

УДК 004.931:517.518.3

*У роботі розглянуто теоретичні засади та практичні аспекти використання поліномів наближення в просторі з породжуючим елементом (просторі Кунченка) для аналізу статистичних даних*

*Ключові слова: порівняння з еталоном, простір із породжуючим елементом, поліном Кунченка, мікрофайл*

*В работе рассмотрены теоретические начала и практические аспекты использования полиномов приближения в пространстве с порождающим элементом (пространстве Кунченко) для анализа статистических данных*

*Ключевые слова: сравнение с эталоном, пространство с порождающим элементом, полином Кунченко, микрофайл*

*In the paper, we discuss theoretical basics and practical aspects of applying approximation polynomials in a space with a generative element (a Kunchenko space) to statistical data analysis*

*Keywords: template matching, space with a generative element, Kunchenko polynomials, microfile*

# ЗАСТОСУВАННЯ ПОЛІНОМІВ КУНЧЕНКА ДЛЯ АНАЛІЗУ СТАТИСТИЧНИХ ДАНИХ

**О.Р. Чертов**

Кандидат технічних наук, доцент\*

**Д.Ю. Тавров**

Аспірант\*

\*Кафедра прикладної математики

Національний технічний університет України «Київський політехнічний інститут»  
пр. Перемоги, 37, м. Київ

## Вступ

В останні десятиліття великого поширення здобула галузь обробки цифрової інформації, пов'язана з пошуком окремих наперед визначених властивостей чи образів у деякому заданому масиві даних. Задачі такого типу належать до розпізнавання образів (pattern recognition) [1, 2].

Потреба мати справу з такого роду завданнями постає в доволі різних сферах людської діяльності: від забезпечення машинного зору та розпізнавання людського почерку до інтелектуального аналізу даних. У даній роботі зосередимо увагу на таких задачах, де образ чи властивість, які ми хочемо віднайти в питомому масиві даних, відомі заздалегідь, причому належний образ називатимемо шаблоном чи еталоном. Відповідний напрям досліджень має назву порівняння чи співставлення з еталоном (template matching).

Зазвичай, постановка такої задачі включає в себе дані для аналізу, деякий шаблон для пошуку та, можливо, область, на якій потрібно здійснювати такий пошук. При цьому потрібно як правильно обрати представлення вхідних даних, так і визначити критерії, за якими відбуватиметься співставлення шаблону з сигналом.

Дані можна представити різними способами. Це може бути звичайний цифровий сигнал або двовимірне графічне зображення (при чому, в останній час було запропоновано низку підходів до зведення двовимірних зображень до одновимірних: див., наприклад, [3]).

У різних джерелах можна знайти розмаїтій набір підходів, метрик та відстаней, які можна використовувати при порівнянні з еталоном. Наприклад, застосовують такі підходи, як принцип оптимальності

Белмана (Bellman) [4], підхід із відстанню редагування для аналізу символічних рядків [5], а також підхід із використанням кореляцій ([6, 7]).

Даний напрям активно розвивається й донині, зокрема, зовсім нещодавно було запропоновано альтернативний спосіб забезпечення пошуку шаблонів із використанням технології найближчого сусіда [8].

У даній роботі ми пропонуємо новий метод пошуку заздалегідь визначеного шаблону в цифровому сигналі, який базується на застосуванні поліномів Кунченка, та обговорюємо можливість використання цього методу для перевірки статистичних гіпотез під час опрацювання статистичних даних.

## Постановка завдання

Маємо (двовимірний) масив статистичних даних, отриманих у результаті визначеного статистичного спостереження, наприклад, перепису населення. Існує припущення про наявність у розподілі значень окремих атрибутів певних закономірностей (шаблонів). Потрібно перевірити справедливості цього припущення шляхом побудови поліномів наближення в просторі з породжуючим елементом.

## 1. Загальні теоретичні відомості

### 1.1. Побудова поліномів наближення в просторі з породжуючою функцією

У роботі [9] Ю.П. Кунченком було докладно представлено процедуру побудови поліномів наближення

в просторі з породжуючою функцією і наведено низку прикладів, які переконливо підтверджують ефективність такого підходу для апроксимації функцій. У даному підрозділі торкнемося основних моментів і наведемо найважливіші результати, отримані Ю.П. Кунченком та необхідні для викладення подальшого матеріалу. За детальнішою інформацією можна звернутися до монографії [9].

Отже, нехай задано так звану породжуючу функцію  $f(x)$ , визначену на проміжку  $[a,d]$ . Тоді, введемо впорядковану множину породжених функцій наступного вигляду:

$$u_v(x) = c_v[f(x)], \quad x \in [a,d], \quad (1)$$

де  $c_v(\cdot)$  – дійсні функції, підібрані певним чином.

Якщо ввести на цій множині лінійні операції додавання та множення на число, то множина всіх лінійних комбінацій породжених функцій сформує лінійний простір.

Оскільки деякі з породжених функцій (1) можуть виявитися лінійно залежними, пропонуємо визначити просторово утворюючу множину функцій, утворену як об'єднання множини лінійно незалежних породжених функцій із (1). Відповідно, лінійний простір, утворений на такій множині, називатимемо лінійним простором над незалежними породженими функціями (лінійним простором Кунченка) та позначатимемо його через LFKu.

Визначимо в просторі LFKu скалярний добуток двох елементів звичним чином:

$$(u_v(x), u_k(x)) \equiv \int_a^d c_v(x) \cdot c_k(x) dx = \int_a^d c_v[f(x)] \cdot c_k[f(x)] dx. \quad (2)$$

Із формули (2) випливає визначення відстані між двома елементами лінійного простору Кунченка як норми різниці двох функцій:

$$c_{v,k}^2 \equiv \|u_v(x) - u_k(x)\|^2 = \int_a^d [u_v(x) - u_k(x)]^2 dx \quad (3)$$

Якщо для всіх породжених функцій, що утворюють простір LFKu, можна ввести скалярний добуток (2), то такий простір називатимемо простором LFKu зі скалярним добутком.

Простір LFKu, у якому множина породжених функцій складається тільки з лінійно незалежних, попарно (повністю чи частково) неортогональних функцій, називатимемо простором (Кунченка) з породжуючим елементом. Під частковою неортогональністю породжених функцій розуміємо ситуацію, коли одна їхня непорожня частина є попарно неортогональною, а інша – попарно ортогональною.

Надалі, значення скалярного добутку (2) також називатимемо корелянтою двох функцій  $u_v(x)$  і  $u_k(x)$  розміру  $(v,k)$ , яку позначатимемо наступним чином:

$$\Psi_{v,k} \equiv \int_a^d u_v(x) \cdot u_k(x) dx. \quad (4)$$

Також, виберемо деяку породжену функцію  $u_b(x)$ , яку назвемо основною. Тоді, можна утворити так званий поліном Кунченка із породжених функцій  $u_v(x)$ ,  $v \neq b$ , які назвемо доповнювальними:

$$P_r(x) \equiv \sum_{\substack{v=0 \\ v \neq b}}^r \beta_v u_v(x) = \beta_0 u_0(x) + \sum_{\substack{v=1 \\ v \neq b}}^r \beta_v u_v(x). \quad (5)$$

Якщо розглянути задачу мінімізації відстані між поліномом Кунченка (5) та основною функцією  $u_b(x)$ :

$$\rho_{bp}^2 = \int_a^d [u_b(x) - P_r(x)]^2 dx, \quad (6)$$

то можна показати [9, с.75-76], що коефіцієнти  $\beta_v$ ,  $v \neq 0$  знаходяться з наступної системи лінійних рівнянь:

$$\sum_{\substack{k=1 \\ k \neq b}}^r \beta_k F_{v,k} = F_{v,b}, \quad v = \overline{1,r}, \quad v \neq b, \quad (7)$$

де центровані кореляти  $F_{v,k} \equiv \Psi_{v,k} - \Psi_v \cdot \Psi_k \cdot \|u_0(x)\|^2$ ,

$$\Psi_v \equiv \frac{\Psi_{v,0}}{\|u_0(x)\|^2} = \frac{\int_a^d u_v(x) \cdot u_0(x) dx}{\int_a^d u_0^2(x) dx}.$$

Коефіцієнт  $\beta_0$  повинен дорівнювати наступному виразу:

$$\beta_0 = \Psi_b - \sum_{\substack{v=1 \\ v \neq b}}^r \beta_v \Psi_v. \quad (8)$$

Якщо підставити коефіцієнти (7-8) у вираз (6), то ми отримаємо:

$$c_{bp}^2 = F_{b,b} - J_r, \quad (9)$$

де так звана інфоркуна полінома (5) має вигляд

$$J_r \equiv \sum_{\substack{v=1 \\ v \neq b}}^r \beta_v F_{v,b} = \sum_{\substack{v=1 \\ v \neq b}}^r \sum_{\substack{k=1 \\ k \neq b}}^r \beta_v \beta_k F_{v,k}.$$

Нарешті, уведемо чисельний показник якості наближення поліномом (5) основної функції  $u_b(x)$ :

$$e_r = \frac{J_r}{\int_a^d (u_b(x) - \Psi_b \cdot u_0(x))^2 dx}. \quad (10)$$

Коефіцієнт (10) називатимемо коефіцієнтом ефективності наближення полінома з доповнювальних функцій до основної функції. Що ближче даний коефіцієнт до 1, то краще є наближення поліномом Кунченка відповідної функції [9, с. 86].

Варто зазначити, що оскільки множина доповнювальних функцій складається з повністю або частково неортогональних функцій, а основна функція не входить до множини доповнювальних, то ситуація, за якої коефіцієнт ефективності (10) дорівнює 1, можлива тільки в граничному випадку, коли порядок полінома Кунченка  $r$  прямує до нескінченності.

### 1.2. Застосування ідеї побудови поліномів Кунченка до задачі розпізнавання шаблону в сигналі

Розглянемо модифікацію описаного вище підходу до побудови полінома Кунченка, виходячи з потреб задачі розпізнавання певного шаблону в сигналі.

Спочатку визначимо, які вхідні дані ми повинні мати для розв'язання поставленої задачі, а потім укажемо для них місце в схемі, запропонованій Ю.П. Кунченком [9].

Отже, для виконання розпізнавання конкретного образу в деякому сигналі ми повинні мати сам сигнал  $s(x)$ , визначений на проміжку  $[a,d]$ , а також певний функціональний шаблон  $ш(x)$ ,  $x \in [a,d]$ . За таких початкових даних нам потрібно отримати відповідну кількісну оцінку схожості обраного шаблону на сигнал для аналізу.

Таким чином,

– у якості основної функції візьмемо початковий сигнал:  $u_b(x) = s(x)$ ;

– у якості доповнювальних функцій потрібно взяти лінійно незалежні та частково неортогональні на проміжку  $[a,d]$  функціональні перетворення шаблону:  $u_v(x) = \psi_v[\psi(x)]$ ;

– кількісною характеристикою схожості шаблону та сигналу може слугувати коефіцієнт ефективності (10).

Проте, дану схему можна застосувати до оцінки схожості двох функцій, визначених тільки на одному й тому ж проміжку. У багатьох практичних задачах довжина шаблону значно коротша за довжину сигналу, тому існує потреба у визначенні ділянок сигналу, які відповідають якнайбільшій близькості до шаблону.

Розглянемо сигнал  $s(x)$ , який визначено на проміжку  $[a,d]$ , та шаблон  $\psi(x)$  визначений на іншому проміжку  $[i,j] \subseteq [a,d]$ . При цьому для нас непринципово, яких конкретно значень набувають числа  $i$  та  $j$ ; важливо лише, що обидва ці числа лежать на проміжку  $[a,d]$ , і довжина проміжку  $[i,j]$  фіксована. Тоді, можна здійснити процедуру побудови полінома Кунченка для кожної ділянки початкового сигналу  $s(x)$  довжиною  $|i-j|$ , яку ми назвемо вікном.

У результаті, ми отримуємо множину значень коефіцієнтів ефективності:

$$E = \{e(x)\} \quad \forall x \in [a, d - i + j + 1], \tag{11}$$

яку назвемо ефективною множиною, а графік, що представляє значення з цієї множини, називатимемо ефектограмою. Аналізуючи ефектограму, можна визначити, які ділянки сигналу і наскільки відповідають запропонованому шаблону.

### 1.3. Використання поліномів Кунченка при аналізі статистичних даних мікрофайлів

Процедуру побудови поліномів наближення Кунченка разом із чисельним показником якості такого наближення в вигляді коефіцієнту ефективності можна застосовувати до аналізу статистичних даних.

Якщо первинні статистичні дані зведено до мікрофайлу (див. табл. 1), де кожному респонденту  $r_k$ ,  $k = \overline{1, \mu}$  співставлено значення  $z_{kn}$  атрибутів  $u_n$ ,  $n = \overline{1, \eta}$ , то можна скористатися визначеннями, запропонованими в роботах [10] та [11]. Зокрема, для нас принциповими є поняття сутнісних (vital), тобто найважливіших із погляду розв'язуваної задачі, атрибутів та їхніх комбінацій і параметризуючих (parameter) атрибутів, тобто атрибутів, за значеннями яких будується розподіл сутнісних комбінацій значень.

Таблиця 1

Дані мікрофайлу, оформлені в вигляді таблиці

		атрибути			
		$u_1$	$u_2$	...	$u_3$
респонденти	$r_1$	$z_{11}$	$z_{12}$	...	$z_{13}$
	$r_2$	$z_{21}$	$z_{22}$	...	$z_{23}$
	...	...	...	...	...
	$r_m$	$z_{m1}$	$z_{m2}$	...	$z_{m3}$

Нехай, із міркувань конкретної практичної задачі, визначено і сутнісні, і параметризуючі атрибути та ком-

бінації. Тоді, відповідно до [10], можна збудувати кількісний сигнал, який складається з кількостей повторень конкретної пари «сутнісна комбінація – параметризуюче значення». У контексті поставленої задачі по розпізнаванню шаблону в деякому сигналі, у якості такого кількісного сигналу візьмемо сигнал  $s(x)$  із тією додатковою умовою, що параметризуючі значення повинні бути впорядковані за певним заздалегідь визначеним правилом (наприклад, якщо параметризуючим атрибутом є «вік», то відповідні значення можна впорядкувати за зростанням). Звісно, окрім кількісного сигналу можна взяти будь-який інший масив даних, який у тій чи іншій мірі задовольняє вимоги конкретного дослідження.

У якості ж шаблону  $\psi(x)$  можна вибирати такі функціональні залежності, які було отримано шляхом певного попереднього аналізу статистичних даних. Тобто, іншими словами, поліноми Кунченка зручно використовувати для перевірки певних статистичних гіпотез по розподілу значень сутнісних атрибутів відповідно до деяких практичних міркувань.

У наступному розділі ми запропонуємо декілька конкретних практичних прикладів застосування викладеної ідеї.

## 2. Практичні результати

### 2.1. Визначення початкового сигналу для аналізу

Для демонстрації можливостей запропонованого методу з застосуванням поліномів Кунченка до пошуку певних шаблонних закономірностей у статистичних даних, ми вирішили взяти 5-відсоткові мікродані, що стосуються перепису населення, проведеного Бюро перепису США [12] у 2000 р. При цьому ми вирішили обмежитися даними по штату Каліфорнія як найчисельнішому за населенням та найпотужнішому за економічним розвитком.

У силу загальних міркувань (велика вибірка незалежних випадкових величин), розподіл середнього доходу домогосподарства в залежності від віку голови останнього повинен бути достатньо близький до нормального. Тому поставимо більш складну задачу: спробуємо відшукати в указаному розподілі ділянки, що мають розподіл, близький до нормального. Існування декількох таких ділянок може свідчити про те, що досліджуваний розподіл утворений декількома нормально розподіленими величинами. Щоб перевірити наявність такої ситуації, нам потрібно спочатку збудувати відповідний сигнал для аналізу.

У якості такого сигналу  $s(x)$  ми вирішили взяти дещо модифіковану версію кількісного сигналу, описаного в попередньому розділі. Для сформування цього сигналу, ми визначили наступні атрибути:

– за сутнісний атрибут ми вибрали атрибут «Family Total Income in 1999» («Сукупний дохід родини в 1999 р.»), котрий повністю відповідає нашій постановці задачі, оскільки містить вичерпну інформацію про рівень доходу конкретної сім'ї за 1999 р.;

– у якості параметризуючого атрибуту ми взяли атрибут «Age» («Вік»), тому що значення саме цього атрибуту нам потрібно використовувати для впорядкування значень кількісного сигналу.

Визначивши відповідні атрибути, ми підрахували середній дохід на родину по всіх респондентах одного

віку, і, впорядкувавши отримані значення за зростанням віку респондента, сформували відповідний сигнал  $s(x)$  (див. рис. 7а). Варто зазначити, що на рис. 7 перший відлік відповідає віку в 15 років (оскільки в мікрофайлі не було виявлено жодного голови домогосподарства молодше). Таким чином, довжина нашого модельного сигналу складає 75 значень.

У наступному підрозділі ми розглянемо ті функціональні шаблони, які було використано для розв'язання поставленої на початку розділу задачі.

**2.2. Шаблони для використання при побудові полінома Кунченка**

Оскільки нашою основною задачею є перевірка гіпотези про нормальний розподіл значень відповідного сигналу  $s(x)$ , то першим і основним шаблоном було взято класичну гаусіану  $\psi_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ . При цьому, через те, що нас цікавлять у першу чергу локальні особливості відповідного сигналу, ми вирішили обмежити носій відповідної експоненти відрізком  $[-3, 3]$ . Відліки з цього відрізка було взято з кроком 0,2. Таким чином, ми отримуємо шаблон довжиною в 31 відлік, що цілком дозволяє, використовуючи підхід із «плаваючим вікном», запропонований у підрозділі 1.2, визначити наявність у початковому сигналі певних ознак нормальності розподілу його окремих ділянок. Графік відповідного шаблону представлено на рис. 1.

Окрім зазначеного шаблону, ми вирішили взяти також два інші шаблони, щоб додатково перевірити коректність застосування поліномів Кунченка до розв'язання поставленої задачі.

Один із цих шаблонів у певному сенсі схожий на відповідник із рис. 1, а другий має зовсім іншу форму та структуру.

У якості шаблону, який за своєю формою та властивостями неістотно відрізняється від гаусіани в контексті поставленої задачі, ми взяли функцію  $\psi_2(x) = 1 - |x|$ , визначену на тій же дискретній множині значень, що й перший шаблон. Відповідний графік представлено на рис. 2.

Шаблони з рис. 1 та рис. 2 мають одну спільну рису – вони унімодальні, із чітко вираженим максимумом. Таким чином, здається майже очевидним, що результати застосування обох шаблонів до початкового сигналу можуть дати близькі результати.

У якості шаблону, який зовсім не схожий на обидва вказані вище, ми взяли функцію  $\psi_3(x) = \sin 4x$ , знову таки, визначену на тому самому носії. Її графік представлено на рис. 3.

Даний шаблон відрізняється від попередніх наявністю одразу декількох екстремумів, таким чином, швидше за все, результати його застосування до побудови полінома Кунченка будуть дещо інакші.

У наступному підрозділі ми розглянемо конкретні результати побудови відповідних поліномів Кунченка, а також спробуємо визначити основні особливості запропонованого методу.

**2.3. Результати побудови поліномів Кунченка на основі запропонованих шаблонів**

У силу невідповідності довжин сигналу  $s(x)$  та шаблонів  $\psi_1(x)$ ,  $\psi_2(x)$ ,  $\psi_3(x)$  потрібно використовувати підхід із «плаваючим вікном», тобто будувати окремий поліном Кунченка в кожному вікні довжиною в 31 відлік. У результаті, згідно з підрозділом 1.2, ми отримаємо

ефектограму, котра покаже, з якою ефективністю було побудовано поліном Кунченка в тому чи іншому вікні.

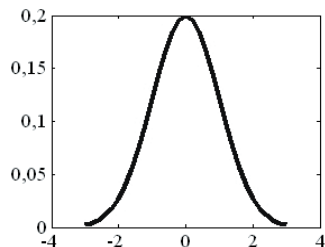


Рис. 1. Графік першого шаблону  $\psi_1(x)$  для побудови полінома Кунченка

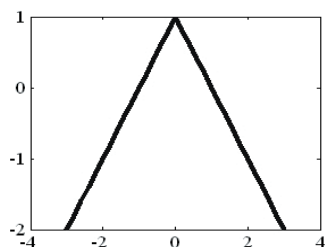


Рис. 2. Графік другого шаблону  $\psi_2(x)$  для побудови полінома Кунченка

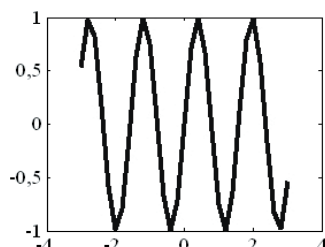


Рис. 3. Графік третього шаблону  $\psi_3(x)$  для побудови полінома Кунченка

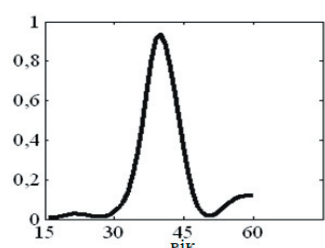


Рис. 4. Ефектограма, отримана за результатами наближення початкового сигналу шаблоном  $\psi_1(x)$

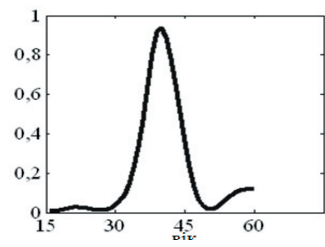


Рис. 5. Ефектограма, отримана за результатами наближення початкового сигналу шаблоном  $\psi_2(x)$

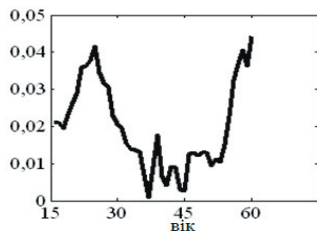


Рис. 6. Ефектограма, отримана за результатами наближення початкового сигналу шаблоном  $\psi_3(x)$

Для першого шаблону  $\psi_1(x)$  ефектограма має вигляд, представлений на рис. 4.

Аналізуючи вигляд даної ефектограми, можна відразу визначити, що найбільшої ефективності було досягнуто в відліку 25 (що відповідає віку в 40 років), при цьому коефіцієнт ефективності дорівнює  $e = 0,9368$  (усі результати в цій роботі презентуються з чотирма знаками після коми, хоча обчислення проводилися зі значно більшою точністю). Якщо звернутися до рис. 7, а, то суто візуально підтвердити цей результат достатньо складно. Це свідчить про велику практичну користь від використання поліномів Кунченка при визначенні відповідних шаблонів.

Також, згідно з рис. 4, фактично, що далі ми віддаємося від зазначеного відліку, то меншими стають відповідні коефіцієнти ефективності. Це дозволяє зробити висновок, що в сигналі  $s(x)$  присутня єдина ділянка, близька до обраного шаблону. А оскільки сам шаблон є гаусіаном, то це свідчить про те, що в створенні початкового сигналу брала участь лише одна нормально розподілена випадкова величина.

Для демонстрації основних кроків із побудови полінома Кунченка, пропонуємо зосередитися на деякому одному вікні в 31 відлік, оскільки процес побудови полінома Кунченка в усіх інших вікнах абсолютно аналогічний. У якості такого вікна пропонуємо взяти вікно, що починається з відліку 25, тобто вікно, що відповідає найбільшому коефіцієнту ефективності.

У даному випадку ми маємо наступний шаблон:

$\psi_1 = (0,0044, 0,0079, 0,0136, 0,0224, 0,0355, 0,0540, 0,0790, 0,1109, 0,1497, 0,1942, 0,2420, 0,2897, 0,3332, 0,3683, 0,3910, 0,3989, 0,3910, 0,3683, 0,3332, 0,2897, 0,2420, 0,1942, 0,1497, 0,1109, 0,0790, 0,0540, 0,0355, 0,0224, 0,0136, 0,0079, 0,0044)$

і таку ділянку сигналу:

$s_{25-55} = 104 \cdot (7,1647, 7,3292, 7,4860, 7,6912, 7,7397, 7,8440, 8,0223, 8,1808, 8,2318, 8,6000, 8,5164, 8,7085, 9,1391, 9,0642, 9,2785, 9,4966, 9,3023, 9,3524, 9,3858, 8,9799, 8,7503, 8,8685, 8,6178, 8,1136, 7,8374, 7,8823, 7,4795, 7,0684, 6,7641, 6,6724, 6,5417)$ .

Отже, збудуємо поліном Кунченка 5-ї степені на основі шаблону  $\psi_1(x)$ . Варто зазначити, що порядок полінома для даної конкретної задачі не становить суттєвого значення. Як було показано в [9, гл. 4], на практиці в більшості випадків вже за 6 членів полінома Кунченка досягається гарна апроксимація, і середньоквадратична відстань між цим поліномом та основною функцією  $s(x)$  близька до нуля.

Для побудови полінома Кунченка потрібно спочатку визначити породжені функції (1). У всіх прикладах із цієї статті ми вирішили брати ступеневі перетворення відповідного шаблону в якості породжених функцій:  $u_v(x) = \psi_v[\psi_1(x)] = \psi_1^v(x)$ ,  $v = 0, 5$ ,  $i = 1, 3$ .

Потрібно розв'язати систему лінійних рівнянь (7), коефіцієнтами якої є центровані корелянти. У рамках даного дослідження при обчисленні скалярного добутку шаблону та ділянки сигналу ми використовували формулу трапецій для підрахунку інтеграла таблично заданої функції. Із цими уточненнями, система рівнянь має наступний вигляд:

$$\begin{pmatrix} 0,5817 & 0,2250 & 0,0824 & 0,0303 & 0,0112 \\ 0,2250 & 0,0924 & 0,0350 & 0,0132 & 0,0050 \\ 0,0824 & 0,0350 & 0,0136 & 0,0052 & 0,0020 \\ 0,0303 & 0,0132 & 0,0052 & 0,0020 & 0,0008 \\ 0,0112 & 0,0050 & 0,0020 & 0,0008 & 0,0003 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} = \begin{pmatrix} 32462,812 \\ 11899,5728 \\ 4251,7946 \\ 1542,2809 \\ 567,3478 \end{pmatrix}$$

Розв'язок даної системи наступний:

$$\bar{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = 10^7 \cdot (0,0272, -0,2071, 0,8779, -1,7866, 1,4059)$$

Тепер, на основі цього розв'язку, згідно з (8), можна отримати коефіцієнт при нульовій степені шаблону:  $\alpha_0 = 0,0068 \cdot 10^7$ .

Таким чином, ми отримали поліном Кунченка, який має наступний вигляд:

$$R_5(x) = \alpha_0 + \sum_{v=1}^5 \alpha_v \psi_1^v(x) = 10^7 \cdot (0,0068 + 0,0272\psi_1(x) - 0,2071\psi_1^2(x) + 0,8779\psi_1^3(x) - 1,7866\psi_1^4(x) + 1,4059\psi_1^5(x)).$$

Тепер ми можемо обчислити інфоркуну полінома Кунченка:

$$J_5 = \sum_{v=1}^5 \alpha_v F_{v,b} = 10^7 \cdot (0,0272 \cdot 32462,812 - 0,2071 \cdot 11899,5728 + 0,8779 \cdot 4251,7946 - 1,7866 \cdot 1542,2809 + 1,4059 \cdot 567,3478) = 1,9360 \cdot 10^9.$$

На основі цієї інфоркуни можна обчислити коефіцієнт ефективності (10):

$$e_5 = \frac{J_5}{\int_a^d (u_b(x) - \psi_b \cdot u_0(x))^2 dx} = \frac{1,9360 \cdot 10^9}{2,0665 \cdot 10^9} = 0,9368$$

Використовуючи вищенаведену схему, можна побудувати поліном Кунченка для кожної ділянки початкового сигналу довжиною в 31 відлік. При цьому, очевидно, буде отримано множину поліномів, носії яких частково перекриваються. Тоді, для сформування результуючого полінома потрібно залишити відліки тих поліномів, які було збудовано з більшою ефективністю.

Якщо застосувати наведену схему до всіх запропонованих шаблонів, то отримаємо відповідні ефектограми. На рис. 4, 5 і 6 зображено ефектограми, отримані за результатами побудови полінома Кунченка з застосуванням шаблонів  $\psi_1(x)$ ,  $\psi_2(x)$  та  $\psi_3(x)$  відповідно.

Співставлення рис. 4 та 5 показує, що шаблони  $\psi_1(x)$  і  $\psi_2(x)$  однаково сильно представлені в початковому сигналі на одній і тій же ділянці. З іншого боку, результати наближення сигналу  $s(x)$  третім шаблоном значно гірші (максимальна ефективність дорівнює 0,0442). Із цього можна зробити висновок, що в початковому сигналі дійсно мали місце певні особливості, схожі на перший та другий шаблони. І це вдалося віднайти за допомогою поліномів Кунченка.

Щоб порівняти поліноми Кунченка, отримані з застосуванням усіх трьох шаблонів, ми представили їх на рис. 7.

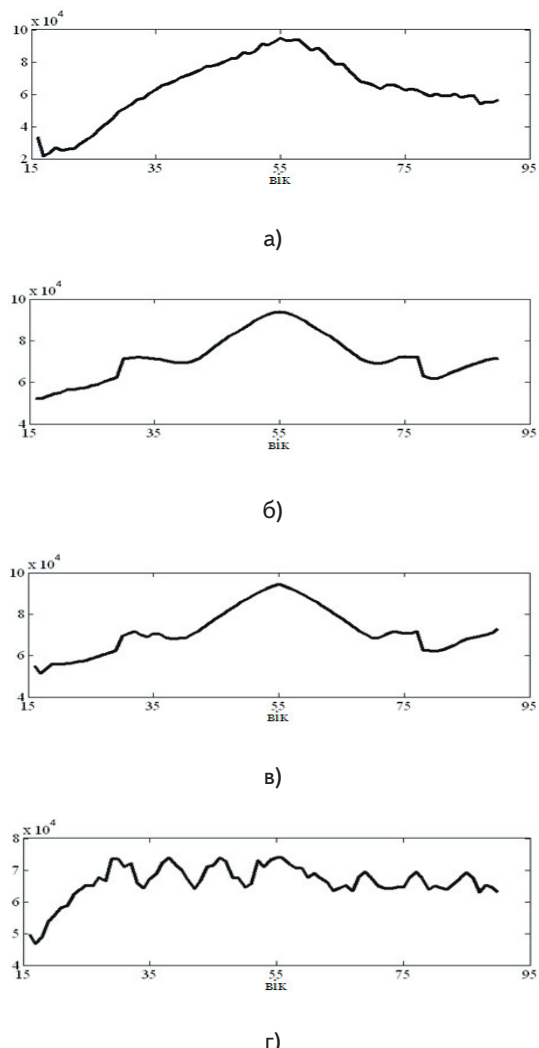


Рис. 7. Початковий сигнал (а) та поліноми Кунченка, отримані з застосуванням шаблону  $\psi_1(x)$  (б),  $\psi_2(x)$  (в) та  $\psi_3(x)$  (г)

Також, варто окремо зазначити, що при виборі шаблонів для побудови поліномів Кунченка основну роль грає форма та структура шаблону, тобто шаблони можна брати з точністю до деякого сталого множника та деякого сталого доданку. Так, зокрема, якщо взяти шаблон, що дорівнює  $\psi_4(x) = \sqrt{2\pi} \cdot \psi_1(x)$ , то нев'язка між відповідними ефектограмами має порядок 10-14. Більше того, у якості сталого множника можна взяти навіть  $-1$ . У такому випадку результати побудови полінома Кунченка будуть узагалі тотожні.

Це в першу чергу пояснюється тим, що основну роль під час пошуку коефіцієнтів відповідного полінома грає розв'язання системи лінійних рівнянь (7), розв'язок якої не змінюється від помноження обох частин системи на константу.

### Висновки

За результатами детального розгляду можливостей поліномів Кунченка по виявленню локальних осо-

бливостей у цифровому сигналі, та на основі аналізу результатів, отриманих на основі використання різних шаблонів, ми дійшли до наступних висновків:

- поліноми Кунченка дійсно є потужним інструментом по виявленню локальних структурних особливостей у цифрових сигналах, що можна з успіхом використовувати для пошуку заздалегідь визначених шаблонів у деякому масиві даних;
- поліноми Кунченка будуються на основі ефективного алгоритму, який легко застосовується на практиці; більше того, в якості шаблонів для поліномів Кунченка можна використовувати шаблони з точністю до сталого множника та сталого доданку;
- поліноми Кунченка можна ефективно застосовувати для аналізу статистичних даних, як було показано в даній роботі, зокрема, для визначення навіть із точністю до відліку ділянок сигналу, розподілених відповідно до деякої статистичної гіпотези;
- поліноми Кунченка не є універсальним механізмом; як було показано в даній статті, важливим етапом застосування поліномів Кунченка є вдалий підбір шаблону, а також коректний аналіз відповідної ефектограми.

### Література

1. Горелик А.Л. Методы распознавания / А.Л. Горелик, В.А. Скрипкин. – М.: Высшая школа, 1989. – 232 с.
2. Theodoridis S. Pattern Recognition / S. Theodoridis, K. Koutroumbas. – Academic Press, 2008. – 984 p.
3. Lin Y.-H. Template matching using the parametric template vector with translation, rotation and scale invariance / Y.-H. Lin, C.-H. Chen // Pattern Recognition. – 2008. – V.41, No.7. – P. 2413-2421.
4. Bellman R.E. Dynamic Programming / R.E. Bellman. – Princeton: University Press, 1957. – 366 p.
5. Damerau F.J. A technique for computer detection and correction of spelling errors / F.J. Damerau // Communications of the ACM. – 1964. – Vol. 7(3). – P. 171–176.
6. Briechle K. Template Matching Using Fast Normalized Cross Correlation / K. Briechle, U.D. Hanebeck // Proceedings of SPIE: Optical Pattern Recognition XII. – 2001. – Vol. 4387. – P. 95-102.
7. Lewis J.P. Fast Template Matching / J.P. Lewis // Vision Interface. – 1995. – P. 120-123.
8. Bhatia N. Survey of Nearest Neighbor Techniques / N. Bhatia, A. Vandana // International Journal of Computer Science and Information Security. – 2010. – Vol. 8, No. 2. – P. 302-305.
9. Кунченко Ю.П. Полиномы приближения в пространстве с порождающим элементом / Ю.П. Кунченко. – К.: Наукова думка, 2003. – 243 с.
10. Chertov O. Group Anonymity / O. Chertov, D. Tavrov // IPMU-2010 [(eds.) E. Hullermeier, R. Kruse, and F. Hoffmann]. – Heidelberg: Springer, 2010. – Communications in Computer and Information Science, vol. 81. – P. 592-601.
11. Чертов О.Р. Групповая анонимность данных / О.Р. Чертов // Зб. наук. праць Військового інституту телекомунікацій та інформатизації НТУУ «КПІ». – К.: ВІТІ НТУУ «КПІ», 2010. – Вип. 1. – С. 130-139.
12. U.S. Census 2000. 5-Percent Public Use Microdata Sample Files [Електронний ресурс]. – Режим доступу: <http://www.census.gov/Press-Release/www/2003/PUMS5.html>.