operator, non-asymptotic estimates of the effectiveness of the method are obtained.

Key words: *variational inequality problem, monotonicity, pseudo-monotonicity, Lipschitz condition, extra-gradient method, Bregman divergence.*

Получено 12.02.2019

УДК 519.615.7
DOI: 10.32626/2308-5878.2019-19.137-141

В. Ю. Семенов*, канд. фіз.-мат. наук,
Є. В. Семенова**, канд. фіз.-мат. наук

***Інститут кібернетики імені В. М. Глушкова НАН України;
ТОВ «Дельта СПЕ», м. Київ,
****Інститут математики НАН України, м. Київ

МЕТОД РОЗВ'ЯЗАННЯ СИСТЕМ БІТОВИХ РІВНЯНЬ НА ОСНОВІ ПРИНЦИПУ ГІЛОК ТА ГРАНИЦЬ

Розв'язання систем рівнянь над бітовими полями є актуальною задачею для галузей криптографії, теорії завадостійкого кодування інформації, роботехніки, астрофізики та інших областей. У даній статті запропоновано метод розв'язування бітових рівнянь, що базується на методології гілок та границь (branch-and-bound). Запропонована методологія вже буда використана авторами для розв'язання систем нелінійних алгебраїчних рівнянь. Метод може бути використаний не тільки для систем бітових рівнянь, а також і для розв'язання систем бітових рівнянь над довільними скінченними полями Галуа $GF(n)$. Важливою особливістю запропонованого методу є те, що він дозволяє знайти усі розв'язки системи бітових рівнянь при будь-якому співвідношенні кількості змінних та кількості рівнянь. У статті наведено алгоритм, що реалізує послідовність дій, необхідну для реалізації запропонованого методу розв'язування систем бітових рівнянь. Алгоритм виконує послідовне зниження порядку системи (кількості змінних). Запропонована методика є спорідненою до методики

Constrained Propagation, що використовується у задачах розв'язання систем нелінійних алгебраїчних рівнянь та задач глобальної мінімізації функцій. Також у статті наведено чисельні приклади, що демонструють роботу метода при розв'язанні систем бітових рівнянь у випадку квадратичних нелінійностей. При цьому також досліджені різні комбінації кількості рівнянь та кількості невідомих, а також окремо розглянуто випадок розрідженої системи рівнянь. Показано, що метод має перевагу в кількості операцій перед методом прямого перебору можливих розв'язків системи рівнянь.

Ключові слова: *бітові рівняння, метод гілок та границь.*

Вступ. Розв'язання рівнянь над бітовими полями — актуальна задача, зокрема для криптографії, теорії завадостійкого кодування, роботехніки та інших областей [1]. Для розв'язання цієї задачі використовуються багато різних підходів, включаючи лінеаризацію систем рівнянь, алгоритми, що базуються на базисі Гребнера [2], методи прямого перебору та інші.

У статті запропоновано метод розвязування бітових рівнянь виду

$$f_i(x_1,...,x_n) = 0, i = 1,...,m \qquad (1)$$

над полем $GF(2)$, тобто $x_i \in \{0,1\}, 1 \le i \le n$.

Запропонований метод розвязання систем (1) базується на методології гілок та границь (branch-and-bound) [3], яка вже була застосована авторами в роботі [4] для розв'язання систем нелінійних алгебраїчних рівнянь. Особливістю методу є те, що він дозволяє знайти усі розв'язки системи бітових рівнянь при будь-якому співвідношенні кількості змінних та кількості рівнянь.

Опис алгоритму. Отже, ми розглядаємо систему рівнянь (1). На початку, пов'яжемо із змінними $x_1,...,x_n$ вектор стану $(x_1,...,x_n)$, кожна з координат якого може приймати значення «0», «1» и «2». Значення «0» та «1» є фіксованими, а значення «2» означає, що відповідна змінна може приймати одно з можливих значень: «0» чи «1».

На початку роботи алгоритму вектор розв'язків цілком складається із значень «2»: $(x_1,...,x_n) = (2,...,2)$. В процесі роботи алгоритму вектор стану, в якому є змінна, що приймає значення «2», замінюється на два вектори, в кожному з яких ця змінна приймає значення «0» та «1» відповідно. У результаті, розмірність (кількість невідомих) системи (1) знижується, щонайменше, на одиницю. Після підстановки значень «0» чи «1» в кожне з рівнянь можливі наступні варіанти.

1. Якщо ми отримали рівняння 1 = 0, то даний вектор змінних має бути відкинутим. Робота алгоритму по даній гілці зупиняється.
2. Якщо ми отримали рівняння 0 = 0, то усі значення змінних, сумісні із даним вектором, є розв'язком для даного рівняння. Виконується перехід до аналізу наступного рівняння.

3. Якщо поточне рівняння можна представити у вигляді $x_j g(x_1,...,x_n)=1$, то робиться висновок, що $x_j=1$.
4. Якщо поточне рівняння можна представити у вигляді $(1-x_j)g(x_1,...,x_n)=1$, то робиться висновок, що $x_j=0$.
5. Якщо перераховані умови не виконуються для жодного з рівнянь та у векторі стану присутнє, щонайменше, одне значення «2» (тобто $x_k=2$ для деякого індексу k), то вектор стану $(x_1,...,x_n)$ замінюється на два вектори, в яких значення k-ї координати дорівнює «0» та «1» відповідно.

Експериментальні результати. У даному дослідженні ми розглядаємо випадок, у якому функції f_i є квадратичними:

$$f_i(x_1,...,x_n)=a_i+\sum_{k=1}^{n}b_{ik}x_k+\sum_{k,l=1}^{n}c_{ikl}x_kx_l;i=1,...,m \qquad (2)$$

де $a_i,b_{ik},c_{ikl}\in\{0,1\}$. Не втрачаючи загальності розглядання, можна припустити, що $c_{ij}=0,i\geq j$.

На початку роздивимось простий приклад, що являє собою систему із одного рівняння з трьома змінними:

$$x_1+x_2+x_1x_2+x_1x_3+x_2x_3+1=0. \qquad (3)$$

Діаграма розв'язання рівняння (3) за допомогою запропонованого алгоритму зображена на рисунку. Як видно, отриманими розв'язками рівняння (3) є $\{x_1=0,x_2=1,x_3=0\}$, $\{x_1=1,x_2=0,\ x_3=0\}$, $\{x_1=1,x_2=1\}$ (для останнього розв'язку третя координата може приймати довільне значення). Також зазначимо, що кількість гілок (тобто підстановок) в дереві складає 4, що є меншим, ніж для прямого перебору розв'язків ($2^3=8$).

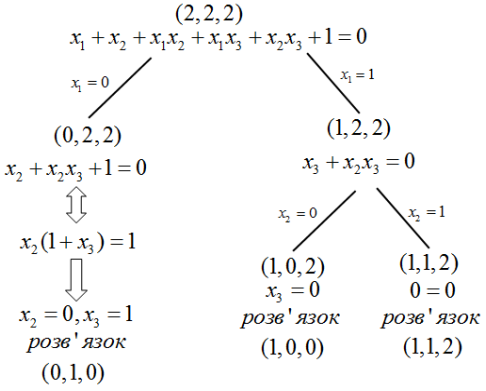

Рисунок. *Діаграма роботи методу при розв'язанні рівняння* (3)

Зазначимо, що якщо додати до системи (3) ще одне рівняння

$$\begin{cases} x_1 + x_2 + x_1x_2 + x_1x_3 + x_2x_3 + 1 = 0; \\ x_1 + x_2 + x_3 + x_1x_2 = 0, \end{cases}$$

то ми отримуємо лише один розв'язок $\{x_1 = 1, x_2 = 1, x_3 = 1\}$ за рахунок усього двох підстановок.

Як інший приклад розглянемо систему (2) з коефіцієнтами, що з рівною ймовірністю приймають значення «0» та «1». Залежність кількості розв'язків (N_s) та кількості підстановок алгоритму (N_b) від параметрів $m = 16,\ n = 16$ наведено у табл. 1.

Таблиця 1

(m, n)	N_b	N_s
(16,16)	12193	1
(12,16)	13133	15
(8,16)	12985	271
(8,16)	12985	271

Зазначимо, що кількість гілок (підстановок) у будь-якому випадку є меншою, ніж для прямого перебору розв'язків ($2^{16} = 65536$).

Тепер розглянемо випадок розрідженої матриці $C = \{c_{ikl}\}$ із формули (2). Коефіцієнт заповнення розрідженої матриці становив 2.6 %. Залежність кількості розв'язків та кількості підстановок алгоритму від параметрів (m, n) наведено у табл. 2.

Таблиця 2

(m, n)	N_b	N_s
(16,16)	2735	1
(12,16)	2185	11
(8,16)	6575	242

У цьому випадку ми також спостерігаємо, що кількість гілок (підстановок) у будь-якому випадку є меншою, ніж для прямого перебору розв'язків ($2^{16} = 65536$). Залежність кількості розв'язків (N_s) та кількості підстановок алгоритму (N_b) від параметрів $m = 16, n = 16$ наведено у табл. 2.

Висновки. У статті запропоновано метод розв'язування бітових рівнянь, що базується на методології гілок та границь (branch-and-bound), а також алгоритм, що реалізує послідовність дій, необхідну для реалізації запропонованого методу. Наведено чисельні приклади, що демонструють роботу методу при розв'язанні систем бітових рівнянь у випадку квадратичних нелінійностей. Показано, що метод

має перевагу у кількості операцій перед метод прямого перебору можливих розв'язків системи рівнянь.

Метод може бути узагальнений для розв'язання рівнянь над полями $GF(n), n > 2$.

Подяка. Семенов В. Ю. висловлює подяку професору Віденського університету А. Ноймайеру за важливі поради при виконанні цього дослідження.

Список використаних джерел:

1. Лидл Р., Нидеррайтер Г. Конечные поля: Т. 1. Пер. с англ. М.: Мир, 1988. 430 с.
2. Faugère J.-C. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). *Proc. Int. Symp. Symbolic and Algebraic Computation*. N.-Y., 2002. P. 75–83.
3. Neumaier A. Complete Search in Continuous Global Optimization and Constraint Satisfaction. *Acta Numerica*. 2004. P. 383–408.
4. Семенов В. Метод нахождения всех корней системы нелинейных алгебраических уравнений, основанный на операторе Кравчика. *Кибернетика и системный анализ.* 2015. Вып. 51, № 5. С. 169–175.

METHOD FOR THE SOLUTION OF SYSTEMS OF BIT EQUATIONS BASED ON BRANCH-AND-BOUND PRINCIPLE

Solution of systems of bit equations is an important task for the cryptography, theory of error-correction coding, information coding, robotechnics, astrophysics and other fields. In this paper we propose a method for the solution of bit equations based on the branch-and-bound methodology. The proposed methodology was already used by the authors for the solution of systems of nonlinear algebraic equations. The method can be applied not just for the systems of bit equations, but also for the equations over arbitrary finite Galois field *GF*(*n*). The important feature of proposed method is that it allows to find all solutions of system of bit equations for any combination of number of equations and number of variables. The algorithm for the implementation of the proposed method is given. The algorithm performs subsequent decreasing of the order of the system (i. e. number of variables). The proposed methodology is close to the method of Constrained Propagation which is used for the solution of the systems of nonlinear equations and global minimization tasks. Numerical examples demonstrating the application of the method to the solution of systems of quadratic bit equations is also shown. Different combinations of the number of variables and the number of equations are considered. It is shown that the method has advantage over the direct search approach for the solution of the system of bit equations.

Key words: *bit equations, branch-and-bound.*

Одержано 14.01.2019

УДК 519.9
DOI: 10.32626/2308-5878.2019-19.142-148

І. В. Сергієнко, д-р фіз.-мат. наук, професор,
В. К. Задірака, д-р фіз.-мат. наук, професор,
І. В. Швідченко, канд. фіз.-мат. наук

Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ВІД ТЕОРІЇ ПОХИБОК ДО СУЧАСНИХ КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ

Робота підводить підсумок розвитку обчислювальної математики за 50 років (1969–2019 рр.) в галузі точності та ефективності обчислювальних алгоритмів.

Зокрема, наголошено на повній похибці обчислювального алгоритму (о. а.), оцінках її якості, постановці задач оптимізації обчислень, оптимальних за точністю та швидкодією о. а., резервах оптимізації обчислень, тестуванні якості прикладного програмного забезпечення, комп'ютерних технологіях розв'язання задач прикладної та обчислювальної математики з заданими значеннями характеристик якості за точністю та швидкодією.

Ключові слова: *теорія обчислень, теорія похибок, апріорна інформація, оптимальні алгоритми, комп'ютерні технології.*

Вступ. У шістдесяті роки минулого сторіччя був бурхливий розвиток обчислювальної математики. Було створено багато методів розв'язання типових класів задач обчислювальної математики: систем лінійних та нелінійних рівнянь, відновлення функцій і функціоналів, задачі Коші для систем диференційних рівнянь, інтегральних та сингулярних рівнянь, математичної фізики, мінімізації функцій і функціоналів, прикладної статистики тощо.

Немає таких методів, які були б завжди кращими за інші. Кожен метод має диференційовану поведінку і є кращим за інші при параметрах о. а. з певної області. На той час ця важлива теза не всіма усвідомлювалась і тому кожний з авторів того чи іншого методу відстоював (часом безпідставно) його якість. Часом на наукових семінарах та захистах дисертацій доходило до бійок.

Чому це було так? Не було чітких критеріїв якості о. а. Це з одного боку. А з іншого — не було оцінок знизу похибки розв'язку задачі (які, до речі, не залежать від о. а., а лише від задачі), не враховувались інші джерела похибок, які реально супроводжують обчислювальний процес, інші характеристики о. а. (обчислювальні ресурси, необхідні для розв'язання задачі) та інше.

Така була реальна ситуація. Тому питання точності й ефективності о. а. були і є досить актуальними [1].

У статті розглядаються шляхи розв'язання цих проблем, методи побудови (при даній інформації про задачу) оптимальних за точністю або швидкодією о. а., застосування теорії похибок, загальної теорії оптимальних алгоритмів, теорії тестування якості прикладного програмного забезпечення для створення сучасних комп'ютерних технологій розв'язання задач прикладної та обчислювальної математики з заданими значеннями характеристик якості за точністю та швидкодією.

Математичні школи, які працюють у цьому напрямку зосереджені в Каліфорнійському, Московському, Варшавському університетах та в інститутах кібернетики та математики НАН України.

1. Комплексний підхід до оцінки якості наближеного розв'язку задач. В шістдесяті роки минулого сторіччя працювали різні математичні школи. Одні з них вивчали похибки метода при розв'язанні тих чи інших задач, інші досліджували неусувну похибку і пропонували методи розв'язання некоректних задач; оцінювали похибку заокруглення алгоритмів. Але кожна з цих шкіл досліджувала лише один вид похибок: метода, неусувної та заокруглення. А в реальній ситуації присутні всі три види похибок. Неврахування хоча б однієї з них на практиці не дає гарантії якості наближеного розв'язку задачі. Наприклад, неврахування похибки заокруглення може призвести до того, що комп'ютерні моделі не мають нічого спільного з фізичними моделями.

Тому на першому Симпозіумі і літній математичній школі 1969 року основна увага була приділена оцінкам повної похибки о. а. для деяких типових класів задач обчислювальної математики. Інтерес до цієї тематики був дуже великий. Про це свідчить кількість учасників Симпозіуму — 462.

Оцінка абсолютної похибки $E(I, X, Y)$ о. а. має вигляд

$$E(I, X, Y) \le \Delta_H(I, X, Y) + \Delta_M(I, X, Y) + \Delta_3(I, X, Y), \qquad (1)$$

де $\Delta_H(I, X, Y)$ — неусувна похибка о. а., $\Delta_M(I, X, Y)$ — похибка методу; $\Delta_3(I, X, Y)$ — похибка заокруглення о. а., I, X, Y — скінченні множини параметрів від яких суттєво залежать задача, о. а. та комп'ютер.

Особливо важливими є питання якості оцінок E та її складових. З позицій обчислювальної математики непокращувані оцінки не завжди нас влаштовують, оскільки вони досягаються на екзотичних задачах, які на практиці, як правило, не зустрічаються. Іноді нас більше влаштовують статистичні оцінки, які більше орієнтовані на реальні задачі [2].

2. Інші характеристики о. а. Окрім характеристики точності $E(I, X, Y)$ в практиці чисельного розв'язання задач за допомого сучасних комп'ютерів розглядають і інші характеристики о. а., які будемо ототожнювати з програмою на комп'ютері.

Нехай задачі $P(I)$ розв'язуються о. а. $A(X)$ на комп'ютері $C(Y)$. Важливе значення (для порівняльного аналізу о. а. та в загальних постановках задач оптимізації обчислень) мають наступні характеристики задач, о. а. та комп'ютерів:

- $T(I,X,Y)$ — час, необхідний для розв'язання задачі $P(I)$ о. а. $A(X)$ на комп'ютері $C(Y)$;
- $M(I,X,Y)$ — необхідна для цього пам'ять комп'ютера.

Відомо, що час T вбирає у себе наступні види робіт: введення та виведення даних T_1; обчислення розв'язку задачі T_2; обмін з зовнішніми накопичувачами T_3; додаткові обчислення T_4 (наприклад, вибір параметрів алгоритму); обчислення оцінок характеристик та інші.

Оскільки деякі з перерахованих робіт можуть виконуватись одночасно, то

$$T \le T_1 + T_2 + T_3 + T_4.$$

Оцінка пам'яті M, необхідної для розв'язання задачі, обчислюється за формулою

$$M \le \sum_{i=1}^{n} k_i N_i + \sum_{j=1}^{m} MF_j + MP,$$

де n — кількість масивів різної розмірності N_i, які використовуються в програмі розв'язання задачі, k_i — число масивів розмірності N_i; m — кількість програмних модулів, які мають бути написані користувачем для розв'язання його задачі; MF_j — оцінка пам'яті, для написаного користувачем модуля; MP — частина оцінки пам'яті (для даного транслятора) для розміщення самої програми. Наведені характеристики о. а., безумовно, не є єдино можливими.

3. Постановка задачі оптимізації обчислень. Оптимізація обчислень полягає в оптимізації однієї з введених характеристик (в загальному випадку по I,X,Y) при дотриманні обмежень на інші характеристики.

Наведемо дві основні постановки задач [2].

Мінімізація часу $T(I,X,Y)$ при дотриманні реальних (Re) обмежень на E і M:

$$T(I,X,Y) = \min_{I,X,Y};$$

$$E(I,X,Y) \le E_{\mathrm{Re}};\ \ M(I,X,Y) \le M_{\mathrm{Re}}.$$

Мінімізація повної похибки $E(I,X,Y)$ при дотриманні обмежень на T і M:

$$E(I, X, Y) = \min_{I,X,Y};$$

$$T(I, X, Y) \le T_{\text{Re}}; \;\; M(I, X, Y) \le M_{\text{Re}}.$$

Можливі й імовірнісні постановки задач оптимізації обчислень [3].

4. Оптимальні за точністю та швидкодією обчислювальні алгоритми. Ця тематика широким фронтом почала розвиватись після робіт М. С. Бахвалова [4] (роботи С. М. Нікольського у 1958 р. присвячені лише чисельному інтегруванню) та В. В. Іванова [5]. З 1971 року ця тематика опанувала наукові форуми «Питання оптимізації обчислень». Монографія [6] з'явилась лише у 1983 році.

Одним з основних критеріїв оптимальності о. а. може слугувати вимога його максимальної точності при наявній інформації про задачу і заданих обчислювальних ресурсах. При цьому вважається, що вихідна інформація задана наближено і застосовуються різні стратегії прийняття рішень (визначення оптимального о. а.): чиста (чебишовський центр області невизначеності розв'язків задачі, метод нев'язки, квазіоптимальний метод), послідовна [4, 7], послідовно-оптимальна [8].

В рамках чистої стратегії частіше всього використовується метод «капелюхів» М. С. Бахвалова та метод граничних функцій (МГФ), розроблений в Інституті кібернетики АН УРСР [7]. МГФ щільно застосовується в умовах найбільш повного використання апріорної інформації про задачу. Там, де її не вистачає, використовуються алгоритми її виявлення [9]. Це дає змогу зменшити похибку оптимального алгоритму на більш вузькому класі задач, який генерується цією додатковою апріорною інформацією [10].

Оптимальні за швидкодією о. а. потрібні в першу чергу для задач, які потребують розв'язання в режимі реального часу або для розв'язання задач трансобчислювальної складності, задач крипто та стеганоаналізу [11]. Багато оптимальних за швидкодією алгоритмів використовують швидкі ортогональні перетворення [2].

5. Тестування якості прикладного програмного забезпечення. Задача розробки якісного програмного забезпечення є найбільш важливою у загальній проблемі створення обчислювальних систем. Тому дослідження питань, пов'язаних з розробкою принципів і методів створення якісного програмного забезпечення, є досить важливим. Перш за все це стосується прикладних програм, призначених для розв'язання типових класів задач обчислювальної математики.

Експериментальні дослідження о. а. полягають у проведенні за різними критеріями числових обчислювальних експериментів (тестування) за допомогою наборів спеціально розроблених задач (тестів) для визначення функціональних можливостей о. а., кількісних показників їх характеристик і областей їх диференційованої поведінки за

цими характеристиками, порівняння програм за різними критеріями якості тощо [12]. Запропонований метод тестування базується на концепції поєднання теоретичних досліджень чисельних методів і о. а., враховуючи теорію оцінок похибок, що супроводжують обчислювальний процес, і експериментальних досліджень.

6. Резерви оптимізації обчислень. На наукових форумах розглядались наступні резерви поліпшення характеристик якості розв'язку задачі та обчислювального процесу [13].

Резерви зменшення похибок:

- за рахунок неточності вхідних даних:
  - уточнення класу задач;
  - підвищення точності вхідної інформації.
- методу:
  - використання оптимальних о. а.;
  - перехід в інший клас інформаційних операторів;
  - повне використання вхідної інформації для звуження класу задач;
- заокруглень:
  - використання схем обчислень, що мінімізують швидкість накопичення похибки заокруглень;
  - збільшення довжини розрядної сітки;
  - вибір та моделювання правила заокруглення.

Резерви зменшення процесорного часу:

- поліпшення точності оцінок похибок методу та заокруглень;
- узгодження о. а. з архітектурою комп'ютера;
- використання «швидкої» арифметики [14];
- розпаралелювання обчислень;
- спеціалізовані обчислювачі.

7. Елементи комп'ютерної технології (КТ) розв'язання задач із заданими значеннями характеристик якості. КТ як загального плану, так і для конкретних класів задач, доповідались на наукових форумах «Питання оптимізації обчислень» з 2000 року [15].

Побудова наближеного ε — розв'язку задачі $P \in \mathrm{P}$ при обмежених обчислювальних ресурсах може бути описана умовами:

$$E(I, X, Y) \le \varepsilon, \tag{2}$$

$$T(I, X, Y, \varepsilon) \le T_0(\varepsilon), \tag{3}$$

$$M(I, X, Y, \varepsilon) \le M_0(\varepsilon), \tag{4}$$

де ε, T_0, M_0 — задані числа.

Концепція КТ полягає у наступному:

- задаються ε, $T_0(\varepsilon)$, $M_0(\varepsilon)$;

- з деякої множини о. а. і програм знаходиться (або розробляється) о. а. і програма, яка може забезпечити якість (2)–(4);
- за допомогою розробленої або наявної програми обчислюється розв'язок задачі із заданими значеннями характеристик якості.

КТ щільно використовує теорію похибок і перелічені вище резерви оптимізації обчислень.

Висновки. Викладені деякі віхи розвитку теорії обчислень, які щільно обговорювались і докладались на міжнародних наукових форумах «Питання оптимізації обчислень» вподовж 1969–2018 років.

Список використаних джерел:

1. Иванов В. В. Вопросы точности и эффективности вычислительных алгоритмов. Киев : Ин-т кибернетики АН УССР, 1969. 135 с.
2. Задирака В. К. Теория вычисления преобразования Фурье. Киев : Наук. думка, 1983. 216 с.
3. Кендалл М. Дж., Стюарт А. Статистические выводы и связи. М. : Наука, 1973. 899 с.
4. Бахвалов Н. С. О свойствах оптимальных методов решения задач математической физики. *Журн. вычисл. математики и мат. физики.* 1970. Т. 10, № 3. С. 555–568.
5. Иванов В. В. Об оптимальных алгоритмах минимизации в классах дифференцируемых функций. Докл. АН СССР. 1971. Т. 201, № 3. С. 527–530.
6. Трауб Дж., Вожняковский Х. Общая теория оптимальных алгоритмов. М. : Мир, 1983. 382 с.
7. Иванов В. В., Задирака В. К. Вопросы оптимизации вычислений. К. : О-во «Знание» УССР, 1978. 34 с.
8. Сухарев А. Г. Оптимальный метод построения наилучших равномерных приближений для функций некоторого класса. *Журн. вычисл. математики и мат. физики*. 1978. Т. 18, № 9. С. 302–313.
9. Сергієнко І. В., Задірака В. К., Литвин О. М. Елементи загальної теорії оптимальних алгоритмів та суміжні питання. Киев : Наукова думка, 2012. 400 с.
10. Сергієнко І. В., Задірака В. К., Швідченко І. В. Наукова тематика міжнародних математичних форумів з питань оптимізації обчислень. *Математичне та комп'ютерне моделювання.* Серія: Фізико-математичні науки. 2017. Вип. 15. С. 189–193.
11. Задірака В. К., Кошкіна Н. В., Швідченко І. В. Комп'ютерній стеганографії 20 років. *Матеріали V міжнар. наук.-техн. конф. «Захист інформації і безпека інформаційних систем»* (2–3 червня 2016 р.). Львів, 2016. С. 98–99.
12. Бабич М. Д., Задирака В. К., Сергиенко И. В. Вычислительный эксперимент в проблеме оптимизации вычислений. II. *Кибернетика и системный анализ.* 1999. № 2. С. 59–79.
13. Бабич М. Д., Задирака В. К., Людвиченко В. А., Сергиенко И. В. Об использовании резервов оптимизации вычислений в компьютерных технологиях решения задач прикладной и вычислительной математики с требуемыми значениями характеристик качества. *Журнал вычислительной математики и математической физики.* 2010. Т. 50, № 12. С. 2285–2295.

14. Задірака В. К., Олексюк О. С. Комп'ютерна арифметика багаторозрядних чисел. Київ, 2003. 264 с.
15. Сергиенко И. В., Задирака В. К., Бабич М. Д., Березовский А. И., Бесараб П. Н., Людвиченко В. А. Компьютерные технологии решения задач прикладной и вычислительной математики с заданными значениями характеристик качества. *Кибернетика и системный анализ*. 2006. № 5. С. 33–41.

FROM ERROR THEORY TO MODERN COMPUTER TECHNOLOGIES

The article summarizes the development of computational mathematics during 50 years' period (1969–2019) in the field of accuracy and efficiency of computational algorithms.

In particular, it is emphasized on the global error of the computational algorithm (c. a.), estimates of its quality, the formulation of computations optimization problems, computational algorithms that are optimal in accuracy and processing speed, reserves of calculations optimization, testing the quality of applied software, computer technologies solving the problems of applied and computational mathematics with the given values of quality characteristics in accuracy and processing speed.

Key words: *theory of computing, error theory, apriori information, optimal algorithms, computer technologies.*

Одержано 31.01.2019

УДК 512.7+512.9,688.321
DOI: 10.32626/2308-5878.2019-19.148-155

Р. В. Скуратовский, преподаватель

Межрегиональная академия управления персоналом, г. Киев

РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ К УДВОЕНИЮ ТОЧКИ СКРУЧЕННОЙ КРИВОЙ ЭДВАРДСА НАД КОНЕЧНЫМ ПОЛЕМ

Получено решение задачи обратной к удвоению точки для кривой представленной в скрученной форме Эдвардса. Получены оценки сложности операции деления на два в сравнении с удвоением точки. Найдено одно из приложений свойств делимости точки на два для определения порядка точки в криптосистеме основаной на проблеме дискретного логарифма.

Найдены необходимые и достаточные условия делимости точки $G=(X,Y)$ кривой $E_{a,d}$ на 2. Исследовано возможность применения данных кривых для генерации криптостойкой последовательности большого периода. Важность операции делимости точки на 2 при криптоанализе уже частично замечена криптографами.

Исследованы все возможные количества результатов от деления точки на два и зависимости этих количеств от делимой точки. Исследованы необходимые и достаточные условия существования 4 разных прообразов точки $G=(X,Y)$ при делении ее на два. Спаривание на дружественных эллиптических кривых простого порядка или почти простого порядка есть очень существенным во многих криптографических протоколах вида короткой цифровой подписи длительного использования.

Ключевые слова: *конечное поле, эллиптическая кривая, кривая Эдвардса, порядок кривой, порядок точки эллиптической кривой, символ Лежандра, квадратичный вычет, квадратичный невычет.*

Введение. Мы рассматриваем алгебраические кривые в форме Эдвардса [1] над простым полем F_p, которые сейчас являются одними из наиболее перспективных носителей групп, используемых в асимметричных криптосистемах.

Цель роботы — получение новых [2, 3] и уточнение старых критериев делимости точки кривой не только напополам, но и на 4 над полем F_{p^n}. Важность операции делимости точки на 2 при криптоанализе уже частично описана в работе А. В. Бессалова [4]. Наша цель найти эти условия и исследовать возможности их применения для скрученной кривой Эдвардса. Пусть E — единичный элемент в группе точек кривой $E_{a,d}$.

Основной результат. Скрученная кривая Эдвардса $E_{a,d}$ имеет вид

$$ax^2+y^2=1+dx^2y^2,\ a,d\in F_p^{\ *},\ ad(a-d)\neq 0,\ d\neq 1,\ p\neq 2,\ a\neq d. \quad (1)$$

Под делимостью точки (X; Y) пополам понимается нахождение ее прообраза, то есть точки (x; y), которая при применении формулы удвоения точки [1] дает в результате точку (X; Y).

Теорема 1. Пусть G = (X; Y) — точка скрученной кривой Эдвардса. Тогда необходимым условием делимости точки G на 2 является условие

$$\left(\frac{1-aX^2}{p}\right)\neq -1.$$

Доказательство. Для скрученной кривой Эдвардса закон удвоения [4, 9] имеет форму

$$2(x_1,y_1)=\left(\frac{2x_1y_1}{y_1^2+ax_1^{\ 2}},\frac{y_1^2-ax_1^{\ 2}}{1-dx_1^{\ 2}y_1^2}\right)=(X,Y), \quad (2)$$

отсюда, воспользовавшись уравнением кривой, мы выводим модифицированную формулу сложения точки с собой:

$$2(x_1, y_1) = \left(\frac{2x_1 y_1}{1 + dx_1^{\,2} y_1^{\,2}}, \frac{y_1^2 - ax_1^{\,2}}{1 - dx_1^{\,2} y_1^{\,2}} \right) = (X, Y) = G \ . \qquad (3)$$

Рассмотрим уравнение $\frac{2x_1 y_1}{1 + dx_1^{\,2} y_1^{\,2}} = X$ равносильное $dXx^2 y^2 - 2xy + X = 0$, при $1 + dx_1^2 y_1^2 \neq 0$, и применим замену $t = x_1 y_1$, после чего получаем уравнение $dXt^2 - 2t + X = 0$, решение, которого существует тогда и только тогда, когда $\left(\frac{1 - dX^2}{p} \right) = 1$ (или если $1 - dX^2 \equiv 0 (\mathrm{mod}\ p)$). Решения имеют вид $t_{1,2} = \frac{1 \pm \sqrt{1 - dX^2}}{dX}$, они существуют если $\left(\frac{1 - dx_1^2}{p} \right) = 1$. Согласно с леммой 1 имеем $\left(\frac{1 - dx_1^2}{p} \right) = \left(\frac{1 - ax_1^2}{p} \right)$. Из уравнения (2) имеем для первой координаты одно уравнение

$$\frac{2x_1 y_1}{y_1^2 + ax_1^{\,2}} = X \ .$$

Сделав замену $u = \frac{y}{x}$, корректность которой следует из не делимости точек $D_{0,1} = (0, \pm 1)$ второго порядка на 2 а также 2 особых точек [3] ввиду того, что это точки для которых задача деления на 2 не имеет смысла а других точек вида $(0, y)$ не существует, получаем $\frac{2u}{u^2 + a} = X$ или $2u = X(u^2 + a)$. Переписав последнее уравнение как квадратное относительно u получаем $Xu^2 - 2u + Xa = 0$, где определитель $D_2 = 4(1 - aX^2)$. Поэтому, согласно лемме 1, имеем уравнения $dXt^2 - 2t + X = 0$, и $Xu^2 - 2u + Xa = 0$ решения которых существуют или не существуют одновременно. Это дает выражения для координат точки $P_j = (x_j, y_j)$: $x_j = \sqrt{t_j u_j^{-1}}, y_j = \sqrt{t_j u_j}$ $j \in \{0,1\}$.

Приравнивая левые части равенств $\frac{2x_1y_1}{1+dx_1^2y_1^2}=X$ и $\frac{2x_1y_1}{y_1^2+ax_1^2}=X$, получаем $ax_1^2+y_1^2=1+dx_1^2y_1^2$, то есть полученные пары координат (x_1,y_1) удовлетворяют уравнению кривой. Заметим, что вместе с (x_1,y_1) выше указанные уравнения удовлетворяют точки $(-x_1,-y_1),\left(-\frac{y_1}{\sqrt{a_1}},-x_1\right),\left(\frac{y_1}{\sqrt{a_1}},x_1\right)$.

Проанализируем, какие из полученных точек удовлетворяют уравнение удвоения точки по 2-ой координате

$$\frac{y_1^2-ax_1^2}{1-dx_1^2y_1^2}=Y\ .$$

Преобразуем уравнение кривой (1) как $Y^2=\frac{1-aX^2}{1-dX^2}$, подставим полученные $X=\frac{2x_1y_1}{1+dx_1^2y_1^2}$ и обозначим $x=x_1,\ y=y_1$, тогда

$$Y^2=\frac{1-aX^2}{1-dX^2}=\frac{1-a\dfrac{4t^2}{(y^2+ax^2)^2}}{1-d\dfrac{4t^2}{y^2+ax^2}}=\frac{(y^2+ax^2)^2-4at^2}{(y^2+ax^2)^2-4dt^2}=\frac{(y^2+ax^2)^2-4at^2}{(1+dt^2)^2-4dt^2}=$$

$$=\frac{(y^2-ax^2)^2}{(1-dt^2)^2}=\frac{(y^2-ax^2)^2}{(1-dx^2y^2)^2}\ .$$

Поэтому получили уравнение, которое задает вторую координату полученную в результате удвоения точки (x_1,y_1). Это уравнение мы используем для выбора правильного из дополнительных корней $(-x_1,-y_1),\left(-\frac{y_1}{\sqrt{a_1}},-x_1\right),\left(\frac{y_1}{\sqrt{a_1}},x_1\right)$ к истинному корню (x_1,y_1). Таким образом, второе уравнение удовлетворяют точки (x_1,y_1) и $(-x_1,-y_1)$. Заметим, что $(-x_1,-y_1)=(x_1,y_1)+D$. Учитывая, что $y_1^2-dx_1^2y_1^2=1-ax_1^2$ откуда $y_1^2(1-dx_1^2)=1-ax_1^2$ получаем

$$\left(\frac{1-ax_1^2}{p}\right)=\left(\frac{1-dx_1^2}{p}\right).$$

Из равенства (2) для второй координаты имеем

$$\frac{y_1^2 - a{x_1}^2}{1 - d{x_1}^2 y_1^2} = Y .$$

Поскольку мы ввели замену переменных $t = x_1 y_1$, то последнее уравнение примет вид $y_1^2 - a{x_1}^2 = Y(1 - dt^2)$. Откуда, получаем

$$\frac{t^2}{x^2} - a{x_1}^2 = Y(1 - dt^2),$$

$$t^2 - ax^4 = Y(1 - dt^2)x^2,$$

$$ax^4 + Y(1 - dt^2)x^2 - t^2 = 0 .$$

Откуда

$$x^2 = \frac{Y(dt^2 - 1) \pm \sqrt{Y^2(1 - dt)^2 + 4dt^2}}{2d} . \qquad (4)$$

После подстановки $t_{1,2} = \frac{1 \pm \sqrt{1 - dX^2}}{dX}$ имеем

$$x_{1,2}^2 = \frac{Y(dt_{1,2}^2 - 1) \pm \sqrt{Y^2(1 - dt_{1,2}^2)^2 + 4dt_{1,2}^2}}{2d}. \qquad (5)$$

Заметим, что знаки $\pm$ перед выражениями $\sqrt{1 - dX^2}$ одинаковы. Поскольку эти корни являются сопряженными иррациональностями, то точка $(\pm x, \pm y)$ удовлетворяют одновременно уравнению кривой, чего достаточно для выполнения условий теоремы. Кроме того, $y^2 = \frac{t^2}{x^2} = \frac{(1 + \sqrt{1 - dx^2})^2}{dx^3}$, то есть элемент dx, где x определяется условием (4), должен быть квадратическим вычетом в $\mathbb{F}_p$. Заметим, что оба корни уравнений (4) и (5) являются сопряженными иррациональностями, поэтому если один из них удовлетворяет уравнению над Z или над $\mathbb{F}_p$, то элементы полученные в результате операций сложения, умножения и возведения его в натуральную степень тоже все ему удовлетворяют. Поэтому все найденные координаты удовлетворяют уравнению кривой (1) и уравнению операции удвоения. Операция деление точки требует 4 умножения и 2 извлечения корня в поле и 1 инверсии.

Обозначим $Y^2(1 - \mathrm{d}\frac{1 \pm \sqrt{1 - dX^2}}{dX})^2 + 4d(\frac{1 \pm \sqrt{1 - dX^2}}{dX})^2$ как g. При этом знаки «+» или «–» подставляются в обеих дробях одинако-

вым образом. А полученные выражения обозначаем как g_1 для «+» и как g_2 для знака «–».

Теорема 2. Для любой точки *A*, допускающей деление надвое, существует столько же точек со свойством $2B = A$, сколько существует на кривой точек D, для которых $2D = E$.

Доказательство. Пусть D_i, $i \in 2,4$ семейство точек удовлетворяющих условию $2D = E$. Тогда каждая из них удовлетворяет и уравнению $2(B + D_i) = A$, которое по сути есть уравнение (2), где точка $B + D_i = (x_1, y_1)$ и удовлетворяет условию $2(x_1, y_1) = \left(\frac{2x_1 y_1}{y_1^2 + a x_1^2}, \frac{y_1^2 - a x_1^2}{1 - d x_1^2 y_1^2} \right) = (X, Y)$. Действительно $2(x_1, y_1) = 2(B + D_i) = 2B + 2D_i = A + E = A.$ Поэтому совокупное количество решений уравнения (2) равно количеству решений $2D = E$.

Замечание. Необходимыми и достаточными условиями делимости точки $G = (X, Y)$ кривой $E_{a,d}$ на 2 является условия

$$\left(\frac{1 - aX^2}{p} \right) \neq -1 \text{ и } (x_1, y_1) \in E_{a,d}.$$

Доказательство. Поскольку полученное в результате деление на 2 точку (x_1, y_1) не обязательно лежит на исходной кривой. Хотя она и удовлетворяет уравнениям удвоения точки (2) и (3), покуда мы можем получить такую пару (*X*; *Y*), которая не на кривой $E_{a,d}$ в силу того, что мы не потребовали выполнения $(X, Y) \in E_{a,d}$. Ввиду того, что группа точек кривой $E_{a,d}(F_{p^n})$ не является ограничено делимой [8] т. е. не для любого $n \in N$, $n < m$ и $g \in G$ уравнение $x^n = h$ имеет решение $h \in G$, то это не выполняется автоматически. Именно поэтому сформулированное в теореме 3 условие является лишь необходимым. Дополнив его условием $(x_1, y_1) \in E_{a,d}$ мы получим следующий критерий. Деление точки требует 4M, 2S и I в F_{p^n}.

Следствие 1. Необходимым и достаточным условием существования 4 разных точек, для которых результат удвоения равен *G,* является: $(\frac{1 - dX^2}{p}) = 1$ и $(\frac{g}{p}) = 1$.

Следствие 2. Если $(\frac{1 - dX^2}{p}) = 1$ но $(\frac{g_1}{p}) = 0$ и $(\frac{g_2}{p}) = 0$, то для точки *A* = (*X*, *Y*) существует либо 2, либо 4 прообраза в зависимости

от количества точек D, со свойством $2D = E$. Последнее определяется условием $\left(\frac{ad}{p}\right) = 1$ [2, 3, 7].

Следствие 3. Если $\left(\frac{1 - aX^2}{p}\right) \neq -1$ и $(x, y) \in E_{a,d}$, то существует 2 прообраза при $\left(\frac{ad}{p}\right) = -1$, либо 4 при $\left(\frac{ad}{p}\right) = 1$.

Доказательство основывается на теореме 2 и условии существования особых точек 2-го порядка записанного как $\left(\frac{ad}{p}\right) = 1$ [2, 3, 7].

Вывод. В работе исследована обратная операция к удвоению точки для скрученной кривой Эдвардса над конечным полем.

Список использованных источников:

1. Bernstein Daniel J., Birkner Peter, Joye Marc, Lange Tanja, Peters Christiane. Twisted Edwards Curves. IST *Programme under Contract IST-2002-507932 ECRYPT*. 2008. P. 1–17.
2. Скуратовський Р. В. Побудова еліптичних кривих з нульовим слідом ендоморфізма Фробеніуса. *Захист інформації*. 2018. Т. 20, № 1, січень-березень 2018. С. 32–45.
3. Скуратовський Р. В. Суперсингулярність еліптичних кривих і кривих Едвардса над F_p^n. Research in mathematics and mechanics. 2018. Т. 31, № 1. С. 17–26.
4. Бессалов А. В., Третьяков Д. Б. Удвоение точки и обратная задача для кривой Эдвардса над простым полем. *Сучасний захист інформаці*ї. 2013. № 3. С. 16–27.
5. Bernstein D. J., Lange Tanja. Faster addition and doubling on elliptic curves. *IST Contract 2002-507932 ECRYPT*. 2007. P. 1–20.
6. Skuratovskii R. V. Employment of Minimal Generating Sets and Structure of Sylow 2-Subgroups of Alternating Groups in Ciphers. *Advances in Computer and Computational Sciences*. P. 351–364.
7. Бессалов А. В. Эллиптические кривые в форме Эдвардса и криптография : монография. Киев : Polytechnika, 2017. 272 с.

A SOLUTION OF THE INVERSE PROBLEM TO DOUBLING OF TWISTED EDWARDS CURVE POINT OVER FINITE FIELD

A solution for the inverse doubling problem is obtained for elliptic curves represented in the twisted Edwards form. Estimates of the complexity of the division operation into two are obtained in comparison with the doubling of the point. One of the applications of the divisibility properties of a curve point into two is considered to determine the order of a point in a cryptosystem based on discrete logarithm problem.

The necessary and sufficient conditions for the divisibility of a point $G=(X,Y)$ of a curve $E_{a,d}$ by 2 are found. The possibility of using these curves to generate a crypto-resistant sequence of a large period is investigated.

All possible numbers of the result of the division of a point into two and the dependence of these quantities on the dividend point are studied. The necessary and sufficient conditions for the existence of 4 different preimages of a point $G=(X,Y)$ when dividing it into two are investigated. Pairing-friendly curves of prime or near-prime order are absolutely essential in certain pairing-based schemes like short signatures with longer useful life.

Key words: *elliptic curve, twisted Edwards curve, curve order, points order, Legendre symbol, square, non-square.*

Получено 21.01.2019

УДК 519.65
DOI: 10.32626/2308-5878.2019-19.155-160

В. М. Старков, д-р фіз.-мат. наук,
П. М. Томчук, д-р фіз.-мат. наук
Інститут фізики НАН України, м. Київ

ПРО ВПЛИВ ПОХИБОК ВИМІРЮВАНЬ НА ІНТЕРПРЕТАЦІЮ РЕЗУЛЬТАТІВ ЛАЗЕРНИХ ЕКСПЕРИМЕНТІВ

На основі врахування і математичного аналізу незначних апаратних похибок вимірювань розглянуті приклади їх впливу на фізичну інтерпретацію лазерних експериментальних досліджень. Проведений нами аналіз показує, що ігнорування факту наявності похибок може призвести до помилкових висновків щодо фізичної суті розглянутих оптичних явищ.

Ключові слова: *похибки вимірювань, експеримент, інтерпретація, апроксимація.*

Вступ. На принципове значення достовірності інтерпретації результатів фізичних досліджень звертали увагу видатні вчені [1, 2]. Так, в роботі [2, с. 3] сказано: «Будь-яке наукове дослідження в галузі фізики (і не тільки в галузі фізики) безсумнівно, пов'язане з інтерпретацією отриманих результатів. Таку інтерпретацію часто називають «з'ясуванням фізичного сенсу» або досяганням «розуміння» тих явищ, які досліджують. Зазвичай, інтерпретація фізичного явища відображає рівень розвитку науки в даний момент часу, і тому вона не є абсолютною, а може змінюватися з плином часу». До останнього зауваження можна лише додати, що інтерпретація відображає, крім усього іншого, рівень інтелекту, освіти, наукового досвіду і т.д. дослідника, який її реалізує. Інтерпретація результатів наукового фізич-

ного експерименту неминуче обумовлена тим, що дослідник повинен мати чітке уявлення про взаємодію всіх складових елементів експериментального процесу. Дуже важливою якісною обставиною в цій роботі є обов'язкове врахування наявності в результатах експерименту похибок вимірювань, оскільки експеримент завжди проводять на реальних установках. Якраз на цій множині даних з сукупністю різного роду похибок виникають нерідко серйозні проблеми.

Математична інтерпретація даних оптичного експерименту. Умовно наш приклад пов'язаний з експериментальними дослідженнями ефектів оптичного обмеження в тонких наноструктурних плівках різних політипів карбіду кремнію. Це середовище є перспективним для використання в екстремальних умовах високих і низьких температур, при значних радіаційних навантаженнях і в хімічно активній атмосфері [3, с. 91]. Результати досліджень показали, зокрема, що в зразку карбіду кремнію, який характеризується в основному аморфною фазою, ефект оптичного обмеження, як на основній довжині хвилі генерації неодимового лазера (λ = 1064 нм), так і на його другій гармоніці (λ = 532 нм), не був виявлений. Аналогічний результат був отриманий і для зразка, який завдяки додатковій обробці відпалом складався майже на 100 % з кристалічної фази (3C) нанорозмірного карбіду кремнію. З цих результатів випливає, що залежність інтенсивності випромінювання, що пройшло через зразок, від інтенсивності падаючого випромінювання носить майже лінійний характер для аморфного і 100 % кристалічного зразка.

Оскільки метою подальшого викладу є виявлення інтерпретації без строгого врахування похибки реєстрованих даних, то будемо використовувати інші лінійні залежності подібного роду, але з більшим на порядок числом експериментальних точок і з більш яскраво вираженими похибками вимірювань [4, с. 485–486].

На рис. 1, а показано типову залежність (експеримент 1) відносної величини зареєстрованого сигналу (наприклад, проходження лазерного пучка через обмежуючу діафрагму за умови відсутності зразка) від вхідного сигналу $u_\delta(x) = I_\delta^{(out)} / I_{\max}$, $x = I^{(in)} / I_{\max}$. Аналогічну залежність (експеримент 2) відносної величини сигналу повного пропускання зразка від інтенсивності лазерного випромінювання приведено на рис. 1, b: $u_{\delta s}(x) = \tilde{I}_\delta^{(out)} / I_{\max}$. З достатнім ступенем впевненості можна стверджувати, що обидві залежності носять лінійний характер. Дійсно, апроксимуючи експериментальні дані лінійними функціями, отримаємо:

$$\hat{u}_\delta(x) = a_1 + b_1 x;\ a_1 = 0.00485;\ b_1 = 1.227; \quad (1)$$

$$\hat{u}_{\delta s}(x) = a_2 + b_2 x;\ a_2 = -\,0.00175;\ b_2 = 1.078. \quad (2)$$

Максимальні похибки наближення в першому і другому випадку виявляються рівними відповідно $\delta\hat{u} = 0.015$, $\delta\hat{u}_s = 0.012$.

Результати апроксимації (1), (2) показані графіками на рис. 2.

Якщо наблизити результати експерименту без зразка (рис. 1) поліномом третього степеню:

$$\bar{u}_\delta(x) = a_3 + b_3 x + c_3 x^2 + d_3 x^3;$$

$$a_3 = -0.0148;\ b_3 = 1.533;\ c_3 = -1.167;\ d_3 = 1.266, \qquad (3)$$

то отримаємо максимальну похибку наближення, що дорівнює $\delta\bar{u} = 0.0145$, тобто менша, ніж в разі лінійної апроксимації.

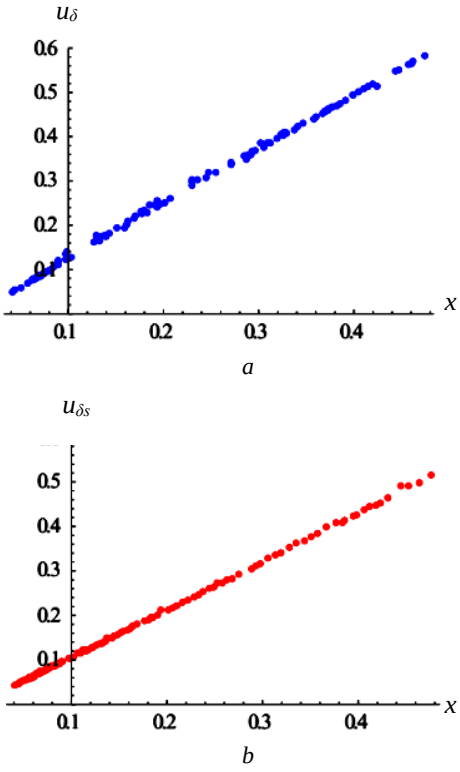

Рис. 1. *Залежність інтенсивності лазерного випромінювання, що пройшло через обмежуючу діафрагму за умови відсутності зразка* (*a*), *та інтенсивності лазерного випромінювання, що пройшло через зразок* (*b*), *від інтенсивності падаючого випромінювання*

Слід зазначити принципове значення того факту, що в результаті всіх наближень (1)–(3) коефіцієнти a_i $(i = 1, 2, 3)$ виявилися ненульо-

вими. Інакше кажучи, при відсутності сигналу на вході в систему вимірювальна апаратура фіксує в першому і другому варіанті експерименту присутність сигналу. Цей сигнал є ніщо інше, як похибка моделювання на початку координат

$$\delta u(0) = u_\delta(0) - u(0) \neq 0 \,. \qquad (4)$$

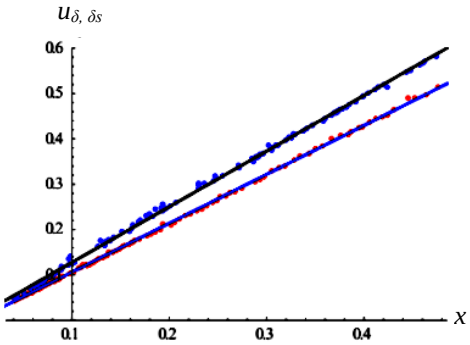

Рис. 2. *Апроксимації експериментальних даних лінійними функціями*

Ігнорування цього факту може призвести до невірної фізичної інтерпретації експериментальних досліджень. Справа в тому, що зіставлення результатів вимірювань у другому експерименті з даними в першому здійснюють нерідко шляхом діленням одних даних на другі: $\left(\hat{u}_{\delta s}(x) / \hat{u}_\delta(x)\right)$. Такий варіант можливий, наприклад, при наявності ідеальної вимірювальної апаратури, яка виключає будь-які похибки. За ідеальних умов усі експериментальні точки лежать строго на прямій, яка, в свою чергу, проходить через початок координат.

Пояснимо сказане простими викладками. Нехай $a_1 = 0$ и $a_2 = 0$, тоді

$$\hat{u}_{\delta s}(x) / \hat{u}_\delta(x) = \hat{b}_2 x / \hat{b}_1 x = const \ \ (\hat{b}_1 \neq 0) \,.$$

Підкреслимо: або це ідеальний варіант, або в процесі попередньої обробки експериментальні дані наближаються лінійними функціями $u_{\delta i}(x) = b_i x$. Зрозуміло, що мова йде про експерименти, результати яких подібні представленим на рис. 1 та 2.

Розглянемо перший, досить простий, але реальний випадок, коли дані вимірювань супроводжують похибки, але пряма лінія перших вимірів проходить строго через нуль ($a_1 = 0$), тобто при відсутності сигналу на вході в систему відсутній і спостережуваний сигнал. Нехай всі інші коефіцієнти мають попередні значення з експериментів 1 і 2. Тоді

$$v_1(x) = \hat{u}_{\delta s}(x) / \tilde{u}_\delta(x) = (a_2 + b_2 x) / (b_1 x) =$$

$$= b_2 / b_1 + a_2 / (b_1 x) = \alpha_1 + \beta_1 x^{-1}, \ \ \alpha_1 = 0.879, \ \beta_1 = -0.00143,$$

так, що отримана звичайна гіпербола.

Якщо розглядати варіант, коли всі коефіцієнти з (1) і (2) відмінні від нуля, то і в цьому випадку буде отримана гіперболічна залежність. Можна тепер уявити, що деякому «абстрактному експериментатору», який з різних причин залишає без уваги (4), більш цікавий варіант, коли результати першого експерименту апроксимовані поліномом третього степеня (3). Тим більше, функції $v_1(x)$ и $v_2(x)$ поводять себе в околі нуля своєрідно. В результаті виходить така функція $v_3(x)$, графічне зображення якої (рис. 3, а) може викликати спокусу пошуку якогось «глибокого» фізичного сенсу.

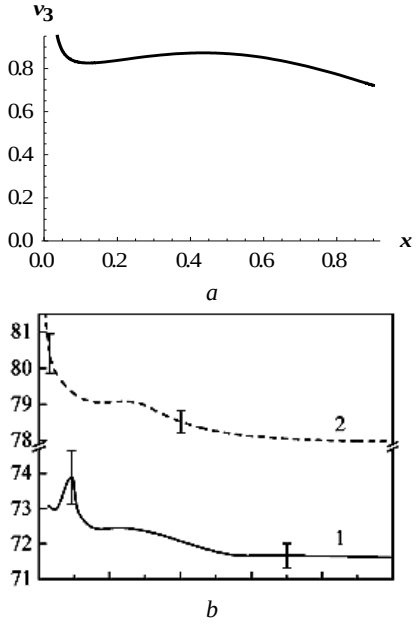

Рис. 3. *Графічне представлення функції* $v_3(x)$ (*a*) *та її аналога* (*b, крива 2 [5, с. 98]*)

Справа в тому, що якщо розкласти $\bar{u}_\delta(x)$ (17) на прості множники

$$\bar{u}_\delta(x) = (x - 0.00972)(1.266\,x^2 - 1.155\,x + 1.522),$$

то можна побачити, що значення $x^* = 0.00972$ є особливою точкою функції

$$v_3(x) = \hat{u}_{\delta s}(x) / \bar{u}_\delta(x) = (1.078x - 0.00175) \times$$
$$\times[(x - 0.00972)(1.266\,x^2 - 1.155\,x + 1.522)]^{-1}$$

і її похідних. Наявність полюса визначає характер поведінки розглянутої функції $v_3(x)$. Щоб уникнути невірної інтерпретації результа-

тів експерименту можна рекомендувати замість поділу функцій використовувати відносини їх похідних $\left(\hat{u}'_{\delta s}(x)/\hat{u}'_{\delta}(x) = b_2/b_1\right)$.

Як приклад такого роду «глибокого» фізичного сенсу можна вказати на результати роботи [5, с. 98] (рис. 3), де крива 2 демонструє наявність фізичного ефекту НЛО. Так що основний результат роботи був сформульований як видатне досягнення дослідника [6, с. 35]: «Вперше було спостережено ефект гігантського нелінійно-оптичного (НЛО) відгуку в пористих шарах нанокристалів TiO2 анатаз модифікації, що на шість порядків перевищує відгук об'ємного матеріалу».

Висновки. На конкретних прикладах результатів наукових експериментальних досліджень ми показали, що апроксимація експериментальних залежностей оптичного експерименту поліномами (як це часто роблять) при неуважному ставленні до наявності похибок у даних експерименту (особливо в околі початку координат) може призвести до невірної фізичної інтерпретації отриманих результатів.

Список використаних джерел:

1. Гейзенберг В. К. Что такое «понимание» в теоретической физике. *Природа.* 1971. № 4. С. 75–77.
2. Давыдов А. С. Интерпретация результатов научных исследований в области физики. *Препринт ИТФ. 80. I3IP.* 1980. 28 с.
3. Borshch A. A., Brodyn M. S., Starkov V. N. et al. Broadband optical limiting in thin nanostructured silicon carbide films and its nature. *Optics Communications.* 2016. № 364. P. 88–92.
4. Starkov V. N., Borshch A. A., Gandzha I. S., Tomchuk P. M. Some Examples of Seemingly Plausible Interpretation of Experimental Results. *Ukr. J. Phys.* 2017. Vol. 62. № 6. P. 481–488.
5. Gayvoronsky V., Galas A., Shepelyavyy E. et al. Giant nonlinear optical response of nanoporous anatase layers. *Appl. Phys. B* 80. 2005. P. 97–100.
6. Гайворонский В. Я. Дослідження нелінійно-оптичних властивостей композитів на основі пористих напівпровідників та наноструктурованих діелектриків. Автореф. дис. … д-ра фіз.-мат. наук. Інститут фізики НАН України. Київ, 2015. 38 с.

ON THE EFFECT OF MEASUREMENT ERRORS ON THE INTERPRETATION OF LASER EXPERIMENTAL RESULTS

On the basis of accounting and mathematical analysis of minor instrumental measurement errors, examples of their influence on the physical interpretation of laser experimental studies are considered. Our analysis shows that ignoring the fact of errors may lead to erroneous conclusions regarding the physical essence of the considered optical phenomena.

Key words: *measurement errors, experiment, interpretation, approximation.*

Одержано 11.02.2019

УДК 519.85
DOI: 10.32626/2308-5878.2019-19.161-167

П. І. Стецюк, д-р фіз.-мат. наук,
О. М. Хом'як, канд. фіз.-мат. наук
Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ПРО УСЕРЕДНЕННЯ ЧИСЕЛ ТА ЛІНІЙНИХ СПЛАЙНІВ

Розглядаються задачі безумовної мінімізації опуклих функцій для знаходження мінімальних за L_p-нормою лінійних сплайнів для випадків $p \geq 1$ та $1 \leq p \leq 2$. Вони побудовані по аналогії з подібними задачами для знаходження числа, яке за L_p-нормою мінімально відрізняється від m заданих чисел $a_1, \ldots, a_m$. Якщо $p \geq 1$, то використовується негладка функція, а якщо $1 < p \leq 2$ — гладка функція. Показано, що при певному виборі параметра p оптимізаційні задачі породжують відомі методи — метод найменших квадратів, метод найменших модулів та мінімаксний чебишевський метод. Наведено властивості розв'язків задач при $1 < p \leq 2$.

Ключові слова: *L_p-норма, опукла функція, негладка функція, метод найменших модулів, метод найменших квадратів.*

Вступ. Обробка експериментальних даних для отримання достовірних результатів на основі проведених вимірювань — задача, з якою сучасний вчений та інженер зустрічаються майже повсякденно. Серед методів обробки експериментальних даних слід відмітити метод найменших модулів [1], використання якого доцільно в тих випадках, коли розподіл помилок вимірювань підпорядкований закону Лапласа, та метод найменших квадратів [2], який використовується, коли розподіл помилок вимірювань підпорядкований закону Гаусса. Меншого поширення набули методи, в яких мінімізується L_p-норма вектора $y = (y_1, \ldots, y_m)^T$, яка визначена як: $\|y\|_p = (\sum_{i=1}^{m} |y_i|^p)^{1/p}$, де $p \geq 1$ — скалярний параметр.

У статті розглянемо оптимізаційні задачі для знаходження числа, яке за L_p-нормою мінімально відрізняється від m заданих чисел $a_1, \ldots, a_m$, та використаємо цю техніку для знаходження мінімального за L_p-нормою лінійного сплайна. Покажемо, що при певному виборі

параметра p оптимізаційні задачі породжують відомі методи обробки експериментальних даних, та дослідимо властивості цих методів для p — такого, що $1 \le p \le 2$.

1. Про L_p -усереднення чисел. Нехай задано m чисел $a_1, \dots, a_m$. Потрібно знайти таке число a_p^*, яке за L_p -нормою мінімально відрізняється від чисел $a_1, \dots, a_m$. Умовимось його називати L_p -усередненим числом, або для зручності усередненим числом. Знаходженню числа a_p^* відповідає задача безумовної мінімізації опуклої негладкої функції: знайти

$$a_p^* = \underset{a \in R}{\operatorname{argmin}} \left\{ f_p(a) = \left(\sum_{i=1}^{m} |a - a_i|^p \right)^{1/p} \right\}, \qquad (1)$$

де $p \in R$ — скалярний параметр такий, що $p \ge 1$. Умова $p \ge 1$ забезпечує опуклість функції $f_p(a)$.

Задача (1) завжди має розв'язок, але не обов'язково він є єдиним. Так, наприклад, якщо $p = 1$ та $m = 2$, то задача (1) є задачею мінімізації функції $f_1(a) = |a - a_1| + |a - a_2|$. Її оптимальним розв'язком є $a_1^* = \lambda a_1 + +(1-\lambda)a_2$, де $0 \le \lambda \le 1$, a_1 та a_2 — довільні числа. При цьому мінімальне значення функції буде рівним $f_1(a_1^*) = |a_2 - a_1|$. Цей випадок для $a_1 = 1$ та $a_2 = 2$ показано на рис. 1, звідки легко бачити, що мінімальне значення функції $f_1(a) = |a-1| + |a-2|$ досягається у кожній із точок інтервалу [1, 2]. Їм відповідає мінімальне значення функції $f_1^* = 1$.

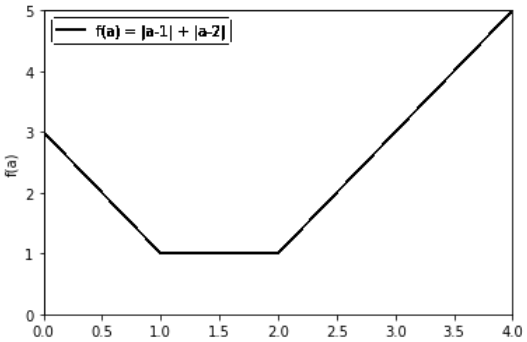

Рис. 1. *Графік функції* $f_1(a) = |a-1| + |a-2|$

Якщо $p=1$, то аналогічна ситуація має місце і для чотирьох чисел $a_1=1$, $a_2=2$, $a_3=3$ та $a_4=6$ (див. рис. 2). Тут задача (1) є задачею мінімізації функції $f_1(a)=|a-1|+|a-2|+|a-3|+|a-6|$, оптимальним розв'язком якої є $a_1^*=2\lambda+3(1-\lambda)=3-\lambda$, де $0\le\lambda\le 1$. При цьому мінімальне значення функції $f_1(a_1^*)=6$.

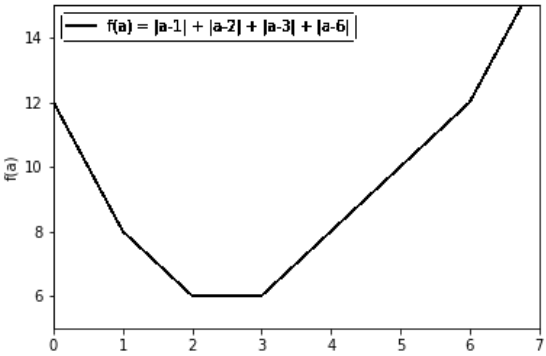

Рис. 2. *Графік функції* $f_1(a)=|a-1|+|a-2|+|a-3|+|a-6|$

Залежно від значень параметра p із задачі (1) витікають три її часткові випадки, що тісно пов'язані з відомими методами обробки експериментальних даних. Якщо $p=1$, то отримуємо функцію

$$f_1(a)=|a-a_1|+\ldots+|a-a_m|, \qquad (1')$$

якій відповідає метод найменших модулів (МНМ), тобто у задачі (1) мінімізуються сумарні абсолютні величини відхилень невідомого числа a від заданих чисел $a_1,\ldots,a_m$. Якщо $p=2$, то отримаємо функцію

$$f_2(a)=\sqrt{(a-a_1)^2+\ldots+(a-a_m)^2}, \qquad (1'')$$

якій з точністю до кореня квадратного відповідає метод найменших квадратів (МНК), тобто у задачі (1) мінімізуються сумарні квадрати відхилень невідомого числа a від заданих чисел $a_1,\ldots,a_m$. Якщо $p=\infty$, то отримаємо функцію

$$f_\infty(a)=\max_{i=1,\ldots,m}\{|a-a_i|\}, \qquad (1''')$$

якій відповідає мінімаксний (чебишевський) метод, тобто у задачі (1) мінімізується максимальна серед абсолютних величин m відхилень невідомого числа a від заданих чисел $a_1,\ldots,a_m$. Отже, кожному з наведених значень параметра p відповідає свій метод розв'язання

задачі (1) — МНМ ($p=1$), МНК ($p=2$), або мінімаксний (чебишевський) метод ($p=\infty$).

2. Властивості L_p -усереднених чисел. Розглянемо їх для значень $1 \le p \le 2$, коли задачу (1) можна спростити, опустивши знак степеня $1/p$. В результаті отримаємо таку задачу мінімізації опуклої функції: знайти

$$a_p^* = \underset{a \in R}{\operatorname{argmin}} \left\{ F_p(a) = \sum_{i=1}^{m} |a - a_i|^p \right\}, \qquad (2)$$

де p — скалярний параметр, такий що $1 \le p \le 2$. Тут $F_p(a)$ — опукла функція, яка є негладкою тільки при $p=1$. При цьому мінімальні значення функції $f_p^*(a_p^*)$ в задачі (1) та функції $F_p^*(a_p^*)$ у задачі (2) зв'язані співвідношенням $F_p^*(a_p^*) = (f_p^*(a_p^*))^p$.

На відміну від того, що для задачі (1) обов'язково використовувати тільки методи мінімізації негладких опуклих функцій [3], для розв'язання задачі (2) підійдуть методи мінімізації гладких опуклих функцій. Окрім того, якщо $p>1$, то функція $F_p(a)$ є строго опуклою, тобто для неї виконується умова

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \quad \forall x \ne y, 0 < \alpha < 1. \qquad (2')$$

Якщо $p=1$, то задача (2) відповідає МНМ та зводиться до задачі лінійного програмування. Якщо $p=2$, то задача (2) відповідає МНК та є задачею мінімізації квадратичної функції

$$F_2(a) = (a - a_1)^2 + \ldots + (a - a_m)^2. \qquad (2'')$$

При цьому оптимальні значення функцій в задачах (1) і (2) зв'язані співвідношенням $F_2^*(a_2^*) = (f_2^*(a_2^*))^2$.

Лема 1. Якщо p — таке, що $1 < p \le 2$, то задача (2) має єдиний розв'язок.

Доведення. Проведемо його методом від супротивного. Нехай a^* та a^{**} — два неспівпадаючі розв'язки задачі (2), яким відповідає оптимальне значення цільової функції $F_p^* = F_p(a^*) = F_p(a^{**})$.

Функція $F_p(a)$ при $1 < p \le 2$ є строго опуклою, тому, враховуючи (2′), для $F_p(a^{***})$ — значення функції $F_p(a)$ в точці $a^{***} = \lambda a^* + (1-\lambda)a^{**}$ справедливі співвідношення

$$F_p(a^{***}) = F_p(\lambda a^* + (1-\lambda)a^{**}) <$$
$$< \lambda F_p(a^*) + (1-\lambda)F_p(a^{**}) = \lambda F_p^* + (1-\lambda)F_p^* = F_p^*,$$

з яких випливає нерівність $F_p(a^{***}) < F_p^*$. Вона суперечить тому, що a^* та a^{**} розв'язки задачі (2), так як в точці a^{***} значення функції $F_p(a^{***})$ є меншим за мінімальне значення F_p^*. Лема 1 доведена.

Зазначимо, що при $p = 2$ задача (2) має аналітичний розв'язок $a_2^* = (a_1 + \ldots + a_m)/m$ — середнє арифметичне чисел $a_1, \ldots, a_m$, який випливає із мінімізації квадратичної функції (2′′).

3. Про L_p-усереднення лінійних сплайнів. Оптимізаційні задачі (1) та (2) використаємо для пошуку лінійного сплайна, який мінімально за L_p-нормою відрізняється від m лінійних сплайнів $y^1, \ldots, y^m$, які визначені значеннями $y_1^i, \ldots, y_n^i$, $i = 1, \ldots, m$ в одних і тих же базових точках $x_1 < \ldots < x_n$ інтервалу $[x_1, x_n]$. Невідомими у задачах будуть значення $y = (y_1, \cdots, y_n)^T$.

Якщо $p \geq 1$, то знаходженню мінімального за L_p-нормою лінійного сплайна y_p^* буде відповідати задача безумовної мінімізації опуклої негладкої функції: знайти

$$y_p^* = \underset{y \in R^n}{\operatorname{argmin}} \left\{ f_p(y) = \left(\sum_{j=1}^{n} \sum_{i=1}^{m} \left| y_j - y_j^i \right|^p \right)^{1/p} \right\}. \qquad (3)$$

Тут умова $p \geq 1$ гарантує опуклість функції $f_p(y_1, \cdots, y_n)$ від n змінних.

Для значень $1 \leq p \leq 2$ задачу (3) можна спростити, аналогічно тому, як це було зроблено для задачі (2). Спрощеній задачі для знаходження мінімального за L_p-нормою лінійного сплайна y_p^* відповідає така задача мінімізації опуклої функції: знайти

$$y_p^* = \underset{y \in R^n}{\operatorname{argmin}} \left\{ F_p(y) = \sum_{j=1}^{n} \sum_{i=1}^{m} \left| y_j - y_j^i \right|^p \right\}, \qquad (4)$$

де p — скалярний параметр, такий що $1 \leq p \leq 2$. Тут $F_p(y)$ — сепарабельна опукла функція, яка є негладкою тільки при $p = 1$. При цьому мінімальні значення функції $f_p^*(y_p^*)$ в задачі (3) та функції $F_p^*(y_p^*)$ у задачі (4) зв'язані співвідношенням $F_p^*(y_p^*) = (f_p^*(y_p^*))^p$.

Лема 2. Якщо p — таке, що $1 < p \le 2$, то задача (4) має єдиний розв'язок.

Доведення. Враховуючи, що функція $F_p(y)$ є сепарабельною за невідомими $y_1,\dots,y_n$, то задачу (4) можна переформулювати як таку задачу мінімізації строго опуклої функції: знайти

$$y_p^* = \sum_{j=1}^{n} \operatorname*{argmin}_{y_j \in R^1} \left\{ F_p(y_j) = \sum_{i=1}^{m} \left| y_j - f_i^j \right|^p \right\}, \qquad (4')$$

де p — скалярний параметр — такий, що $1 < p \le 2$. Застосовуючи до останньої n раз лему 1, отримаємо, що задача (4) має єдиний розв'язок. Лема 2 доведена.

При $p = 2$ аналогічно, як і у випадку задачі (2), задача (4) має аналітичний розв'язок $y_2^* = (y^1 + \dots + y^m)/m$ — середнє арифметичне m лінійних сплайнів $y^1,\dots,y^m$.

Висновки. В роботі розглянуто задачі безумовної мінімізації опуклих функцій для знаходження мінімальних за L_p-нормою лінійних сплайнів, використовуючи при цьому оптимізаційні задачі для знаходження числа, яке за L_p-нормою мінімально відрізняється від m заданих чисел $a_1,\dots,a_m$. У випадку, якщо $p \ge 1$, у задачах використовується негладка функція, а у випадку, якщо $1 < p \le 2$ — гладка функція. Якщо задачі (3) та (4) доповнити обмеженнями на властивості лінійного сплайна на окремих ділянках інтервалу (монотонність, опуклість, увігнутість, кривина), то можна автоматизувати вибір сплайн функцій для інтерполяції чи апроксимації кривих ліній гладкими функціями необхідної степені гладкості. Це може бути використано для побудови аеродинамічних профілів з заданими властивостями (кривина, ізогеометрія і т. п.).

Отримані в результаті оптимізаційні задачі можна ефективно розв'язувати за допомогою сучасних модифікацій *r*-алгоритмів — субградієнтних методів з розтягом простору в напрямку різниці двох послідовних субградієнтів [3, 4]. Використання методів мінімізації негладких функцій дає можливість будувати негладкі цільові функції, що значно розширює набір критеріїв оптимальності профілів, та включати негладкі функції в обмеження оптимізаційної задачі.

Матрично-векторні операції *r*-алгоритмів роблять їх перспективними в системах паралельних та розподілених обчислень. Наявні бібліотеки стандартних програм для паралельних матрично-векторних операцій дозволяють за короткий термін адаптувати алгоритми для ефективного

розв'язання оптимізаційних задач для лінійних сплайнів з використанням векторних процесорів на основі графічних прискорювачів (GPU).

Робота виконана за фінансової підтримки НАН України (проект № 0118U005227) та Volkswagen Foundation (грант No 90 306).

Список використаних джерел:

1. Мудров В. И., Кушко В. Л. Метод наименьших модулей. М. : Знание, 1971. 64 с.
2. Зоркальцев В. И. Метод наименьших квадратов: геометрические свойства, альтернативные подходы, приложения. Новосибирск : Наука, 1995. 220 с.
3. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. Киев : Наук. думка, 1979. 200 с.
4. Стецюк П. И. Теория и программные реализации *r*-алгоритмов Шора. *Кибернетика и системный анализ*. 2017. № 5. С. 43–57.

ON AVERAGING NUMBERS AND LINEAR SPLINES

Problems of unconstrained minimization of convex functions for finding the minimal linear splines in L_p -norm for cases $p \geq 1$ and $1 \leq p \leq 2$ are considered. They are constructed analogically with similar problems for finding a number that is different minimally in L_p -norm from the m given numbers $a_1, \ldots, a_m$. If $p \geq 1$, then the non-smooth function is used, and if $1 < p \leq 2$ then the smooth function is used. It is shown, that with a certain choice of parameter p, the optimization problems generate the known methods: the method of least squares, the method of least absolute deviations, and the Chebyshev minimax method. The properties of solutions of problems with $1 < p \leq 2$ are given.

Key words: *L_p -norm, the convex function, the non-smooth function, the method of least absolute deviations, the method of least squares.*

Одержано 15.02.2019

УДК 519.816
DOI: 10.32626/2308-5878.2019-19.168-174

Н. К. Тимофієва, д-р техн. наук
Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, м. Київ

КРИТЕРІЇ ПОДІБНОСТІ ДИНАМІЧНИХ ЗАДАЧ КОМБІНАТОРНОЇ ОПТИМІЗАЦІЇ

У комбінаторній оптимізації можна навести багато прикладів, коли задачі з різних класів розв'язуються за однією і тією ж обчислювальною схемою. Це пов'язано з тим що оговореним задачам властива подібність, завдяки якій вони розв'язуються одним методом або модифікацією одного і того ж алгоритму. Вона відрізняється від геометричної та описаної в теорії подібності. Для її встановлення проводиться аналіз задач різних класів з метою виявлення спільних ознак (критеріїв), за якими визначається їхня подібність. Використання цієї властивості дозволяє розробляти однакові методи та алгоритми для їхнього розв'язання. Задачі комбінаторної оптимізації, як правило, подібні за аргументом цільової функції, а задачі з комбінаторики — за способом утворення та впорядкування комбінаторних конфігурацій. Завдяки цій властивості їхні множини генеруються одним і тим же алгоритмом або його модифікацією.

У статті описано ознаки, за якими встановлюється подібність динамічних задач, що відносяться до різних класів. Задачі комбінаторної оптимізації, в яких у процесі їхнього розв'язання генерується поточна інформація, за якою оцінюється результат, а пошук оптимального розв'язку проводиться поетапно з обчисленням часткових сум цільової функції, названо динамічними. Основними ознаками подібності для них є зміна результату розв'язання в часі та для його поточного відліку необхідність обчислення часткової цільової функції. Процес їхнього розв'язання описується орієнтованим ациклічним графом, а часткові значення цільової функції змінюються в часі та обчислюються за рекурентними правилами. При знаходженні їхніх оптимальних значень виконується принцип Беллмана. Виявлені властивості подібності, які характерні для задач цього класу, визначають їхню універсальність, завдяки якій вони розв'язуються одним і тим же методом. Для розв'язання цих задач, як правило, використовують динамічне програмування. Вивчення та використання цієї властивості в комбінаторній оптимізації в подальшому дозволить зводити нерозв'язні задачі до розв'язних. Наведено приклади деяких динамічних задач комбінаторної оптимізації.

Ключові слова: *комбінаторна оптимізація, комбінаторна конфігурація, динамічні задачі комбінаторної оптимізації, подібність задач комбінаторної оптимізації, цільова функція.*

Вступ. Властивість подібності вивчають в геометрії, але вона характерна і для різноманітних фізичних явищ. В комбінаторній оптимізації також має місце подібність, яка пов'язана з тим, що для розв'язання задач різних класів використовують універсальні методи та алгоритми. Для її встановлення необхідно провести аналіз задач комбінаторної оптимізації різних класів та виявити ознаки, за якими вони розв'язуються за однією і тією ж обчислювальною схемою.

Аналіз останніх досліджень та публікацій за темою. В комбінаториці та комбінаторній оптимізації можна навести багато прикладів, коли задачі з різних класів розв'язуються за однією і тією ж обчислювальною схемою, наприклад [1, 2]. Ця властивість в літературі достатньою мірою не висвітлена, хоча існуючі універсальні методи орієнтовані на розв'язання різноманітних таких задач. У роботі [3] наведено деякі ознаки, за якими встановлюється подібність задач в комбінаторній оптимізації, що дає можливість розробляти універсальні методи та алгоритми. Тому однією з проблем у теорії комбінаторної оптимізації є виявлення критеріїв подібності з метою узагальнення та використання для їхнього розв'язання універсальних підходів.

Загальна математична постановка задачі комбінаторної оптимізації. Задачі комбінаторної оптимізації, як правило, задаються на одній або кількох множинах, наприклад $A=\{a_1,...,a_n\}$ та $B=\{b_1,...,b_{\tilde{n}}\}$, елементи яких мають будь-яку природу [4]. Назвемо ці множини *базовими*. Наявні два типи задач. В *першому* типі кожну з цих множин подамо у вигляді графа, вершинами якого є її елементи, а кожному ребру поставлено у відповідність число $c_{lt}\in R$, яке називають вагою ребра (R — множина дійсних чисел); $l\in\{1,...,n\}$, $t\in\{1,...,\tilde{n}\}$, n — кількість елементів множини A, $\tilde{n}$ — кількість елементів множини B. Покладемо, що $n=\tilde{n}$. Між елементами цих множин існують зв'язки, числове значення яких назвемо вагами. Величини $c_{lt}\in R$ — *вхідні* дані, які задамо матрицями. В *другому* типі задач між елементами заданої множини зв'язків не існує, а вагами є числа $v_j\in R$, $j\in\{1,...,n\}$, яким у відповідність поставлено деякі властивості цих елементів, числові значення яких задаються скінченними послідовностями, що також є вхідними даними.

Для обох типів задач із елементів однієї або кількох базових множин утворюється комбінаторна множина W — сукупність комбінаторних конфігурацій певного типу (перестановки, вибірки різних типів, розбиття тощо). На елементах w комбінаторної множини W вводиться цільова функція $F(w)$. Необхідно знайти елемент w^* множини W, для якого $F(w)$ набуває екстремального значення при виконанні заданих обмежень.

За способом обчислення цільової функції виділимо задачі, в яких для певного варіанту розв'язку її значення обчислюється одночасно. Такі задачі назвемо статичними. Задачі, в яких в процесі їхнього розв'язання генерується поточна інформація, за якою оцінюється результат, а пошук оптимального розв'язку проводиться поетапно з обчисленням часткових сум цільової функції, назвемо динамічними.

Для моделювання прикладних задач в рамках теорії комбінаторної оптимізації необхідно:

1) визначити вид задачі (статична або динамічна);
2) визначити базові множини, якими задається певна задача;
3) за вхідними даними визначити її тип;
4) визначити аргумент цільової функції (комбінаторну конфігурацію);
5) змоделювати цільову функцію.

Під комбінаторною конфігурацією розуміємо будь-яку сукупність елементів, яка утворюється з усіх або з деяких елементів заданої базової множини $A = \{a_1, ..., a_n\}$ [6]. Позначимо її впорядкованою множиною $w^k = (w_1^k, ..., w_\eta^k)$, де $\eta \in \{1, ..., n\}$ — кількість елементів у w^k, $W = \{w^k\}_1^q$ — множина комбінаторних конфігурацій. Верхній індекс k ($k \in \{1, ..., q\}$) у w^k позначає порядковий номер w^k у W, q — кількість w^k у W.

Ознаки подібності динамічних задач комбінаторної оптимізації. Основними ознаками подібності для динамічних задач є зміна результату розв'язання в часі та для його поточного відліку обчислення часткової цільової функції. Процес їхнього розв'язання описується орієнтованим ациклічним графом, а часткові значення цільової функції змінюються в часі та обчислюються за рекурентними правилами. При знаходженні оптимального значення часткової цільової функції виконується принцип Беллмана. Аргументом цільової функції в них є вибірки різних типів, а також розбиття n-елементної множини на підмножини. Вони, як правило, розв'язуються одним і тим же методом — динамічним програмуванням.

Динамічні задачі комбінаторної оптимізації. До динамічних задач відносяться такі задачі: сегментація та розпізнавання мовленнєвих сигналів, задача Джонсона з теорії розкладів, задача класифікації, здача збереження довкілля та ін.

Задача Джонсона з теорії розкладів. Найпростіша задача з теорії розкладів (задача Джонсона) формулюється так [2, 5]. Задано n деталей. Кожна з деталей повинна пройти послідовну обробітку на m машинах. Кожна машина також виконує одну операцію. Необхідно скласти такий

розклад обробітку деталей, щоб затрачений на ці операції час був би мінімальний за умови, що він не перевищує заданої величини T.

У цій задачі задано дві множини A і B, між елементами яких існує певна залежність, числові значення якої назвемо вагами. Подамо їх несиметричною матрицею C розмірністю $\tilde{n} \times n$, де величина c_{sl} відповідає значенню часу, який необхідно затратити на обробку l-ї деталі s-ю машиною. Час послідовної обробки всіх n елементів множини A за будь-якого розкладу, який би не перевищував заданої величини T, невідомий, тому спочатку для вибірки із n елементів $a_l \in A$ по n знаходимо перестановку, для якої значення цільової функції — мінімальне і не перевищує величини T. Якщо одержаний розв'язок не задовольняє цій умові, то задача розв'язується для вибірки із n елементів $a_l \in A$ по η. З цього випливає, що аргументом цільової функції в розглянутій задачі є розміщення без повторень, яке утворюється шляхом знаходження сполучення із n елементів по η, для якого генеруються $\eta!$ перестановок, $\eta \in \{1, ..., n\}$.

Для i-го сполучення уведемо комбінаторну матрицю $Q(\mu^i)$ розмірністю $\tilde{n} \times \eta$, в яку входять стовпці матриці C, номера яких збігаються з номерами елементів множини A, з яких утворено сполучення без повторень $\mu^i \in \mathrm{M}$, $i \in \{1, ..., 2^n - 1\}$, M — множина сполучень. Із фіксованої матриці $Q(\mu^{i*})$ утворимо $\eta!$ комбінаторних матриць $Q'(\mu^{i*}, \omega^k)$, які залежать від перестановки $\omega^k = (\omega_1^k, ..., \omega_\eta^k) \in \Omega$, $k \in \{1, ..., \eta!\}$, $\eta \in \{1, ..., n\}$, Ω — множина перестановок. Цільова функція в задачі планування з теорії розкладу набуде вигляду

$$F(\mu^{i*}, \omega^k) = \sum_{l=1}^{\eta} \sum_{s=1}^{\tilde{n}} g_{sl}(\mu^{i*}) + \sum_{l=1}^{\eta-1} \sum_{s=2}^{\tilde{n}} \left| g'_{sl}(\mu^{i*}, \omega^k) - g'_{s-1\,l+1}(\mu^{i*}, \omega^k) \right|, \quad (1)$$

де $\sum_{l=1}^{\eta} \sum_{s=1}^{\tilde{n}} g_{sl}(\mu^{i*})$ — постійна для будь-якої з $\eta!$ перестановок величина, що визначає затрачений час на обробку деталей, який задано за умовою. Вона не залежить від перестановки, а змінюється в залежності від варіанту сполучення μ^i; $\sum_{l=1}^{\eta-1} \sum_{s=2}^{\tilde{n}} \left| g'_{sl}(\mu^{i*}, \omega^k) - g'_{s-1\,l+1}(\mu^{i*}, \omega^k) \right|$ — сумарний час простою машин. Ця величина — змінна і залежить як від варіанту сполучення μ^i так і від перестановки $\omega^k = (\omega_1^k, ..., \omega_\eta^k)$. За виразом (1) визначається сумарне значення цільової функції.

Задача Джонсона полягає у знаходженні таких μ^{i*} і ω^{k*}, для яких значення $F(\mu^{i*}, \omega^{k*})$ було б мінімальним і $F(\mu^{i*}, \omega^{k*}) \le T$. Процес її розв'язання описується орієнтованим ациклічним графом, а часткові значення цільової функції змінюються в часі і обчислюються за рекурентними правилами. При обчисленні часткової цільової функції для неї виконується принцип Беллмана.

Задача розпізнавання мовлення та задача сегментації мовленнєвого сигналу [6]. Задача сегментації мовленнєвих сигналів полягає у виділенні на заданому відрізку вхідного сигналу майже періодичних та неперіодичних ділянок, а в майже періодичних визначаються довжини поточного майже періоду. Розпізнавання мовлення — це процес автоматичної обробки мовленнєвого сигналу з метою визначення послідовності слів, яка передається цим сигналом. Вона полягає у знаходженні для вхідного сигналу найбільш правдоподібного еталону з усіх можливих еталонних сигналів.

Мовленнєвий сигнал передає мовлення людини в якому спостерігаються ділянки майже періодичні, які моделюють голосні та приголосні звуки, та неперіодичні (шумні звуки). Подамо мовленнєвий сигнал дискретною функцією $f(j)|_1^m$, де m — кількість її значень (відліків сигналу) та проведемо його сегментацію на майже періодичні та неперіодичні ділянки, а в майже періодичних визначимо довжини поточного майже періоду.

Відрізок сигналу, що досліджується, розіб'ємо на ділянки довжиною $L \in \{L_{\min}, L_{\min} + \Delta, L_{\min} + 2\Delta, ..., L_{\max}\}$ з наступним визначенням періодичності сусідніх ділянок, $L_{\min}$ — мінімально можлива довжина майже періоду, $L_{\max}$ — максимально можлива довжина майже періоду, Δ — значення приросту майже періоду (визначається експериментально). За еталонний сигнал приймемо попередню ділянку. При розпізнаванні мовлення для вхідного сигналу знаходиться в бібліотеці подібний еталонний сигнал. Оскільки задача сегментації мовленнєвого сигналу та розпізнавання мовлення — динамічні, то їх розв'язують динамічним програмуванням з використанням кореляції функції $f(j)|_1^m$. Ці задачі описуються орієнтованим ациклічним графом, часткові значення цільової функції в ній змінюються в часі та обчислюються за рекурентними правилами. При знаходженні оптимального значення часткової цільової функції виконується принцип Беллмана.

Висновки. Отже, основними ознаками подібності для динамічних задач комбінаторної оптимізації є зміна результату розв'язання в часі та для нього обчислення часткової цільової функції. Процес їх-

нього розв'язання описується орієнтованим ациклічним графом, а часткові значення цільової функції змінюються в часі та обчислюються за рекурентними правилами. При знаходженні оптимального значення часткової цільової функції виконується принцип Беллмана.

Список використаних джерел:

1. Липский В. Комбинаторика для программистов. Пер. с польск. М. : Мир, 1988. 213 с.
2. Сергиенко И. В., Каспшицкая М. Ф. Модели и методы решения на ЭВМ комбинаторных задач оптимизации. Киев : Наук. думка, 1981. 281 с
3. Тимофієва Н. К. Про подібність задач комбінаторної оптимізації та універсальність алгоритмів. *Системні дослідження та інформаційні технології*. 2013. № 4. С. 27–37.
4. Тимофієва Н. К. Теоретико-числові методи розв'язання задач комбінаторної оптимізації. Автореф. дис. ... докт. техн. наук. Ін-т кібернетики ім. В. М. Глушкова НАН України. Київ, 2007. 32 с.
5. Тимофієва Н. К., Гриценко В. И. Розв'язання задачі планування з теорії розкладу методом структурно-алфавітного пошуку та гібридним алгоритмом. *УСиМ.* 2011. № 3 С. 21–36.
6. Винцюк Т. К. Анализ, распознавание и интерпретация речевых сигналов. Киев : Наук. думка, 1987. 262 с.

CRITERIA OF SIMILARITY OF DYNAMIC PROBLEMS OF COMBINATORIAL OPTIMIZATION

In combinatorial optimization, you can cite many examples when the problems from different classes are solved according to the same computational scheme. This is due to the fact that the specified problems have similarities, due to which they are solved by one method or modification of the same algorithm. It differs from geometric and described in the theory of similarity. For its establishment, an analysis of the problems of different classes is conducted in order to identify common features (criteria) that determine their similarity. Using this property allows you to develop the same methods and algorithms for their solution. The problems of combinatorial optimization, as a rule, are similar of the argument of the objective function, and the problems of combinatorics are similar on the creating and arranging of combinatorial configurations. Due to this property, their sets are generated by the same algorithm or it modification.

The article describes the criterias by which the similarity of dynamic problems relating to different classes is established. The problems of combinatorial optimization, in which in the process of their solution, the current information by which the result is evaluated is generated, and the search for an optimal solution is carried out in stages with the calculation of partial amounts of the objective function, is called dynamic. The main criterias of similarity to them is the change in the result of the solution in time and for its current reference, the need to calculate the partial objective function.

The process of their solution is described by a directed acyclic graph, and the partial values of the objective function change over time and are

calculated according to the recurrent rules. When finding their optimal values, the Bellman principle is followed. The revealed properties of similarity, which are characteristic of the problems of this class, determine their universality, through which they are solved by the same method. Typically, dynamic programming is used to solve these problems. The study and use of this property in combinatorial optimization in the future will allow solving insoluble problems for solvable ones. Examples of some dynamic combinatorial optimization problems are given.

Key words: *combinatorial optimization, combinatorial configuration, dynamic combinatorial optimization problems, similarity of combinatorial optimization problems, objective function.*

Одержано 24.01.2019

УДК 519.1,514.128
DOI: 10.32626/2308-5878.2019-19.174-180

В. О. Устименко*,**, д-р фіз.-мат. наук, професор,
О. С. Пустовіт**

*Університет Марії Кюрі-Склодовської, м. Люблін, Республіка Польща,
**Інститут телекомунікацій і глобального інформаційного простору НАН України, м. Київ

ПРО НОВІ ПОТОКОВІ АЛГОРИТМИ СТВОРЕННЯ ДАЙДЖЕСТІВ ЕЛЕКТРОННИХ ДОКУМЕНТІВ З ВИСОКОРІВНЕВИМ АВАЛАНЧ ЕФЕКТОМ

Пропонується родина залежних від ключа швидких алгоритмів створення дайджестів електронних документів Комп'ютерна симуляція дозволяє дослідити високий рівень аваланч ефекту, що виникає. Нехай K — вільно обране скінчене комутативне кільце, m — додатне ціле число. Алгоритми використовують нещодавно знайдені гомоморфні відображення компресії функцій вільної напівгрупи потенційно нескінчених текстів у алфавіті K на скінчену групу кубічних поліноміальних перетворень m вимірного афінного простору K_m.

Криптографічна стабільність функцій хешування пов'язується зі складними алгебраїчними проблемами, такими як дослідження систем алгебраїчних рівнянь великої степені та задача розкладу нелінійного відображення вільного модуля за заданими твірними.

Для пришвидшення алгоритму дайджестом слова $p = (p_1, p_2, ..., p_n)$, $p_i \in K$ вважатимемо не саме кубічне перетворення $F = \psi(p)$, але його значення $F(w(p))$ на деякому за-

лежному від p векторі $w(p)$ деформоване множенням на псевдовипадкову матрицю M. Алгоритми імплементовано у випадках скінченних полів $F_2^8, F_2^{16}, F_2^{32}$, кілець Z_{256} та $B(32)$ (булеве кільце порядку 2^{32}).

Пропоновані алгоритми можуть працювати з даними у вигляді тексту, відео та аудіо файлів, фільму тощо. Розроблені методи створення дайджестів мають потоковий характер — швидкодія при сталому m лінійно залежить від n. Зростання n збільшує криптографічну стабільність. Імплементації у блоковому режимі можлива, але не вмотивована, бо розмір блоку обмежує кількість змінних системи нелінійних рівнянь.

Необхідність подальших досліджень і технологічних розробок по створенню нових залежних від ключа хеш-функцій пов'язана із викликами кібербезпеки, зростанням глобального інформаційного простору, очікування появи квантового комп'ютера та розвитком технологій bitcoins, де потрібно хешувати вхідні дані довільного розміру, перетворюючи їх у послідовність бітів, що є дайджестом так званих blockchains. Запропоновані алгоритми створення чутливих до змін дайджестів документів будуть використані для виявлення кібератак та аудиту усіх файлів системи після зареєстрованого втручання.

Ключові слова: *кібербезпека, хеш функції, автентифікаційні коди повідомлень, гомоморфізм компресії, високо нелінійна криптографія від багатьох змінних, некомутативна криптографія.*

1. Про верифікацію електронних документів. Важливою категорією інформаційного простору є довіра до документів. Легко побачити, що навіть користування надійними засобами шифрування не забезпечує повної довіри до документів, тому що треба рахуватися із шумами у каналах та проблемами безпечного збереження файлів у електронних сховищах, де документи можуть бути підроблені, пошкоджені комп'ютерними вірусами, технічними помилками в роботі обчислювальної техніки та інше. Зазначимо, що останнім часом постійно зростає загроза потужних кібертерористичних атак на сховища, їх наслідки це не тільки виток інформації, але й ушкодження або фальсифікування документів. Зрозуміло, що після виявлення кібератаки потрібно робити аудит усіх файлів системи.

Для задач виявлення кібератак, верифікації та автентифікації документів потрібні так звані залежні від ключів хеш-функції (автентифікаційні коди повідомлень або МАСи) які залежать. від гасла [2 с. 244–257]. Хеш-функція потрібна для генерації скомпенсованої форми оригінального документа довільно обраного розміру. Таку форму називають хешем або дайджестом документа, її використовують у різних криптографічних застосуваннях. Хеш-функція h працює з файлом довільного розміру n, її значення має фіксований розмір.

Для інших задач захисту інформації потрібна загальна хеш-функція, що не потребує ключа або ж гасла. Нещодавно сертифіковано загальну хеш-функцію Купина, як новий державний стандарт України [1].

2. Вимоги до дайджесту документів. Криптографічно стабільна функція хешування f має забезпечувати: практичну неможливість вибору пари посилань x та z таким самим значенням хеш-функції. Для дайджесту документа, створеного залежною від ключа хеш-функцію (MAC) використовують символ HMAC. Коли користувачі хочуть безпечно обмінятися кореспонденцією, перевіряючи хто є дійсним автором листа, так і відсутність змін при пересилці, вони разом обирають спільний MAC. При цьому користуються спільною схемою симетричного шифрування.

Крім криптографічної стабільності дуже важлива швидкодія та високий показник аваланч ефекту. Цей ефект вимірюється таким чином. Обчислюється HMAC для генерованого файлу, змінюється довільний його біт та обчислюється HMAC для зміненого файлу, після цього робиться побітове порівняння отриманих дайджестів. Для практичного вживання MACу потрібно, щоб статистичні дослідження показали, що поєдина зміна символу приводить до зміни 40% бітів HMACу незалежно від розміру файлів, що генеруються.

3. Математичне підгрунтя хеш-функції, що пропонується. Нехай $F(K)$ — простір потенційно нескінченних текстів в алфавіті K, який являє сукупність всіх кортежів виду $(a_1, a_2, ..., a_k), a_i \in K$ різної довжини k. Будемо вважати, що K є скінченним комутативним кільцем та ототожнювати $F(K)$ з напівгрупою із наступним множенням $(a_1, a_2, ..., a_k) \circ (b_1, b_2, ..., b_s) = (a_1, a_2, ..., a_k, b_1 + a_k, b_2 + a_k, b_s + a_k)$. Нехай $F'(K)$ буде підпівгрупою всіх слів парної довжини. Позначимо $S(K^n)$ скінченну напівгрупу всіх поліноміальних відображень простору K^n в себе.

Наш алгоритм грунтується на наступному математичному твердженні.

Теорема. ([3]) Для кожного натурального $m \geq 2$ існує гомоморфне відображення $\psi : F'(K) \to S(K^m)$ таке, що його образ $\psi(F'(K))$ утворює групу G кубічних поліномів ступеня 3.

Нагадаємо, що властивість гомоморфного відображення для $\psi = \psi_m$ записується як $\psi(a \circ b) = \psi(a) \circ \psi(b)$.

Відображення, що задовольняє умовам теореми будується конструктивно в термінах теорії дискретних динамічних систем, визначе-

них за алгебраїчними графами з екстремальними властивостями [4]. Ці методи дозволяють отримати таку нижню оцінку порядку конструктивно побудованої групи: $|G| \geq 2^{4n}$. Зазначимо, що твердження визначає рідкісний математичний об'єкт. Суперпозиція двох кубічних відображень з великою ймовірністю буде мати ступінь 9, трьох — 27, чотирьох — 81, а у побудованій групі всі ці добутки обмежені числом 3. Ця група вже вживалася для побудови криптографічних алгоритмів з приватним ключем [5, 6], та протоколів обміну ключами [4, 7, 10].

Для створення МАСу [9] використано не саму групу *G*, а відображення ψ, що її визначає, разом з афінними *A* та *B* перетвореннями групи Кремони за правилом $g: x \to A\psi(x)B$. Не важко побачити, що ψ — природній оператор компресії даних який відображає нескінченну множину $F'(K)$ усіх слів парної довжини в алфавіті *K* на скінченну множину $S(K^m)$. На вихід подається список координат *g*(*x*), до яких двічі застосовано оператор повного диференціалу. Комп'ютерна симуляція дозволила обчислити дуже високий аваланч ефект у межах 97–99 %. Для прикладу в МАСу російських дослідників інтервал аваланч ефекту оцінюється як 47–50 % [8].

4. Пришвидшення алгоритмів. У доповіді на симпозіумі буде представлено модифікацію описаного алгоритму у випадку $K = Z_{256}$, що дозволяє зберегти (або ж поліпшити) рівень аваланч ефекту при значному підвищенні швидкодії.

Нехай $(a_1, a_2, ..., a_n)$ документ представлений в алфавіті *K* після перемішування з деяким псевдовипадковим словом сталої довжини. Будемо вважати, що число *n* парне. Користувачі обирають розмір дайджесту $m, m < n$ та $m = O(1)$ або ж $m = O(n)$ разом з ключем, що складається зі зростаючої послідовності натуральних чисел $i(1), i(2), ..., i(m-1)$ та невиродженої матриці *M* складеної з елементів кільця лишків Z_{256}. Вони утворюють вектор $u = (v_1, v_2, ..., v_m)$, де $v_1 = a_1 + a_2 + ... + a_n, ..., v_j = v_{j-1} - a_{i(j-1)}$. Потім обчислюється кубічне відображення $F = \psi_m(a_1, a_2, ..., a_n)$, яке кореспонденти застосовують до вектора *u*. Отриманий вектор-рядок $F(u)$ множиться на матрицю *M*. Вектор $w = F(u)M$ вважаємо дайджестом документу.

Зазначимо, що значення $F(u)$ обчислюється за допомогою рекурсивного алгоритму, його складність визначається як $O(mn)$ і співпадає зі складністю створення дайджесту.

Цей базовий алгоритм легко модифікувати без змінення складності обчислень. Зокрема:

1) можна представити слово $(a_1, a_2, ..., a_n)$ у вигляді конкатенації скінченої кількості слів $z_1, z_2, ..., z_t$ парної довжини. Потім обрати послідовність слів вигляду $u_1, u_2, ..., u_k$, де $u_i \in \{z_1, z_2, ..., z_t\}$ таку, що кожне z_i у цій послідовності зустрічається не менше ніж один раз. Далі обчислюється значення y добутку $u_1, u_2, ..., u_k$ у розглянутій вище напівгрупі слів $F'(K)$. Алгоритм модифікується заміною кубічного відображення $\psi(a)$ на $\psi(y)$. При умові всім відомого розбиття файлу криптографічна стабільність такого дайджесту буде залежною від проблеми розкладу $\psi(y)$ у добуток перетворень $\psi(z_i)$ з аффінної групи Кремони. Зазначимо, що поліноміального алгоритму для розв'язання цієї проблеми на звичайному або квантовому комп'ютері на сьогоднішній день не знайдено. Насправді ця задача виникає за умов неповної визначеності, бо відоме тільки значення $\psi(y)$ на деякому залежному від файла векторі. Зрозуміло що розбиття a на підслова z_i та послідовність u_j слід вважати частиною спільного ключа для кореспондентів;
2) можна обчислювати v_1 як добуток виразів $2a_i + 1$ та отримувати v_i діленням v_{i-1} на $2a_{i(j-1)} + 1$;
3) у варіанті 2 можна замінювати v_i на його непарні степені $k, k < 128$. Тоді ці степені слід вважати параметрами ключа.

Імплементовані випадки зручні для їх використання у технології *blockchain*, де потрібні дайджести у вигляді послідовності бітів або ж нулів та одиниць.

Висновки. Зазначимо, що добрі властивості функції компресії грунтуються на конструкціях гомоморфізмів нескінченної напівгрупи слів парної довжини у напівгрупи Кремони, дискретних динамічних систем, визначених родинами алгебраїчних графів з екстремальними властивостями та комп'ютерних моделях конденсованих систем.

Список використаних джерел:

1. Oliynykov R., Gorbenko I., Kazymyrov O., Ruzhentsev V., Kuznetsov O., Gorbenko Yu., Dyrda O., Dolgov V., Pushkaryov A., Mordvinov R., Kaidalov D. Data Security. Symmetric block transformation algorithm. Ministry of Economical Development and Trade of Ukraine. DSTU 7624:2014. National Standard of Ukraine. Information technologies. Cryptographic. 2015.

2. Aumasson J. Ph, Serious Cryptography: A Practical Introduction to Modern Encryption, No Starch Press. 2017. 312 p.
3. Ustimenko V. On new symbolic key exchange protocols and cryptosystems based on hidden tame homomorphism. *Dopov. Nac. akad. nauk Ukraine.* 2018. N 10. P. 26–36.
4. Устименко В. А. Об экстремальной теории графов и символьных вычислениях. *Докл. НАН Украины*. 2012. № 11. С. 15–21.
5. Пустовіт О., Устименко В., Про застосування алгебраїчної комбінаторики до проблем кодування та криптографії. *Математичне моделювання в економіці*. Київ, 2017. № 3. С. 31–46.
6. Ustimenko V, Roma´nczuk-Polubiec U., Wr´oblewska A., Polak M., Zhupa E., On the implementation of new symmetric ciphers based on non-bijective multivariate maps. *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems. ACSIS*. 2018. Vol. 15. P. 397–405.
7. Устименко В. О., Пустовіт О. С. Про нову концепцію електронного підпису та засоби її реалізації. *Колективна монографія за матеріалами XVI Міжнародно-практичної конференції.* м. Київ (Пуща-Водиця). 2017. С. 86–89.
8. Krendelev S., Sazonova P., Parametric Hash Function Resistant to Attack by Quantum Computer, Based on Problem of Solving a System of Polynomial Equations in Integers. *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems. ACSIS.* 2018. Vol. 15. P. 387–390.
9. Устименко В. О., Пустовіт О. С. Про нові алгоритми аудиту електронних документів, їх імплементацію та застосування у кібербезпеці. *Колективна монографія за матеріалами XVII Міжнародно-практичної конференції.* м. Київ (Пуща-Водиця). 2018. С. 170–174.
10. Ustimenko V., Klisowski M. On Noncommutative Cryptography with cubical multivariate maps of predictable density. *Proceedings of the 2019 Computing Conference.* London. July, 2019 (to appear).

ON NEW STREAM ALGORITHMS FOR GENERATING DOCUMENTS DIGESTS WITH HIGH AVALANCHE EFFECT

The family of key dependent algorithms for generating digests of electronic documents is proposed. Computer simulation allows to investigate high level of corresponding avalanche effect. Let K be a freely chosen finite commutative ring and m be a positive integer. Algorithm uses recently discovered homomorphic compression maps of free semigroup of potentially infinite texts written in the alphabet K onto finite group of cubic polynomial transformations of affine space K_m.

Cryptographic stability of proposed hash functions is connected with hard algebraic problems such as investigation of systems of algebraic equalities or decomposition of nonlinear map on free module into given generators.

To make algorithm faster instead of cubical transformation $F = \psi(p)$ we take as digest its value $F(w(p))$ on some depending from $p = (p_1, p_2, ..., p_n)$ word $w(p)$ additionally transformed by multiplication on pseudorandom matrix M. The algorithms are implemented in the cases of finite fields $F_2^8, F_2^{16}, F_2^{32}$, commutative ring Z_{256} and Boolean ring $B(32)$ of order 2^{32}.

Proposed algorithms can work with data in the form of texts, audio and video files, files with various extensions such as .avi, .tif, .pdf and etc. Algorithms can generate digests of already encrypted files, this option gives a possibility to check the integrity of files without their decryption. Suggested methods of digest generation have a stream nature, the speed for constant m is linearly dependent on variable n. Growth of n increases the cryptographic stability. The implementation in the form of block by block compression is possible but it has a lack of motivation because the size of the block restricts the number of variables in the system of nonlinear equations.

The necessity of a further research and technological solutions on the constructions of key dependent hash functions is caused by cybersecurity calls, the increase of global information space, expectations of quantum computers appearance and development of bitcoins technology, which requires hashing of data of arbitrary size with its transformation into sequences of bits which form digests of the so called blockchains. Proposed algorithms of generation of sensitive for changes of documents digests will be used for cyberattacks detection and for the auditing of all files after registered intrusion.

Key words: *cybersecurity, hash functions, message authentication codes, homomorphism of compression, highly nonlinear multivariate cryptography, noncommutative cryptography.*

Одержано 27.01.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.180-187

О. М. Хіміч, член-кореспондент НАН України, д-р фіз.-мат. наук,
В. А. Сидорук, канд. фіз.-мат. наук

Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ВИКОРИСТАННЯ МІШАНОЇ РОЗРЯДНОСТІ У МАТЕМАТИЧНОМУ МОДЕЛЮВАННІ

У роботі пропонується методика, за допомогою якої можна прискорити час математичного моделювання складних систем, використовуючи мішану розрядність в обчисленнях. Мішана розрядність дозволяє підвищити продуктивність обчислень, а також зекономити пам'ять. Показано застосування такого підходу при побудові алгоритму розв'язання систем лінійних алгебраїчних рівнянь з розрідженими матрицями.

Ключові слова: *математичне моделювання, мішана розрядність, паралельні алгоритми, розріджені матриці.*

Вступ. Проведення обчислень на довільній розрядності — один з основних інструментів підвищення ефективності програм і ідентифікації лінійних систем у комп'ютерному середовищі [1], зокрема,

виявлення поганообумовлених чи некоректних задач, задач з близькими чи кратним власними значеннями. Обчислення на довільній розрядності реалізуються програмно, а час виконання обчислень експоненціально залежить від розрядності числа з плаваючою комою, що використовується. Програмні бібліотеки такі як GNU Mpfr [2], GNU GMP [3] служать для реалізації обчислень на довільній розрядності. У бібліотеці mpfr, реалізовано ряд blas функцій.

В обчислювальних системах різних архітектур апаратно реалізовані половинна, одинарна, подвійна та розширена точність обчислень. У наукових розрахунках переважна більшість обчислень проводиться з подвійною точністю, тому в багатьох високопродуктивних процесорах одинична точність була усунена на користь емуляції операцій з одинарною точністю з використанням схеми подвійної точності. Водночас, використання даних у форматі з одинарною точністю дозволяє зберігати вдвічі більше даних на кожному рівні ієрархії пам'яті, включаючи кеш-пам'ять та регістрову пам'ять. До того ж, обробка значення одинарної точності вимагає використання меншої ширини смуги пропускання між різними рівнями пам'яті і зменшує кількість необхідного кешу і величину TLB промахів.

Враховуючи наведене вище, можна зробити висновок, що одним з шляхів підвищення ефективності обчислень і скорочення часу математичного моделювання є побудова паралельних алгоритмів, які б в процесі роботи виконували б обчислення на різній розрядності.

Розглянемо застосування такого підходу при побудові алгоритму розв'язання систем лінійних алгебраїчних рівнянь з розрідженими матрицями.

Постановка задачі. Розглянемо задачу

$$Ax = b \tag{1}$$

з симетричною додатно-визначеною розрідженою матрицею порядку n.

Однією з передумов розв'язання задачі (1) на комп'ютерах MIMD-архітектури з багатоядерними процесорами (CPU), (зокрема, з процесорами Intel Xeon Phi) є приведення матриці до такого виду, який дозволяє працювати з більш щільними групами ненульових елементів. Перетворення матриці зумовлене особливостями ієрархії пам'яті для систем з процесорами Intel Xeon Phi.

На даний момент існує велика кількість алгоритмів перевпорядкування елементів матриць: метод мінімальної степені, метод вкладених перерізів, метод паралельних перерізів і т. д. Кожен з цих алгоритмів дозволяє привести довільну розріджену структуру до більш регулярного вигляду. Найбільш зручну для паралельної обробки структуру матриці отримується після застосування до матриці методу вкладених, або паралельних, перерізів.

$$\tilde{A} = P^T AP = \begin{pmatrix} A_{11} & 0 & 0 & A_{1p} \\ 0 & A_{22} & 0 & A_{2p} \\ 0 & 0 & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{pmatrix},$$

де P — матриця перестановок, p — кількість діагональних блоків у матриці, блоки A_{pp}, A_{ip}, A_{pi}, A_{ii}, $i = \overline{1, p-1}$ зберігають розріджену структуру.

Таким чином, задача розв'язання (1) зводиться до розв'язування еквівалентної системи

$$\tilde{A}\tilde{x} = \tilde{b}, \tag{2}$$

де $\tilde{x} = P^T x$, $\tilde{b} = P^T b$.

Найбільш ефективним прямим методом розв'язання (2) є метод Холецького [4–6]. В статті буде розглянуто паралельний алгоритм який відповідає саме етапу факторизації матриці.

Паралельний алгоритм. Розіб'ємо матрицю A на блоки розмірністю $c \times c$. Далі для факторизації блочно-діагональної матриці застосуємо алгоритм запропонований в [7] для щільних матриць.

Для факторизації матриці на k-му кроці використаємо наступне співвідношення:

$$A^k = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix}, \tag{3}$$

де розмірності блоків A_{11} — $c \times c$, A_{12} — $(n-kc)c$, A_{22} — $(n-kc)(n-kc)$, блоки A_{12} та A_{22} враховують структуру діагональних блоків та блоків обрамлення.

Звідси отримаємо алгоритм, за яким проводиться розвинення на k кроці:

$$A_{11} = L_{11} * L_{11}^T; \tag{4}$$

$$L_{21} = A_{21} * \left(L_{11}^T\right)^{-1}; \tag{5}$$

$$\tilde{A}_{22} = A_{22} - L_{21} * L_{21}^T. \tag{6}$$

Зазначимо, що реалізація (4)–(6) на кожному кроці модифікує тільки блоки D_{ii}, C_{pi}, $i = \overline{1, p-1}$, D_{pp}.

Нехай для розв'язування задачі на комп'ютері MIMD-архітектури маємо CPU з p процесорними ядрами. Для роботи алго-

ритму на *к* кроці реалізовується наступна декомпозиція даних: у пам'яті CPU (*i*) зберігається плитка A_{11} та плитки необхідні для модифікації підматриці A_{22}. $A_{pp}^{(i)}$ — набір плиток для модифікації діагонального блоку D_{pp}. На рис. 1. показано блочний розподіл даних на *к*-му кроці факторизації блочно-діагональної матриці з обрамленням, враховуючи запропоновану вище декомпозицію.

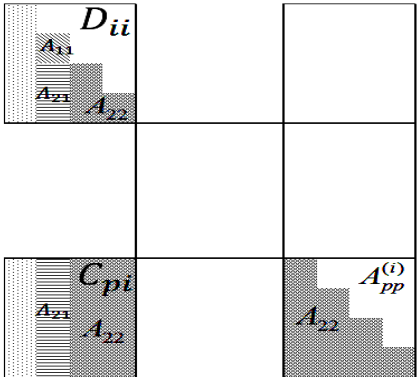

Рис. 1. *Декомпозиція даних в CPU на к-му кроці факторизації*

Враховуючи декомпозицію даних, приведену вище, плитковий алгоритм факторизації записується у наступній формі:

- над всіма діагональними блоками крім D_{pp} та відповідними блоками обрамлення послідовно виконуються такі операції:
  - на CPU факторизуємо A_{11} $A_{11} = L_{11} * L_{11}^T$;
  - паралельно в різних потоках модифікуємо стовпчик блоків L_{21} $L_{21} = A_{21}\left(L_{11}^T\right)^{-1}$;
  - незалежно в кількох потоках модифікуємо блоки матриці A_{22} за формулою:
$$\tilde{A}_{22} = A_{22} - L_{21}L_{21}^T;$$
- використовуючи операцію мультизбирання модифікуємо блок D_{pp} $\tilde{D}_{pp} = D_{pp} - \sum_{i=1}^{p-1} A_{pp}^{(i)}$;
- факторизуємо блок $\tilde{D}_{pp}$, тим самим завершуючи процес факторизації матриці A.

Оцінка прискорення. Для оцінки якості паралельних алгоритмів будемо використовувати коефіцієнт прискорення S_p, що обчислюється за формулою

$$S_p = T_1 / T_p,$$

де T_1 — час розв'язування задачі на комп'ютері з одним CPU, T_p — час розв'язування тієї ж задачі на комп'ютері з багатоядерним CPU.

Будемо вважати, що порядки всіх діагональних блоків приблизно рівні

$$q_i \approx q = \frac{n-s}{p-1},$$

де s — порядок останнього діагонального блоку.

Оскільки (4)–(6) виконуються паралельно і незалежно у всіх $p-1$ CPU і максимальна кількість операцій припадає на етап (6), то оцінка кількості операцій виконуваних на одному CPU визначається саме складністю етапу (6).

Кількість операцій необхідних для виконання (6) можна оцінити величиною

$$N_p \approx \frac{q^3}{3} + sq^2 = \frac{q^2}{3}(q+3s).$$

Обчислимо значення T_1 і T_p, використовуючи значення N_p знайдене вище.

$$T_1 \approx (p-1) N_p t \ , \ T_p \approx N_p t \ + \frac{(p-1)s^2}{2} t_{opp},$$

де t — час виконання однієї арифметичної операції, t_{opp} — час обміну між двома процесами.

Коефіцієнт прискорення для алгоритму оцінюється величиною

$$S_p \approx (p-1)\left(1 + \frac{3}{q^2(q+3s)}\left(\frac{(p-1)s^2}{2}\tau_{opp}\right)\right)^{-1}, \qquad (7)$$

де $\tau_{opp} = \dfrac{t_{opp}}{t}$.

Програмна реалізація та чисельні експерименти. При програмній реалізації алгоритму на комп'ютерах гібридної архітектури доцільно використовувати функції оптимізованих програмних бібліотек. Зокрема, до таких бібліотек відносяться Intel MKL [8]. Реалізація обмінів між процесорами відбувається за допомогою функцій MPI [9].

Розглянемо функції, що використовувались при програмній реалізації алгоритму:

- MPI_Reduce — функція глобальної редукції зі збереженням результату у вказаному процесорі;
- LAPACKE_dlange — повертає значення 1-норми, норми Фробеніуса, норми нескінченності або найбільшого абсолютного значення будь-якого елемента прямокутної матриці;
- dpotrf — знаходження LL^{T} розвинення щільної матриці;
- cblas_strsm, cblas_dtrsm — розв'язання трикутної системи з багатьма правими частинами на одинарній та подвійній точності, відповідно;
- cblas_sgemm, cblas_dgemm — знаходження добутку двох матриць на одинарній та подвійній точності, відповідно.

Розрядність на якій проводяться обчислення визначається наступним чином:

- вибір розрядності для обчислення (4) базується на числі обумовленості матриці відповідного блоку;
- у (5), (6) розрядність вибирається на основі евклідових норм відповідних блоків.

Обчислення проводились на вузлах суперкомп'ютера СКІТ [10], що мають наступні характеристики: 2 чотирьох ядерних Intel Xeon 5345 з тактовою частотою 2,2 ГГЦ, 16 ГБ оперативної пам'яті.

Для проведення чисельних експериментів вибрано ряд тестових матриць з колекції розріджених матриць університету Флориди [11]. Характеристики матриць приводяться в таблиці.

Таблиця

Набір тестових матриць

№ п./п.	Назва	Проблемна область	Порядок	Кількість ненульових елементів
1.	ecology2	2D/3D problem	999999	4,995,991
2.	apache2	structural problem	715 176	4 817 870
3.	thermomech_dM	thermal problem	204,316	1,423,116
4.	G2_circuit	circuit simulation problem	150 102	726 624
5.	Dubcova3	2D/3D problem	146,689	3,636,643
6.	cvxbqp1	optimization problem	50 000	349 968
7.	minsurfo	optimization problem	40 806	203 622

На рис. 2. показано залежність продуктивності від розміру плитки та при використанні різної розрядності. Результати приведені для матриці apache2 при використанні 6 потоків.

Рис. 2. *Залежність продуктивності від розрядності і розміру плитки*

Висновки. В статті розглянуто проблеми, що виникають при обчисленнях на довільній розрядності. Запропоновано алгоритм факторизації розрідженої матриці, який у ході розвинення матриці проводить обчислення на мішаній розрядності. Запропоновано критерії вибору розрядності, які в даному алгоритмі дозволяють зберігати точність обчислень. Використання запропонованого підходу при розв'язанні практичних задач показало значний приріст продуктивності, в порівнянні з обчисленнями на подвійній точності. Також на практиці використання мішаної розрядності в обчисленнях, дозволило значно економити пам'ять системи.

Список використаних джерел:

1. Nikolaevskaya E., Khimich A., Chystyakova T. Programming with Multiple Precision. Springer-Verlag Berlin Heidelberg. 2012. 234 p.
2. URL: https://www.mpfr.org.
3. URL: https://gmplib.org.
4. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. М. : Мир, 1984. 334 с.
5. Химич А. Н., Попов А. В., Полянко В. В. Алгоритмы параллельных вычислений для задач линейной алгебры с матрицами нерегулярной структуры. *Кібернетика і системний аналіз*. 2011. Вип. 47, № 6. С. 159–174.
6. Хіміч О. М., Сидорук В. А. Гібридний алгоритм розв'язування лінійних систем з розрідженими матрицями на основі блочного LL^T методу. *Комп'ютерна математика*. 2015. Вип. 1. С. 67–74.
7. Buttari A., Langou J., Kurzak J., Dongarra J. A Class of Parallel Tiled Linear Algebra Algorithms for Multicore Architectures. *Parallel Computing.* 2009. Vol. 35, Issue 1. P. 38–53.
8. Intel® Math Kernel Library (Intel® MKL). URL: https://software.intel.com/-en-us/intel-mkl.
9. Gropp W., Lusk E. and Thakur R. Using MPI-2: Advanced Features of the Message-Passing Interface. Cambridge : MIT Press. 1999. 382 p.

10. URL: http://icybcluster.org.ua.
11. URL: https://sparse.tamu.edu.

USE OF MIXED PRECISION IN MATHEMATICAL MODELING

The work proposes a method by which it is possible to accelerate the time of mathematical modeling of complex systems, using a mixed precision in calculations. The mixed precision allows you to improve the computing performance and save memory. Showing this approach in constructing an algorithm for solving systems of linear algebraic equations with sparse matrices.

Key words: *mathematical modeling, mixed precision, parallel algorithms, sparse matrices.*

Одержано 15.02.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.187-192

О. В. Чистяков, канд. фіз.-мат. наук

Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ГІБРИДНИЙ ІТЕРАЦІЙНИЙ АЛГОРИТМ ДЛЯ РОЗВ'ЯЗУВАННЯ ЧАСТКОВОЇ ПРОБЛЕМИ ВЛАСНИХ ЗНАЧЕНЬ

Розглядається паралельний алгоритм поперемінно-трикутного методу для розв'язування на багатоядерному комп'ютері з графічними прискорювачами часткової узагальненої алгебраїчної проблеми власних значень розріджених симетричних матриць на основі їх зведення до блочно-діагонального виду з обрамленням.

Ключові слова: *гібридний комп'ютер, поперемінно-трикутний метод, розріджена матриця, алгебраїчна проблема власних значень.*

Вступ. Багато прикладних задач зводяться до розв'язування часткової узагальненої алгебраїчної проблеми власних значень (АПВЗ) для розріджених матриць великих розмірів. Наприклад, такі задачі виникають в електричних та механічних системах, в яких власні значення відповідають власним частотам коливань, а власні вектори характеризують відповідні форми (моди) коливань [1]. Визначення власних значень та векторів дають можливість аналізувати процеси та управляти ними.

З метою підвищення ефективності розв'язування задач на власні значення розріджених матриць великих розмірів у гібридному алгоритмі поперемінно-трикутного методу, який пропонується, використано ідею попереднього зведення вихідної розрідженої матриці A задачі виду

$Ax = \lambda Bx$ за допомогою методу паралельних перерізів до блочно-діагональної матриці з обрамленням [1]. Таке представлення розрідженої матриці дає можливість більш ефективно виконувати розпаралелення обчислень на багатоядерному комп'ютері з графічними процесами (гібридному комп'ютері) як на CPU, так і на GPU. Слід також зазначити, що виконання ітерацій у методі, що розглядається, здійснюється послідовно, тому доцільно розглядати математичні операції, які можуть бути розпаралелені на кожній ітерації. Причому, найбільш трудомісткими математичними операціями та підзадачами (логічно завершенні частини алгоритму) на кожній ітерації щодо витрат обчислювальних ресурсів і часу виконання є множення розрідженої матриці на вектор та розв'язування систем лінійних алгебраїчних рівнянь з трикутними матрицями. Тому саме ці підзадачі підлягають розпаралелюванню на графічних процесорах, що значно прискорює процес обчислень.

Постановка задачі. Розглядаємо розв'язування задачі на власні значення

$$Ax = \lambda Bx, \qquad (1)$$

де A — розріджена додатно визначена матриця порядку n, B — діагональна матриця з додатними елементами, що діють в n-вимірному евклідовому просторі H із скалярним добутком $(\ ,\ )$, λ та x — відповідно власне значення та власний вектор.

Застосуємо до задачі (1) наступну канонічну ітераційну однокрокову схему знаходження λ_1 та x_1 [2]:

$$Z(y_{k+1} - y_k) + \tau_{k+1} r_k = 0, \text{ для } k = 0, 1, 2, 3, \ldots,$$

де y_0 — довільне початкове значення, $r_k = Ay_k - \mu_k By_k$ — нев'язка; $\mu_k = (Ay_k, y_k)(\mathrm{B}\, y_k, y_k)^{-1}$ — наближення до власного значення; y_k — нормоване наближення до власного вектора; $\tau_к$ — ітераційний параметр, що обчислюється за формулою $\tau_{k+1} = (w_k, r_k) / (Aw_k, w_k)$; Z — матриця (регуляризатор), яка впливає на швидкість збіжності ітераційного процесу.

В залежності від вибраної матриці Z та наборів параметрів τ та α можуть бути отримані різноманітні схеми ітераційних методів, наприклад, для попереміно-трикутного методу:

$$w_k = B^{-1} r_k, \ r_k = Ay_k - \mu_k y_k.$$

Матриця Z має вигляд

$$Z = (E + \omega\tilde{L})(E + \omega\tilde{L}^T).$$

Матриця A представляється у вигляді $A = \tilde{L} + \tilde{L}^T$, де $\tilde{L}$ та $\tilde{L}^T$ — відповідно нижня та верхня трикутні матриці, сформовані з умови, що діагональні елементи $l_{ii} = \frac{1}{2} a_{ii}$ при розв'язуванні задачі

$$(E + \omega_{(k)}\tilde{L})(E + \omega_{(k)}\tilde{L}^{T})\, w_{(k)} = r^{(k)}.$$

Зведення матриці до блочно-діагонального вигляду з обрамленням. Застосуємо метод паралельних перерізів для зведення матриці A задачі (1) до блочно-діагонального вигляду з обрамленням [3]:

$$\tilde{A} = P^{T}AP = \begin{pmatrix} D_1 & 0 & 0 & \dots & 0 & C_1 \\ 0 & D_2 & 0 & \dots & 0 & C_2 \\ 0 & 0 & D_3 & & 0 & C_3 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & & D_{p-1} & C_{p-1} \\ C_1 & C_2 & C_3 & \dots & C_{p-1} & D_p \end{pmatrix}, \qquad (2)$$

де P — матриця перестановок, а блоки D_i і C_i зберігають розрідженість.

Таким чином, вихідна задача (1) зводиться до $\tilde{A}y = \lambda\tilde{B}y$, де $\tilde{A}$ — блочно-діагональна матриця з обрамленням, $\tilde{B} = P^{T}BP$ — розріджена додатно визначена матриця.

Оскільки тут використано перетворення подібності, то власні значення отриманої матриці $\tilde{A}$ і вихідної A співпадають.

Передобумовлювач (матриця Z) для попереміннo-трикутного методу має вигляд

$$Z = (E + \omega\tilde{L})(E + \omega\tilde{L}^{T}),\ A = \tilde{L} + \tilde{L}^{T}. \qquad (3)$$

При цьому матриця $\tilde{L}$ зберігає блочно-трикутну структуру:

$$\tilde{L} = \begin{pmatrix} \tilde{L}_1 \\ \tilde{L}_2 \\ \tilde{L}_3 \\ \vdots \\ \tilde{L}_{p-1} \\ \tilde{L}p \end{pmatrix} = \begin{pmatrix} \tilde{D}_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \tilde{D}_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \tilde{D}_3 & \dots & 0 & 0 \\ \vdots & \vdots & & \ddots & & \dots \\ 0 & 0 & 0 & & \tilde{D}_{p-1} & 0 \\ C_1 & C_2 & C_3 & \dots & C_{p-1} & \tilde{D}_p \end{pmatrix}, \qquad (4)$$

де блоки $\tilde{D}_i$ — нижні трикутні матриці D_i, C_i — прямокутні матриці, $1 \le i \le p$.

Розглядається гібридний комп'ютер, архітектура якого складається з багатоядерного комп'ютера MIMD-архітектури з топологією комунікацій між процесорами — «гіперкуб» та декількох графічних процесорів (GPU) SIMT-архітектури (Single Instruction, Multiple Threads). Тобто задача розв'язується на гібридному комп'ютері при використанні p процесів на CPU, позначимо $p\,\mathrm{CPU}$, та p відповідних до них графічних процесорів — $p\,\mathrm{GPU}$. Під $\mathrm{CPU}(q)$ будемо ро-

зуміти паралельний процес, що виконується на ядрі CPU з логічним номером q ($q = 1, 2..., p$) при використанні графічного процесора з таким же номером — GPU(q).

Для розподілу блоків (підматриць) $\tilde{L}$ між процесами CPU гібридного комп'ютера використовуємо блочну схему. З огляду на структуру $\tilde{L}$ це означає, що процеси з номерами $1 \le i < p$ зберігають блоки $\tilde{D}_i$ та C_i, а процес з номером p зберігає блок $\tilde{D}_p$.

Попереміннo-трикутний метод належить до класу методів, ітераційні параметри до яких вибираються з урахуванням апріорної інформації про оператори ітераційної схеми.

Для цього методу такою інформацією є величини δ та Δ. За умов $\|\hat{x}\|_{\hat{A}}^2 \ge \delta \|\hat{x}\|^2$, $\|\hat{L}^{\mathrm{T}} \hat{x}\| \le \Delta \|\hat{x}\|_{\hat{L}}^2 / 4$ параметр $\omega = \dfrac{2}{\sqrt{\delta\Delta}}$ забезпечує збіжність ітераційного процесу зі швидкістю геометричної прогресії [4].

Реалізація гібридного алгоритму попереміннo-трикутного методу. Гібридна реалізація попереміннo-трикутного методу на багатоядерних комп'ютерах з графічними прискорювачами визначається блочно-трикутною структурою матриць $\tilde{L}_i$ та $\tilde{L}_i^{\,T}$ при розв'язуванні системи $Zw = r$, де матриця Z визначається за формулою (3). Отже, розв'язування задачі (3) зводиться до розв'язування на гібридній комп'ютерній архітектурі, з використанням p CPU та відповідних p GPU, двох систем:

$$(E + \omega\tilde{L})\, y = r \quad \text{та} \quad (E + \omega\tilde{L}^T)\, w = y$$

за наступною обчислювальною схемою.

Гібридний алгоритм розв'язування системи з нижньою трикутною матрицею

$$(E + \omega\tilde{L})\, y = r \tag{5}$$

зводиться до виконання таких кроків:

- одночасно і незалежно процеси CPU(q), виконуючи на GPU(q) багатопоточні матрично-векторні операції, розв'язують трикутні системи $(E + \omega\tilde{D}_q) y_q = r_q$, де $1 \le q < p$, та обчислюють вектори $\tilde{y}_q$ за формулою $\tilde{y}_q = C_q y_q$;
- обчислені вектори $\tilde{y}_q$ одночасно пересилаються від процесів з номерами $1 \le q < p$ в процес CPU(p), в якому на GPU(p) розв'язується система $(E + \omega\tilde{D}_p) y_p = r_p - \sum_{q=1}^{p-1} \tilde{y}_q$.

Гібридний алгоритм розв'язування системи з верхньою трикутною матрицею

$$(E + \omega \tilde{L}^T)\, w = y \qquad (6)$$

зводиться до виконання таких кроків:

- процес CPU(p), використовуючи GPU(p), розв'язує систему $(E + \omega \tilde{D}_p^T) w_p = y_p$ та розсилає отриманий вектор y_p процесам CPU(q), $1 \le q < p$;
- процеси CPU(q), використовуючи GPU(q), одночасно та незалежно розв'язують трикутні системи $(E + \omega\, \tilde{D}_q^T)\, w_q = y_q - C_q^T w_p$;
- обчислені вектори y_q пересилаються від процесів CPU(q) $(1 \le q < p)$ в процес CPU(p), в якому, на GPU(p), розв'язується система $(E + \omega\, \tilde{D}_p)\, y_p = r_p - \sum_{q=1}^{p-1} \tilde{y}_q$.

Експериментальне дослідження гібридного алгоритму проводилось на інтелектуальній робочій станції гібридної архітектури Інпарком_g з такими технічними характеристиками: CPU — серії Intel(R) Xeon(R) E5606; тактова частота 2.13 GHz; швидкість 4,8 GT/s; кеш-пам'ять 8 Mb; у вузлі — 2 CPU по 4 ядра, Max Memory Size 288 Gb; графічні процесори — Nvidia Tesla M2090; пам'ять 6 Gb.

На графіках (рисунок) показано прискорення розробленого алгоритму при розв'язуванні на різній гібридній архітектурі (CPU + GPU) часткової АПВЗ для розріджених (стрічкових) матриць порядку 250 000 з різною напівшириною стрічки, які отримано шляхом дискретизації методом скінченних елементів змішаної крайової задачі для оператора Лапласа в прямокутному паралелепіпеді.

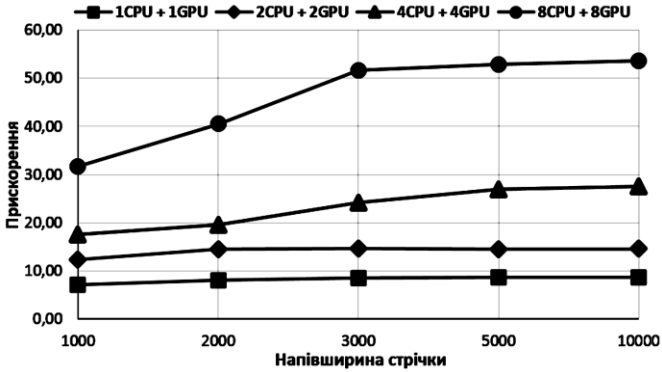

Рисунок. *Прискорення гібридного алгоритму*

З рисунку видно, що створений гібридний алгоритм добре масштабований, прискорення значно зростає при збільшенні кількості CPU та GPU в гібридній архітектурі комп'ютера у порівнянні з прискоренням на архітектурі $1CPU + 1GPU$ для всіх задач.

Найбільше прискорення отримується для задач з напівшириною стрічки 3000 і більше. Це пояснюється зростанням заповненості GPU-процесорів при виконанні матрично-векторних операцій. Як відомо, GPU здатний миттєво обробляти до 1024 потоків, кожному потоку ставиться у відповідність один елемент матриці або компонента вектора.

Отже, найефективніше можна використати продуктивність графічних процесорів, якщо одна і та ж послідовність математичних операцій виконується над великим обсягом даних та кожен GPU буде максимально заповнений.

Висновки. Розглянуто гібридний алгоритм поперемінно-трикутного методу на основі методу паралельних перерізів з передобумовлювачем для розв'язування часткової узагальненої АПВЗ розріджених додатно визначених матриць. Дослідження і апробація створеного гібридного алгоритму показали, що застосування передобумовлювача зводить задачу до розв'язування лінійних систем з блочно-діагональними трикутними матрицями, що визначає високу ступінь паралелізму та збалансованості паралельного процесу. Ефективність алгоритму можна значно покращити за рахунок багатопоточного виконання матрично-векторних операцій з великими обсягами даних на графічних процесорах синхронно з копіюванням даних між CPU та GPU.

Список використаних джерел:

1. Писанецки С. Технология разреженных матриц. М. : Мир, 1988, 410 с.
2. Приказчиков В. Г., Химич А. Н. Итерационные методы решения задач устойчивости и колебания пластин и оболочек. *Прикладная механика.* 1984. Т. 20, № 1. С. 88–94.
3. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. М. : Мир, 1984. 334 с.
4. Самарский А. А., Николаев Е. С. Методы решения сеточных уравнений. М. : Наука, 1978. 592 с.

HYBRID ITERATIVE ALGORITHM FOR SOLVING PARTIAL EIGENVALUE PROBLEM

A parallel alternating-triangular method is considered for solving a partial generalized algebraic problem of eigenvalues of sparse symmetric matrices on a multi-core computer with graphic accelerators based on their reduction to block-diagonal form with a frame.

Key words: *hybrid computer, alternating-triangular method, rarefied matrix, algebraic problem of eigenvalues.*

Одержано 17.02.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.193-198

Т. В. Чистякова, канд. фіз.-мат. наук,
П. С. Єршов, аспірант
Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ПРО ВИБІР РОЗРЯДНОСТІ ОБЧИСЛЕНЬ В ІНТЕЛЕКТУАЛЬНІЙ СИСТЕМІ ОБРОБКИ МАТРИЦЬ

Описується модель прийняття рішень щодо вибору необхідного алгоритму та розрядності обчислень в інтелектуальній системі обробки матриць для достовірного розв'язування систем лінійних алгебраїчних рівнянь на багатоядерному комп'ютері з графічними процесорами.

Ключові слова: *гібридний комп'ютер, інтелектуальна система, система лінійних алгебраїчних рівнянь, підвищена розрядність.*

Вступ. Багато науково-технічних задач зводяться до розв'язування систем лінійних алгебраїчних рівнянь (СЛАР) великої розмірності з наближеними даними та невизначеними математичними властивостями. Для ефективного розв'язування таких великих задач необхідно використовувати сучасні потужні комп'ютери, наприклад, багатоядерні комп'ютери з графічними прискорювачами (гібридні комп'ютери). Вихідні дані математичних моделей більшості прикладних задач задаються наближено, тому в результаті заокруглень вхідних комп'ютерних даних задач відповідні комп'ютерні моделі можуть мати зовсім інші математичні властивості і результати розв'язування. Крім того, комп'ютерні методи розрахунку математичних моделей у свою чергу мають також наближений характер і вносять додаткові похибки в отримані машинні результати. Таким чином, проблема достовірності містить два природних аспекти: достовірність математичної моделі, яка описує прикладну задачу, і достовірність комп'ютерного розв'язку математичної моделі.

Похибку у розв'язку математичної задачі, яка обумовлена похибкою в заданні вихідних даних, називають спадковою похибкою. Якщо спадкова похибка розв'язку математичної задачі велика, то отриманий математичний розв'язок може не мати фізичного змісту, тобто такий розв'язок не буде містити в собі розв'язок фізичної задачі. Тому не всяка математична модель (математична задача з наближено заданими вихідними даними) буде містити розв'язок, що має фізичний зміст. Таким чином, виникає проблема визначення області дії математичної моделі, адже спадкову похибку не можна виправити математичними

методами розв'язування задачі. Для зменшення спадкової похибки необхідно або підвищити точність задання вихідних даних або переформулювати задачу щодо інших параметрів. Обчислювальна похибка для прямих методів виникає внаслідок заокруглень у ході реалізації алгоритму розв'язування математичної задачі в комп'ютері. Одним із способів мінімізації помилок, пов'язаних із заокругленням у комп'ютері даних та результатів обчислень, а також втратою точності через обмеженість представлення чисел в сучасних комп'ютерах, є подальше збільшення розрядності комп'ютерних обчислень [1].

Таким чином, необхідні нові підходи до створення програмного забезпечення на гібридні комп'ютери, які б забезпечували достовірним комп'ютерним розв'язком задачі при ефективному використанні обчислювальних ресурсів гібридних комп'ютерів.

Функціональні можливості інтелектуальної системи для обробки матриць. Пропонується інтелектуальна система для достовірного розв'язування СЛАР на багатоядерному комп'ютері з графічними процесорами, яка здатна самостійно досліджувати математичні властивості комп'ютерних моделей задач та приймати рішення щодо ефективного їх розв'язування. При створенні такої інтелектуальної системи передбачена обробка трьох її складових (дані про задачу, ознаки та властивості, алгоритми).

Дані вводяться в комп'ютер за допомогою інтелектуального інтерфейсу: вид матриці, порядок матриці, кількість правих частин тощо. Передбачено розв'язування задач з матрицями різного виду (щільними, стрічковими, розрідженими довільної структури).

Ознаки та властивості. До них відносяться математичні властивості комп'ютерної моделі задачі, які визначено в результаті дослідження за розробленою технологією.

Алгоритми. Передбачено набір гібридних алгоритмів ефективних методів розв'язування СЛАР з різними матрицями — додатно визначеними, невиродженими, погано обумовленими, виродженими.

На основі проведених досліджень вхідних даних та аналізу результатів досліджень інтелектуальна система приймає рішення щодо вирішення таких питань:

- визначення оптимальної кількості процесорів та побудова ефективної топології міжпроцесорних зв'язків гібридного комп'ютера;
- дворівневе розпаралелення обчислень на необхідних обчислювальних ресурсах центрального процесора та графічних процесорів;
- автоматичне визначення необхідної розрядності обчислень;
- аналіз достовірності комп'ютерних результатів.

Технологія дослідження та розв'язування СЛАР. Розглянемо технологію розв'язування лінійних систем з наближеними даними,

яку реалізовано інтелектуальній системі, на прикладі розв'язування СЛАР із щільними несиметричними квадратними матрицями.

Дослідження та розв'язування СЛАР виду $Ax = b$ розпочинається з обчислення оцінки числа обумовленості матриці на подвійній розрядності в ході реалізації гібридного алгоритму LU-розвинення матриці за схемою [2, 3]:

$$A \cong LU \quad \|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|,\ Uw = e,\ L^T y - w,\ Lv = y,\ Uz = v,$$

$$\mathrm{cond}A = \|A\|_1 \|z\|_1 / \|y\|_1,\ \|z\|_1 = \sum_{i=1}^{n} |z_i|,\ \|y\|_1 = \sum_{i=1}^{n} |y_i|.$$

Якщо матриця системи виявилася в комп'ютері невиродженою, то розв'язування задачі продовжується на подвійній розрядності з оцінками достовірності результатів: оцінка близькості машинного розв'язку до математичного та оцінка спадкової похибки.

Верхня межа відносної спадкової похибки обчислюється за формулою:

$$E_{спадкова} = \frac{\|x - \tilde{x}\|}{\|\tilde{x}\|} \le \mathrm{cond}A \times \frac{\varepsilon_A + \varepsilon_b}{1 - \varepsilon_b},$$

де $\tilde{x}$ — точний розв'язок системи з точно заданими вихідними даними; x — точний розв'язок системи з наближено заданими вихідними даними; ε_A, ε_b — максимальні відносні похибки елементів матриці та правої частини СЛАР відповідно.

Характеристикою близькості машинного розв'язку до математичного є оцінка обчислювальної похибки розв'язку. Для її обчислення використовується один крок процедури ітераційного уточнення розв'язку.

Ітераційне уточнення розв'язку системи $Ax = b$ реалізується за схемою [2]:

$$x^{(0)} = x,$$
$$r^{(s)} = b - Ax^{(s)},$$
$$A \times \Delta x^s + r^{(s)},$$
$$Ax^{(s+1)} = x^{(s)} + \Delta x^{(s)},\ s = 0, 1, 2, \ldots$$

Обчислення нев'язки $r_i^{(s)}$ виконується на підвищеній розрядності (на регістрах процесора, де розрядність на декілька знаків більша у порівнянні з основною подвійною розрядністю).

Оцінка обчислювальної похибки розв'язку визначається за формулою:

$$E_{\text{обчис.}} < \frac{\|\Delta x_1\|}{\|x_2\|},$$

де x_1 — обчислений розв'язок системи, x_2 — наближення до точного розв'язку, яке отримано за один крок ітераційного уточнення.

Якщо значення cond$A$ задовольняє в комп'ютері умові:

$$1.0 + 1.0/\mathrm{cond}A = 1.0\,, \qquad (1)$$

то матриця вважається виродженою у межах подвійної точності. В цьому випадку розв'язування автоматично продовжується з використанням розрядності обчислень 128. По закінченню обчислювального процесу видається отриманий вектор розв'язку системи.

Якщо матриця СЛАР, яка введена в комп'ютер, не кваліфікується за (1) як вироджена, але

$$\varepsilon_A \times \mathrm{cond}A \geq 1\,,$$

(ε_A — максимальна відносна похибка елементів матриці), то така матриця СЛАР в комп'ютері виявляється виродженою у межах подвійної точності задання її елементів, тому не можна гарантувати достовірність обчисленого розв'язку. Оскільки навіть при введені в комп'ютер точно заданих елементів СЛАР здійснюється їх заокруглення, тому ε_A присвоюється значення *macheps* — найменшого числа з плаваючою комою, для якого в комп'ютері виконується умова [4]:

$$1 + macheps > 1\,.$$

В цьому випадку розв'язування автоматично продовжується з використанням підвищеної розрядності обчислень, наприклад 128. Оцінка обчислювальної похибки розв'язку системи отримується за формулою:

$$E_{обчис.} = \mathrm{cond}A \times macheps\,.$$

Демонстраційна задача. Дослідити та розв'язати в інтелектуальній системі СЛАУ виду $Ax = b$, де

$$A = (a_{ij})\,,\ i, j = 1 \div n\,,\ n = 3w + 1\,,\ w = 1, 2, \ldots$$

$$a_{ii} = n - i\,,\ a_{ij} = n + 1 - \max(i, j)\,.$$

Тобто щільна симетрична матриця A має вигляд:

$$A = \begin{pmatrix} n-1 & n-1 & n-2 & \ldots & 2 & 1 \\ n-1 & n-2 & n-2 & \ldots & 2 & 1 \\ n-2 & n-2 & n-3 & \ldots & 2 & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 2 & 2 & 2 & \ldots & 1 & 1 \\ 1 & 1 & 1 & \ldots & 1 & 0 \end{pmatrix}.$$

Елементи правої частини системи обчислюються за формулами:

$$b = \{b\}_1^n\,,\ b_i = n - i,\ \text{якщо } i \leq 2;$$

$$b_i = n + 1 - i,\ \text{якщо } i > 2.$$

Точний розв'язок системи:

$x = (0 \quad 1 \quad 0 \quad ... \quad 0)^T$, порядок системи: $n = 1000$.

Лістинг протоколу дослідження та розв'язування задачі

```
    PROBLEM:
  solving of the linear algebraic system
  with a symmetic positive defined matrix
Data:
   - number of matrix's rows             = 1000
   - number of matrix's columns          = 1000
   - number of the right-hand side
     of the systems                      = 1
   - maximum relative error
     of the matrix elements              =
   0.00000e+00
   - maximum relative error
     of elements of the right-hand sides =
0.00000e+00
   Process of investigating and solving
   Double precision
   METHOD:
      Choletsky decomposition
   RESULTS:
   !!! THE MATRIX IS NOT POSITIVE DEFINED !!!
   METHOD:
       Gauss elimination with partial pivoting
   RESULTS:
   !!! THE MATRIX IS MACHINE-SINGULAR !!!
   PRECISION: 128
   SOLUTION
   first 4 components of solution are:
      0             1             0             0
Computational error in the solution: 0.00000e+00
The vecror of solution are successfully stored in
the file result.out
```

Висновки. Розглянуто інтелектуальну систему для дослідження та розв'язування систем лінійних алгебраїчних рівнянь з використанням підвищеної розрядності обчислень та аналізом достовірності комп'ютерних результатів. Забезпечується не тільки гарантія досто-

вірності результатів, але також звільнення користувачів від проблем розпаралелення обчислень на складній гібридній архітектурі.

Список використаних джерел:

1. Nikolaevskaja E. A., Khimich A. N, Chistyakova T. V. Programming with Multiple Precision. Springer-Verlag. Studies in Computational Intelligence. Berlin, Heidelberg. 2012. Vol. 397 233 p.
2. Молчанов И. Н. Машинные методы решения прикладных задач. Алгебра, приближение функций. Киев : Наук. думка, 1987. 285 с.
3. Химич А. Н., Молчанов И. Н., Попов А. В. и др. Параллельные алгоритмы решения задач вычислительной математики. Киев : Наук. думка, 2008. 248 с.
4. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. М. : Мир, 1980. 275 с.

ABOUT CHOICE OF DETERMINATION OF THE CALCULATIONS IN THE INTELLECTUAL MATRIX PROCESSING SYSTEM

The decision-making model for choosing the required algorithm and the digit capacity of calculations in a intelligente system for the reliable solution of linear algebraic equations systems on a multi-core computer with graphic processors is described.

Key words: *hybrid computer, intelligente system, a system of linear algebraic equations, increased bit depth.*

Одержано 17.02.2019

ВІДОМОСТІ ПРО АВТОРІВ

Барболіна Тетяна Миколаївна — кандидат фізико-математичних наук, доцент, завідувач кафедри, Полтавський національний педагогічний університет імені В. Г. Короленка, м. Полтава, tm-b@ukr.net

Бардадим Тамара Олексіївна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, tamara.bardadym@gmail.com

Варенюк Наталія Анатоліївна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, nvareniuk@ukr.net

Галба Євген Федорович — доктор фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, e.f.galba@ukr.net

Дараган Катерина Володимирівна — аспірантка, Українська інженерно-педагогічна академія, м. Харків, keitakaterina@gmail.com

Emmenegger Jean-Francois — Doctor, Lecturer, University of Fribourg, Switzerland, Jean-Francois.Emmenegger@unifr.ch

Єршов Павло Сергійович — аспірант, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, jershov.pavel.WSR@gmail.com

Задірака Валерій Костянтинович — академік НАН України, доктор фізико-математичних наук, професор, завідуючий відділом, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, zvk140@ukr.net

Кандій Сергій Олегович — студент, Харківський національний університет імені В. Н. Каразіна, м. Харків, kandy.sergey@yandex.ua

Каргапольцева Ганна Вікторівна — пошукувач, Українська інженерно-педагогічна академія, м. Харків, kargapoltseva@ukr.net

Качко Олена Григорівна — заступник генерального конструктору, АТ «Інститут інформаційних технологій», м. Харків, elena_kachko@nure.ua

Knolle Helmut — Dr. math., formerly at the Institute for Social and Preventive Medicine, University of Bern, Switzerland, Prof. at the National University of Colombia, helmut.knolle@bluewin.ch

Когутич Оксана Іванівна — магістрант, Ужгородський національний університет, м. Ужгород, oksanakogutych97@gmail.com

Козін Ігор Вікторович — доктор фізико-математичних наук, професор, Запорізький національний університет, м. Запоріжжя, ainc00@gmail.com

Коломис Олена Миколаївна — кандидат фізико-математичних наук, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, kolomys@ukr.net

Крот Александр Михайлович — доктор технических наук, профессор, заведующий лабораторией, Объединенный институт проблем информатики НАН Беларуси, г. Минск, Республика Беларусь, alxkrot@newman.bas-net.by

Лаптін Юрій Петрович — доктор фізико-математичних наук, завідувач відділом, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, yu.p.laptin@gmail.com

Литвин Олег Миколайович — доктор фізико-математичних наук, професор, професор кафедри, Українська інженерно-педагогічна академія, м. Харків, academ_mail@ukr.net

Литвин Олег Олегович — доктор фізико-математичних наук, доцент, декан, Українська інженерно-педагогічна академія, м. Харків, olegolitvin55@gmail.com

Луц Лілія Володимирівна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, lv1@ukr.net

Макаренко Олександр Сергійович — доктор фізико-математичних наук, завідувач відділом, Інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, makalex51@gmail.com

Маринець Василь Васильович — доктор фізико-математичних наук, професор, завідувач кафедри, Ужгородський національний університет, м. Ужгород, vasyl.marynets@uzhnu.edu.ua

Недашковський Микола Олександрович — доктор фізико-математичних наук, професор, Університет Казимира Великого, м. Бидгощ, Республіка Польща, m.nedashkovskyy@gmail.com

Нестеренко Алла Никифорівна — молодший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, alla.nest1958@gmail.com

Нечуйвітер Олеся Петрівна — доктор фізико-математичних наук, доцент, завідувач кафедри, Українська інженерно-педагогічна академія, м. Харків, olesya@email.com

Nour Eldin Hassan Ahmed — Prof. Dr., formerly Prof. of Automatic Control and Technical Cybernetics, University of Wuppertal, Germany, eldin@uni-wuppertal.de

Острянська Єлізавета Вадимівна — студент, Харківський національний університет імені В. Н. Каразіна, м. Харків, antelizza@gmail.com

Пасічник Валентина Олексіївна — кандидат технічних наук, доцент, Харківська державна академія дизайну і мистецтв, м. Харків, pasechnik.va@gmail.com

Першина Юлія Ігорівна — доктор фізико-математичних наук, доцент, професор кафедри, Українська інженерно-педагогічна академія, м. Харків, yuliapershina78@gmail.com

Петренюк Володимир Ілліч — кандидат фізико-математичних наук, доцент, доцент кафедри, Центральноукраїнський національний технічний університет, м. Кропивницький, petrenjukvi@i.ua

Покутний Олександр Олексійович — доктор фізико-математичних наук, старший науковий співробітник, Інститут математики НАН України, м. Київ, lenasas@gmail.com

Поліщук Олександр Дмитрович — кандидат фізико-математичних наук, старший науковий співробітник, Інститут прикладних проблем механіки і математики імені Я. С. Підстригача НАН України, м. Львів, od_polishchuk@ukr.net

Полюга Світлана Ігорівна — кандидат фізико-математичних наук, старший викладач, Запорізький національний університет, м. Запоріжжя, veta9932@gmail.com

Попов Олександр Володимирович — кандидат фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, alex50popov@gmail.com

Пустовіт Олександр Сергійович — молодший науковий співробітник, Інститут телекомунікацій і глобального інформаційного простору НАН України, м. Київ, sanyk_set@ukr.net

Рудич Ольга Василівна — науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, olgar2006@ukr.net

Савкіна Марта Юріївна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут математики НАН України, м. Київ, marta@imath.kiev.ua

Сардак Вікторія Ігорівна — аспірант, Запорізький національний університет, м. Запоріжжя, vsardak85@gmail.com

Семенов Володимир Вікторович — доктор фізико-математичних наук, професор, Київський національний університет імені Тараса Шевченка, м. Київ, semenov.volodya@gmail.com

Семенов Василь Юрійович — кандидат фізико-математичних наук, докторант, Інститут кібернетики імені В. М. Глушкова НАН України, завідуючий науково-дослідного відділу, ТОВ «Дельта СПЕ», м. Київ, vasyl.semenov@gmail.com

Семенова Євгенія Вікторівна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут математики НАН України, м. Київ, semenovaevgen@gmail.com

Сергієнко Іван Васильович — академік НАН України, доктор фізико-математичних наук, професор, директор, Інститут кібернетики імені В.М. Глушкова НАН України, м. Київ, vasilchuk.sp@gmail.com

Сидорук Володимир Антонович — кандидат фізико-математичних наук, науковий співробітник відділу, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, wolodymyr.sydoruk@gmail.com

Скуратовский Руслан В'ячеславович — викладач, Міжрегіональна академія управління персоналом, м. Київ, ruslan@unicyb.kiev.ua

Старков Вячеслав Миколайович — доктор фізико-математичних наук, старший науковий співробітник, провідний науковий співробітник, Інститут фізики НАН України, м. Київ, vjach-nikstar@gmail.com

Стецюк Петро Іванович — доктор фізико-математичних наук, старший науковий співробітник, завідуючий відділом, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, stetsyukp@gmail.com

Тимофієва Надія Костянтинівна — доктор технічних наук, старший науковий співробітник, провідний науковий співробітник, Міжнародний науково-навчальний центру інформаційних технологій та систем НАН та МОН України, м. Київ, TymNad@gmail.com

Ткаченко Олександр Володимирович — кандидат фізико-математичних наук, начальник відділу, ДП «Івченко-Прогрес», м. Запоріжжя, avt2007@outlook.com

Томчук Петро Михайлович — доктор фізико-математичних наук, професор, завідуючий відділом, Інститут фізики НАН України, м. Київ, ptomchuk@iop.kiev.ua

Тукалевська Нелля Іванівна — кандидат фізико-математичних наук, завідувач відділу, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, nvareniuk@ukr.net

Устименко Василь Олександрович — доктор фізико-математичних наук, професор, завідуючий відділом, Інститут телекомунікацій і глобального інформаційного простору НАН України, м. Київ; Університет Марії Кюрі-Склодовської, м. Люблін, Республіка Польща, vasylustimenko@yahoo.pl

Хіміч Олександр Миколайович — доктор фізико-математичних наук, професор, член-кореспондент НАН України, завідувач відділу, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, khimich505@gmail.com

Хом'як Ольга Миколаївна — кандидат фізико-математичних наук, науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, khomiak.olha@gmail.com

Чистяков Олексій Валерійович — кандидат фізико-математичних наук, науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, alexej.chystyakov@gmail.com

Чистякова Тамара Василівна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, tamara.chistjakova@gmail.com

Chable Daniel — Dipl. math. ETH Zurich, formerly Chief Actuary, Nestlé, Vevey, Switzerland, julien@bluewin.ch

Швідченко Інна Віталіївна — кандидат фізико-математичних наук, старший науковий співробітник, Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ, inetsheva@gmail.com

АЛФАВІТНИЙ ПОКАЖЧИК АВТОРІВ

ЗМІСТ

Інститут кібернетики імені В. М. Глушкова
Національної академії наук України
Кам'янець-Подільський національний університет
імені Івана Огієнка

НАУКОВЕ ВИДАННЯ

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

Серія: Фізико-математичні науки

Збірник наукових праць

Випуск 19

Редактор **В. П. Замула**
Комп'ютерна верстка **О. М. Коломис**

Підписано до друку 25.06.2019 р. Гарнітура «Таймс».
Папір офісний. Друк різографічний.
Формат 60x84/16. Умовн. друк. арк. 12,1. Обл.-вид. арк. 13,2.
Тираж 100. Зам. № 863.

Кам'янець-Подільський національний університет імені Івана Огієнка,
вул. Огієнка, 61, м. Кам'янець-Подільський, 32300.
Свідоцтво серії ДК № 3382 від 05.02.2009 р.

Надруковано в Кам'янець-Подільському національному
університеті імені Івана Огієнка,
вул. Огієнка, 61, м. Кам'янець-Подільський, 32300.