

О. М. Ткаченко, к. т. н., доц.; О. Д. Феферман

## МЕТОДИ КЛАСТЕРИЗАЦІЇ ДАНИХ У ЦИФРОВІЙ ОБРОБЦІ МОВЛЕННЄВИХ СИГНАЛІВ

*У статті розглянуто основні методи кластеризації даних, проаналізовано їх переваги та недоліки, сформульовано напрямки подальшого дослідження.*

**Ключові слова:** мовленнєві сигнали, ущільнення, кластеризація.

### Вступ

При параметричному кодуванні мовленнєвого сигналу в сучасних системах цифрового зв'язку замість безпосередньо коефіцієнтів лінійного прогнозування чи лінійних спектральних частот в канал зв'язку передаються індекси цих коефіцієнтів в таблиці – кодовій книзі. На приймальній стороні ці індекси декодуються із застосуванням такої самої кодової книги, а з отриманих параметрів синтезується мовленнєвий сигнал.

У кодовій книзі не можна зберігати всі можливі значення параметрів, оскільки в цьому випадку ми не отримаємо більшого ступеня ущільнення, тому в ній зберігаються лише найбільш репрезентативні значення, тобто такі, за допомогою яких можна закодувати весь спектр можливих значень параметрів. Таким чином при кодуванні сигналу після отримання певного вектору параметрів в кодовій книзі здійснюється пошук найближчого (найбільш схожого) вектора, і в канал зв'язку передається вже його індекс.

Оскільки необхідно, щоб значення в кодовій книзі були репрезентативні, створення кодової книги для ущільнення мовленнєвих сигналів відбувається таким чином: спочатку підбирається репрезентативний фонетичний матеріал, тобто такий, який містить всі можливі звуки, потім він кодується і отримується набір параметрів для кожного фрейма, а вже із застосуванням цього набору обраховуються значення для кодової книги. Саме для обрахунку значень, що зберігаються в кодовій книзі, на основі наявного матеріалу і застосовуються різноманітні методи кластеризації.

Кодові книги можуть бути скалярними або векторними. Скалярна кодова книга містить набір значень для кожного елемента вектора параметрів окремо. Векторна кодова книга містить набір векторів параметрів. Скалярні кодові книги більш прості в обчисленнях і потребують менше пам'яті для зберігання. Обчислювальна складність і необхідний обсяг пам'яті при застосуванні векторних кодових книг набагато більші порівняно зі скалярними, але вони забезпечують більший ступінь ущільнення.

Різні методи кластеризації застосовуються як для скалярних, так і для векторних кодових книг. Для того, щоб перейти безпосередньо до розгляду методів кластеризації, спочатку необхідно визначити, що таке кластеризація.

### Задача кластеризації при побудові кодових книг

Кластеризація – це розбиття певної множини об'єктів на підмножини (кластери), що не перетинаються, таким чином, щоб кожен кластер містив схожі об'єкти, а об'єкти різних кластерів відрізнялися між собою [1].

Нехай  $X$  – множина об'єктів,  $X^m$  – кінцева вхідна вибірка об'єктів,  $X^m = \{x_1, x_2, \dots, x_m\} \subset X$ .

Задано певну функцію відстані між об'єктами  $d(x, x')$ . Кластеризація, базуючись на вхідній вибірці об'єктів, визначає множину кластерів  $Y = \{y_1, y_2, \dots, y_n\}$ .

При цьому кожному  $x_i \in X^m$  ставиться у відповідність  $y_j \in Y$ . Кожен кластер складається з об'єктів, близьких за метрикою  $d$ . Визначається функція  $\alpha: X \rightarrow Y$ , яка кожному  $x \in X$  ставить у відповідність  $y \in Y$ .

У випадку, якщо об'єктами є числа або вектори чисел, кластеризацію називають квантуванням.

Скалярний квантизатор  $Q$  розміру  $N$  – це перетворення множини дійсних чисел  $x \in R$  у кінцеву множину  $Y$ , що містить  $N$  значень, які називаються кодовими векторами або центроїдами. Тобто  $Q: R \rightarrow Y$ , де  $(y_1, y_2, \dots, y_n) \in Y$  [2].

$Y$  називають кодовою книгою квантизатора. Операцію квантування можна також записати таким чином:

$$Q(x) = y_i, x \in R, i = 1, \dots, N. \quad (1)$$

Кодовим регіоном для скалярного квантизатора називатимемо частину простору дійсних чисел, таку, що

$$R_i = \{x \in R, Q(x) = y_i\} = Q^{-1}(y_i), i = 1, 2, \dots, N. \quad (2)$$

Також зрозуміло, що:

$$i \neq j \Rightarrow R_i \cap R_j = \emptyset. \quad (3)$$

А також:

$$\bigcup_i R_i = R. \quad (4)$$

Кодові регіони, що мають межі, називаються обмеженими, а такі, що не мають межі – необмеженими.

У галузі обробки сигналів операцію квантування можна розглядати як сукупність операцій кодування ( $E$ ) та декодування ( $D$ ):

$$Q(x) = y_i \Rightarrow E(x) = i, D(i) = y_i. \quad (5)$$

Квантизатор вважається оптимальним, якщо він мінімізує значення спотворення  $D$ :

$$D = \int_{-\infty}^{\infty} d(x, Q(x)) f_x(x) dx, \quad (6)$$

де  $d$  – функція відстані, визначена на множині дійсних чисел.

Умова найближчого сусіда для оптимальності квантизатора формулюється так:

$$R_i = \{x: d(x, y_i) \leq d(x, y_j)\}, \forall j \neq i. \quad (7)$$

Тобто для будь-якого вектора, що потрапляє в кодовий регіон, відстань до відповідного центроїда повинна бути менша, ніж відстань до будь-якого іншого центроїда у кодовій книзі.

Оскільки  $Q(x) = y_i$  тільки якщо  $d(x, y_i) \leq d(x, y_j)$ , то можна говорити, що:

$$d(x, Q(x)) = \min_i (d(x, y_i)). \quad (8)$$

Таким чином відстань від певного вектора до його квантованого значення – це мінімальне значення з відстаней до кожного центроїда у кодовій книзі.

Умова оптимальності центроїдів для кодового регіону  $R_i, i = 1, 2, \dots, N$ :

$$\sum d(x, y_i) \rightarrow \min, x \in R_i. \quad (9)$$

Тобто центроїд повинен обиратися таким чином, щоб він мінімізував спотворення для кожного вектора, що належить даному кодовому регіону. Дана умова повинна виконуватися

для всіх регіонів кодової книги.

Демо визначення векторному квантизатору. Векторний квантизатор  $Q$  розмірності  $K$  та розміру  $N$  – це перетворення  $K$ -мірного вектора  $x$  з Евклідового простору  $R^K$  у кінцевий набір  $Y$ , що містить  $N$   $K$ -вимірних векторів, які називаються кодовими векторами або центроїдами  $Q: R^K \rightarrow Y$ , де  $x = [x_1, x_2, \dots, x_K]$ ,  $(y_1, y_2, \dots, y_N) \in Y$ ,  $y_i = [y_i^1, y_i^2, \dots, y_i^K]$ ,  $i = 1, 2, \dots, N$ .

Умови оптимальності для векторного квантизатора визначаються аналогічно до таких же умов для скалярного.

### Чітка та нечітка кластеризація. Метод $k$ -середніх

При чіткій кластеризації об'єкти розділяються на окремі кластери таким чином, що кожен об'єкт може належати лише одному кластеру. При нечіткій кластеризації об'єкт може належати до декількох кластерів, при цьому зазначається ступінь його приналежності для кожного з них, який характеризує ступінь зв'язку об'єкта з певним кластером.

Прикладом чіткого методу кластеризації є алгоритм  $k$ -середніх [3] ( $k$ -means). Спочатку обирається кількість кластерів  $N$  та проводиться ініціалізація центроїдів  $y_1, y_2, \dots, y_N$ . Ініціалізація може проводитись як випадковим чином або просто рівномірним розбиттям, так і застосуванням певних більш простих алгоритмів кластеризації. Далі проводиться класифікація вхідних значень із застосуванням розрахованих центроїдів:

$$d(x, y_i) = \min_j d(x, y_j) \Rightarrow Q(x) = y_i, j = 1, 2, \dots, N. \quad (10)$$

На наступному кроці кожен центроїд замінюється середнім значенням об'єктів, що відносяться до відповідного кластеру.

$$y_i = \frac{\sum_{k=1}^{K_i} x_k}{K_i}, Q(x_k) = y_i, i = 1, 2, \dots, N, \quad (11)$$

де  $K_i$  – кількість векторів у  $i$ -му кластері.

Дані кроки повторюються, поки на певному кроці жоден центроїд не буде змінено.

Прикладом нечіткого методу кластеризації є нечіткий метод  $c$ -середніх (fuzzy  $c$ -means) [4].

Мета даного методу полягає у мінімізації показника  $J_m$ :

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m d^2(x_i, c_j), 1 \leq m < \infty, \quad (12)$$

де  $C$  – кількість центроїдів;  $N$  – кількість вхідних об'єктів;  $x_i$  –  $i$ -й вхідний об'єкт;  $c_j$  –  $j$ -й центроїд;  $d$  – функція відстані;  $u_{ij}$  – ступінь належності об'єкта  $x_i$   $j$ -му кластеру;  $m$  – так званий параметр нечіткості.

Алгоритм працює таким чином. На першому кроці проводиться ініціалізація матриці приналежності векторів  $U$ . Номеру кроку  $k$  присвоюється початкове значення.

На кожному  $k$ -му кроці обчислюємо значення центроїдів:

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m}. \quad (13)$$

Далі після оновлення центроїдів оновлюється матриця приналежності центроїдів:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{d(x_i - c_j)}{d(x_i - c_k)} \right)^{\frac{2}{m-1}}} . \quad (14)$$

Після цього обраховується умова закінчення роботи алгоритму. Якщо  $d(U^{(k+1)}, U^k) < \varepsilon$ , то робота алгоритму закінчується, інакше – відбувається перехід до кроку  $k+1$ .  $\varepsilon$  – додатне число в межах від 0 до 1.

Алгоритми нечіткої кластеризації, такі як c-means, рідко застосовуються в задачах ущільнення мовленнєвих сигналів, а частіше в задачах їх розпізнавання, а також у задачах цифрової обробки зображень. Метод k-means досить широко застосовується при ущільненні мовленнєвих сигналів. Зокрема його застосовують для побудови скалярних кодових книг для вокодера CELP. Основною перевагою даного методу є його простота та низька обчислювальна складність. Недоліком є те, що результат його роботи залежить від початкового вибору центроїдів і в загальному випадку метод не дає оптимальних результатів, а дозволяє знайти тільки субоптимальні рішення.

### Кластеризація з відомою кількістю кластерів. Метод LBG

Деякі методи потребують заздалегідь задати кількість кластерів, на які необхідно розбити тренувальну послідовність. Таким, наприклад, є метод k-середніх, розглянутий вище. Спочатку задається число кластерів і центроїдам присвоюються початкові значення, а після цього відбувається оптимізація кодової книги шляхом модифікації центроїдів та кодових регіонів. Існують і інші методи, як наприклад LBG (Linde, Buzo, Gray) [5]. В даному методі не потрібно задавати число кластерів та присвоювати початкові значення центроїдам. У роботі методу можна виділити декілька етапів.

Перш за все проводиться ініціалізація кодової книги. Обраховується середнє значення всіх векторів:

$$c_1^* = \frac{1}{M} \sum_{m=1}^M x_m , \quad (15)$$

де  $c_1^*$  – початковий кодовий вектор,  $M$  – кількість тренувальних векторів,  $x_m$  – тренувальний вектор.

Таким чином це значення і є першим кодовим вектором, що міститься в кодовій книзі. Кількість векторів в кодовій книзі встановлюється  $N=1$ . Після цього обраховується значення середньоквадратичного спотворення:

$$D_{ave}^* = \frac{1}{Mk} \sum_{m=1}^M D(x_m, c_1^*) , \quad (16)$$

де  $k$  – розмірність векторів, а функція  $D$  визначається так:

$$D(x, c) = (x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_k - c_k)^2 . \quad (17)$$

Наступний етап – це розділення векторів. Тобто кожен вектор, що вже є в кодовій книзі, поділяється на 2 вектори таким чином:

Для  $i = 1, 2, \dots, N$ :

$$c_i^{(0)} = (1 + \varepsilon) c_i^* , \quad (18)$$

$$c_{N+i}^{(0)} = (1 - \varepsilon) c_i^* . \quad (19)$$

Далі виконується ряд кроків, поки не буде отримано бажану кількість векторів (LBG Наукові праці ВНТУ, 2010, № 1

дозволяє отримати  $2^n$  векторів). На кожному кроці спочатку відбувається оновлення кодових регіонів. При цьому для кожного тренувального вектора знаходиться кодовий регіон, до якого він знаходиться ближче всього і тренувальний вектор переноситься до відповідного кодового регіону. Тобто, якщо для  $m = 1, 2, \dots, M$  та  $n = 1, 2, \dots, N$  :

$$\min(D(x_m, c_n^{(i)})) = D(x_m, c_{n^*}^{(i)}), \quad (20)$$

де  $i$  – номер ітерації, то:

$$Q(x_m) = c_{n^*}^{(i)}. \quad (21)$$

Після цього оновлюються значення кодових векторів. Кожен вектор прирівнюється до середнього значення всіх векторів, що належать до відповідного кодового регіону:

$$c_n^{(i+1)} = \frac{\sum_{Q(x_m)=c_n^{(i)}} x_m}{\sum_{Q(x_m)=c_n^{(i)}} 1}, \quad n = 1, 2, \dots, N. \quad (22)$$

Після цього проводиться розрахунок середньоквадратичного спотворення:

$$D_{ave}^{(i)} = \frac{1}{Mk} \sum_{m=1}^M D(x_m, Q(x_m)). \quad (23)$$

Далі з'ясується, як це значення змінилося відносно попереднього. Якщо необхідно, операція оновлення кодових регіонів та центроїдів повторюється і лише після цього відбувається перехід до наступного кроку.

Метод LBG є досить поширеним в обробці мовленнєвих сигналів та зображень, зокрема при ущільненні мовленнєвих сигналів. Основними перевагами даного методу є відносно невисока обчислювальна складність, простота реалізації, незалежність результатів роботи методу від початкових даних (оскільки центроїди формуються динамічно на основі тренувальної послідовності і нема потреби задавати їм певні початкові значення). Взагалі метод дає непогані результати, але має і певні недоліки, що пов'язані зі специфікою методу, а саме – нерівномірне розташування центроїдів у кодовій книзі відносно розподілу об'єктів тренувальної послідовності. Цей недолік зумовлено тим, що значення кожного наступного центроїда обраховується на основі попереднього, і таким чином неможливо кардинально змінити положення центроїда у випадку необхідності, оскільки він може змінюватися лише в певних межах.

### Роздільне та багатоетапне квантування

При застосуванні векторних кодових книг не завжди можливо зберігати у кодовій книзі вектори параметрів такої ж самої розмірності, як і вектори тренувальної послідовності. Представлення всіх необхідних варіантів векторів у кодовій книзі потребувало б надзвичайно великих витрат пам'яті, що ускладнило б реалізацію таких систем на персональних комп'ютерах та взагалі майже унеможливило б їх апаратну реалізацію.

Тому існує два різних підходи до застосування векторних кодових книг – роздільне (split) та багатоетапне (multistage) квантування.

При роздільному квантуванні вектор параметрів ділиться на підвектори і кожен підвектор кодується за допомогою своєї векторної кодової книги. Векторна кодова книга будується для кожного підвектора незалежно із застосуванням одного з алгоритмів кластеризації. При цьому необхідно визначити, на скільки векторів розбивати початкові вектори, скільки елементів має бути в кожному підвекторі та скільки значень зберігати в кодовій книзі для кожного підвектора.

Багатоетапне квантування побудовано на зовсім іншій ідеї по відношенню до роздільного

квантування. Основна відмінність полягає в тому, що в кодовій книзі зберігаються вектори, розмірність яких збігається із векторами кодової послідовності. При цьому даний метод створює декілька кодових книг. Проте при квантизації певного вектора він замінюється не просто вектором із однієї з цих кодових книг, а комбінацією векторів із різних кодових книг, сформованих даним методом.

У кодері вхідний вектор  $x$  порівнюється з вектором  $\hat{x}$ :

$$\hat{x} = y_{i1}^{(1)} + y_{i2}^{(2)} + \dots + y_{iK}^{(K)}, \quad (24)$$

де  $y_i^{(l)}$  – це  $i$ -й кодовий вектор з кодової книги, отриманої на  $l$ -му етапі. Всі кодові вектори мають таку саму розмірність, як і вхідний вектор.

Шляхом вибору різних індексів кодер намагається мінімізувати відстань між  $x$  та  $\hat{x}$ . Після знаходження такого набору  $\{i_1, i_2, \dots, i_K\}$ , який мінімізує відстань, індекси передаються на приймальну сторону, де декодер із застосуванням таких самих кодових книг відновлює сигнал.

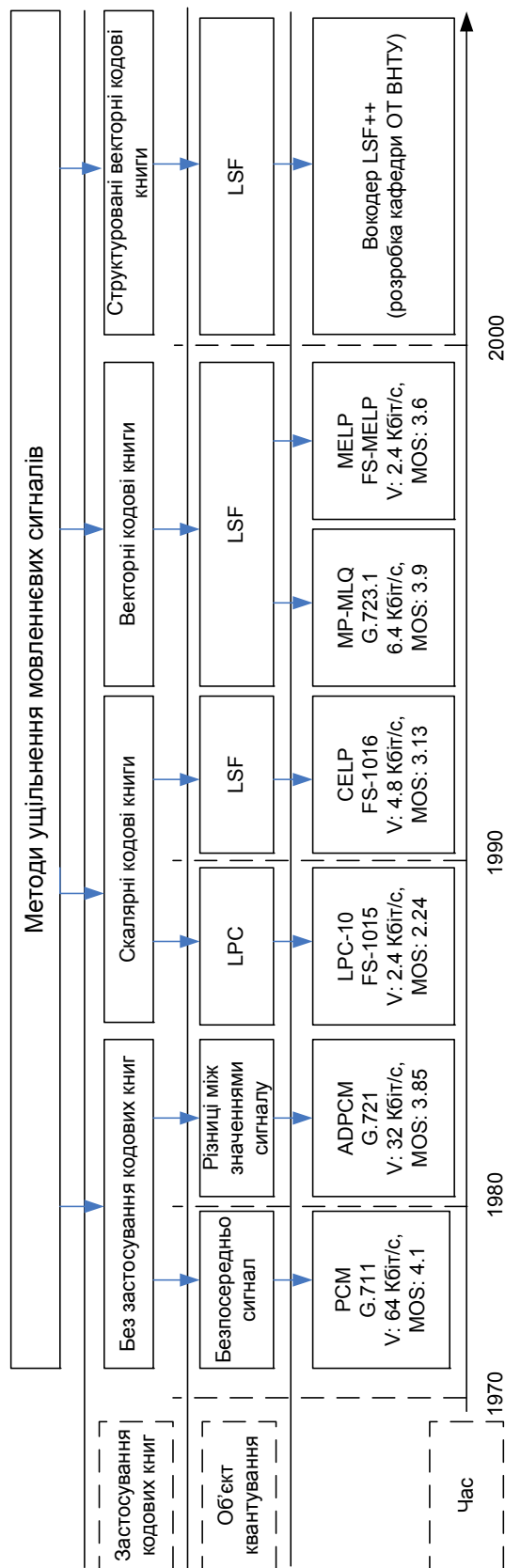
Одна з основних переваг багатоетапного квантування порівняно з роздільним – це зменшення обсягів пам'яті, необхідних для зберігання кодових книг. В плані обчислювальної складності багатоетапне квантування потребує більше ресурсів порівняно з роздільним.

Основний недолік роздільного квантування – це те, що в процесі квантування може порушуватись стійкість системи параметрів. Між елементами квантованого вектора параметрів є певні залежності, які в процесі квантування можуть порушуватися. Відповідно це призводить до значного спотворення ущільненого сигналу. При багатоетапному квантуванні цей недолік відсутній, оскільки вектори, що зберігаються в кодових книгах багатоетапного квантизатора мають таку саму розмірність, як і тренувальні вектори, і згадані вище залежності в них не порушуються. Багатоетапне квантування застосовується для побудови векторних кодових книг в алгоритмі MELP.

### **Класифікація методів ущільнення мовленнєвих сигналів**

Як правило, вокодери класифікують лише за одним параметром – швидкістю передавання мовленнєвого сигналу в канал зв'язку. Таким чином їх поділяють на низькошвидкісні, середньошвидкісні та високошвидкісні [6]. Проте така класифікація не дає можливості оцінити закономірності розвитку методів ущільнення мовленнєвих сигналів у часовій перспективі.

Вище було розглянуто різні методи кластеризації, які можуть бути застосовані при ущільненні мовленнєвих сигналів. Оскільки в кожному з методів ущільнення мовленнєвих сигналів застосовується квантування в тому чи іншому вигляді, то це можна застосувати як одну з ознак їх класифікації. У даній роботі пропонується класифікація методів ущільнення мовленнєвих сигналів, наведена на рис. 1.



\* V – пропускна спроможність, необхідна для створення одностороннього мовленнєвого каналу.  
 \*\* MOS (Mean opinion score) – середня суб'єктивна оцінка якості звуку.

Рис. 1. Класифікація методів ущільнення мовленнєвих сигналів

На першому етапі всі методи діляться за типом кодових книг, які вони застосовують, а саме на такі, які не застосовують ніяких кодових книг, тобто просто квантують значення за певним законом, такі, які застосовують скалярні кодові книги, та такі, що застосовують векторні кодові книги. На наступному етапі методи поділяються за об'єктом квантування, тобто за тим, що зберігається в кодовій книзі. Можна виділити методи, які квантують безпосередньо сигнал, методи, що квантують прості похідні від сигналу (як, наприклад, різницю між попереднім та наступним значенням сигналу), методи, що квантують коефіцієнти лінійного прогнозування (LPC) та методи, що квантують лінійні спектральні частоти (LSF).

### Напрямки подальшого дослідження

Наведена вище класифікація дозволяє побачити закономірності розвитку методів ущільнення мовленнєвих сигналів і спрогнозувати напрямки їх подальшого розвитку. Оскільки в даній класифікації чітко прослідковується часова послідовність розвитку методів ущільнення мовленнєвих сигналів, можна зробити висновки про шляхи подальшого розвитку даних методів.

1. Дослідження та розробка загальних принципів побудови структурованих векторних кодових книг, які мають замінити неструктуровані кодові книги.
2. Розробка критеріїв оцінювання якості кластеризації для структурованих кодових книг.
3. Розробка методів упорядкування та пошуку векторів у структурованих кодових книгах.
4. Об'єктом квантування у векторних кодових книгах можуть залишатися лінійні спектральні частоти, проте перспективним видається дослідження альтернативних моделей представлення спектральної інформації про сигнал.

### СПИСОК ЛІТЕРАТУРИ

1. A. K. Jain. Data clustering: a review / A. K. Jain, M. N. Murty, P. J. Flynn.– ACM Comput. Surv.–1999. – №31. – 60 с.
2. J. Makhoul. Vector quantization in speech coding / J. Makhoul, S. Houcos, H. Gish . – Proc. IEEE. – 1985 – №73. – pp. 1551 – 1558.
3. P. Hedelin. Vector Quantization for Speech Transmission / P. Hedelin, P. Knagenhjelm, M. Skoglund. – Speech Coding and Synthesis. – 1995. – pp. 311 – 346.
4. J. C. Dunn A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well Separated Clusters / J. C. Dunn. – Journal of Cybernetics. – 1973. – №3. – pp. 32 – 57.
5. Linde Y. An Algorithm for Vector Quantizer Design / Linde, Y., Buzo, A., Gray, R. – IEEE Transactions on Communications. – 1980. – №28., pp. 84 – 94.
6. Рабинер Л. Р. Цифровая обработка речевых сигналов: Пер. с англ. / Рабинер Л. Р., Шафер Р. В. – М.: Радио и связь, 1981. – 496 с.

**Ткаченко Олександр Миколайович** – доцент кафедри обчислювальної техніки.

**Феферман Олег Дмитрович** – аспірант кафедри обчислювальної техніки.  
Вінницький національний технічний університет.