

ПЕРВИННЕ ТА ВТОРИННЕ ПЕРЕТВОРЕННЯ І ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ ПОТОКІВ В ІС

© Якушев В.С., 2002

The incoming data flow transformation and processing by primary and secondary information system elements, mathematical apparatus and cluster analysis methods are reviewed in this article.

Розглядаються перетворення і опрацювання вхідних потоків даних первинними і вторинними елементами ІС, математичний теоретико-числовий апарат, а також апарат кластерного аналізу.

Синтез, аналіз та дослідження інформаційних систем (ІС) належать до інформаційних технологій.

Інформаційні технології – це сукупність методів (математичних, статистичних тощо), засобів (апаратних, мікропрограмних, програмних та ін.), процесів, у тому числі виробничих, які об'єднані в деякий інформаційних ланцюг, що забезпечує збирання, перетворення, опрацювання, накопичення (бази даних і знань, сховища), прийняття рішень, виведення і передавання інформації [1,2].

Розглянемо вказаний інформаційний ланцюг, а саме – первинне та вторинне перетворення і опрацювання інформації, математичний апарат, який пропонується застосувати для підвищення точності, швидкодії і завадостійкості процесів, що реалізуються.

Процедура підготовки вхідних потоків даних, на основі яких відбувається фактичне формування та інформаційне наповнення відповідної моделі процесу чи явища, реалізується саме в ланках первинного та вторинного перетворення, опрацювання інформації і розглядається як адаптивна інтелектуальна компонента входу ІС.

У зв'язку з великою кількістю параметрів вхідних потоків даних, зокрема щодо матеріальних (фізичних) величин, їх, як правило, уніфікують за допомогою первинних перетворювачів. Оскільки перелік уніфікованих величин залишається все ж значним, в основу їх поділу доцільно покласти деякі узагальнюючі інформаційні ознаки S, D, P [3].

Фізичні величини, що підлягають, наприклад, вимірювальному перетворенню, змінюються і взаємно перетворюються у часі та в просторі. Тому така інформація може міститися або у фіксованому в часі та просторі рівні деякої величини (статична інформаційна ознака S), або у визначеному (невизначеному) локальному розподілі величин у часі (динамічна інформаційна ознака D), або у визначеному (невизначеному) локальному

розподілі величин у просторі (просторова інформаційна ознака P), а також у комбінації вказаних ознак.

Прийняті узагальнюючі інформаційні ознаки (S, D, P) , сукупність первинних та вторинних перетворень, що здійснюються над відповідними уніфікованими фізичними величинами, дають змогу послідовно і достатньо повно розглянути вказаний інформаційний ланцюг входу ІС.

Враховуючи прийняту послідовність перетворення і опрацювання матеріальних (фізичних) величин, за первинними, переважно неоднорідними перетвореннями (перетворювачами), розташовуються однорідні вторинні. У вторинних ланках здійснюється дискретизація, квантування і кодування фізичних величин, що характеризуються S, D, P -ознаками, а також їх подальше опрацювання, а саме, інтерполяція та екстраполяція, лінеаризація характеристик первинних перетворювачів, корекція похибок, кластерний аналіз, стиснення: структурне, статистичне, семантичне (змістове), прагматичне (цільове) тощо.

Зауважимо, що в загальній теорії систем досі не виділений єдиний об'єкт досліджень, а також не встановлені верхня і нижня границі системності [4]. Тому пропонується у визначенні об'єму загального поняття систем, зокрема ІС, враховувати аналіз тих суттєвих ознак якісно різних процесів у об'єктах, для яких системи є придатними, але не обов'язково адекватними, фізичними і математичними моделями.

Враховуючи таку методологічну постанову, виділимо об'єкт O , що досліджується, з оточуючого середовища C . Опишемо об'єкт O деяким набором інформаційних параметрів $X(x_1, \dots, x_n)$, а середовище C набором неінформаційних змінних $Q(q_1, \dots, q_m)$, де $x_i (i=1, \dots, n)$ та $q_j (j=1, \dots, m)$ – функції часових (динамічних) і просторових параметрів, відповідно, $D(d_1, \dots, d_\nu)$ та $P(p_1, \dots, p_\lambda)$, де $\nu=1, 2, 3, \dots$, $\lambda=1, 2, 3, \dots$. Окрім скінченного числа змінних у наборах X та Q , приймемо також, що кожна змінна характеризується скінченною кількістю станів або рівнів, тобто запишемо

$$X = X[x_1^{k_1}(d_\nu, p_\lambda), \dots, x_n^{k_n}(d_\nu, p_\lambda)];$$

$$Q = Q[q_1^{\varepsilon_1}(d_\nu, p_\lambda), \dots, q_m^{\varepsilon_m}(d_\nu, p_\lambda)],$$

де k_i – кількість станів змінної x_i , а ε_m – кількість станів змінної q_j .

Таке припущення відображає факт обмеженої подільності енергетичних і просторово-часових параметрів носіїв інформації, а також наявність верхньої границі швидкості зміни процесу або верхньої частоти у спектрі, що дозволяє відображати змінні x_i, q_j та їх подальше перетворення цілими аналітичними функціями [5].

Зазвичай, в ІС інформаційні сигнали (x_i) різної природи, зокрема різної фізичної модальності, надходять на первинні перетворювачі (давачі), де перетворюються в уніфіковані, переважно електричні, сигнали. Однак, оскільки первинні перетворювачі не ізольовані від середовища C , на них надходять також неінформаційні сигнали q_j .

Нехай кожен первинний перетворювач деякої ІС здійснює перетворення

$$Y_i = Y_i[d_\nu, p_\lambda, x_i^{k_i}(d_\nu, p_\lambda); q_1^{\varepsilon_1}(d_\nu, p_\lambda), \dots, q_m^{\varepsilon_m}(d_\nu, p_\lambda)]. \quad (1)$$

де залежність від d_i, p_i враховує динамічні властивості i -го давача, а залежність від q_i – вплив середовища S .

Подальше опрацювання $Y_i (i = 1, \dots, n)$ може бути здійснене у спільному каналі з часовим або частотним розподілом відповідних сигналів, або ж у паралельних каналах.

Другий варіант забезпечує найбільшу швидкість і надійність спрацювання, однак ставить жорсткіші вимоги до ступеня уніфікації і однорідності елементарних перетворювачів, а також збільшує апаратні витрати.

У загальному випадку змінні x_i об'єкта O є взаємно незалежними, у частковому випадку – корельованими.

Враховуючи сказане, розглянемо опрацювання результатів Y_i у вторинних ланках з метою отримання необхідної інформації про X_i . Як варіант, використаємо для цього спряжене розгортаюче опрацювання Y_i як порівняння з деякою однорідною мірою M [6]. Мірою, що є деякою розгортаючою функцією (сигналом) вигляду $Y_p^k(d, p) = G_k(d, p)$, де $k = 1, \dots, k_{\max}$ (k – кількість станів змінної X_i). При цьому

$$Y_i = G_k^{-1} \{g_i; d_i, p_i, x_i^k(d_i, p_i); q_i^1(d_i, p_i), \dots, q_i^{e_m}(d_i, p_i)\}, \quad (2)$$

де G_k^{-1} – функція, що обернена до G_k .

З рівності (2) видно, що якщо залежність g_i від d_i, p_i та q_i за час проходження цілої розгортки G_k^{-1} допустимо мала, а залежності g_i від x_i ($1 \leq i \leq n$) практично співпадають і дорівнюють у деякий момент y_i , то отримуємо лінеаризацію функції опрацювання вторинної ланки.

У цьому випадку

$$y_i = G_k^{-1} \{g_i; 0, 0, x_i^k(0, 0); 0, \dots, 0\}, \quad (3)$$

тобто лінеаризацію функції опрацювання отримуємо з похибкою апроксимації ідеальної міри $M(G_k^{-1})k \rightarrow \infty$ за допомогою реальної $M(G_k^{-1})k = 1, \dots, k_{\max}$. Існує також похибка, що зумовлена нерівностями $d_i \neq 0, p_i \neq 0, q_i \neq 0$. Перша похибка може бути зменшена за рахунок збільшення k , а друга – за рахунок введення корегуючого зворотного зв'язку та усереднення результатів опрацювання. Зауважимо, що за наявності декількох паралельних вторинних ланок опрацювання інформації, за умови їх уніфікації та однорідності, корегуючий зворотний зв'язок може бути єдиним для них всіх. Необхідне також виконання важливої умови, коли вплив змінних d_i, p_i, q_i за час корекції у паралельних ланках знаходиться у допустимих межах похибки.

На сучасному етапі для перетворення і попереднього опрацювання інформаційних потоків даних в основному використовуються системи числення з цілочисельною основою.

У цій роботі для розв'язання поставленої задачі пропонується використовувати відображення дійсних чисел (д.ч.) швидкозбіжними рядами Енгеля, Остроградського, комбінованими [1,2], а також апарату кластерного аналізу [7].

Ефективність сучасних алгоритмів (E) аналого-цифрового, цифроаналогового перетворення інформаційних потоків даних можна оцінити добутком показників точності (T), швидкодії ($Ш$) і завадостійкості ($З$): $E = T \cdot Ш \cdot З$.

Покращення одного з показників, як правило, здійснюється за рахунок інших. Окрім того, існує принципова границя, тобто $E_{\text{гп}}$, для відомих алгоритмів, що обмежує можливість їх застосування.

Запропоновані методи дозволяють збільшити $E_{\text{гп}}$ перетворення інформаційних потоків, а застосування апарату кластерного аналізу для попереднього опрацювання – зменшити навантаження на канали передачі вказаних потоків.

Стосовно аналого-цифрового, цифроаналогового перетворення теоретико-числові алгоритми Енгеля, Остроградського, комбіновані можуть бути поанані у вигляді таких рядів [2]:

$$\gamma_x = \gamma_{E1} = \left[C + 1 - \left(\frac{1}{g_1} + \sum_{v=1}^{\infty} \frac{1}{g_1 g_2 \dots g_{v+1}} \right) \right] X_0, \gamma_x = \gamma_{E2} = \left[C + 1 - \left(\frac{1}{g_1} + \sum_{v=1}^{\infty} \frac{1}{g_{v+1}} \right) \right] X_0, \quad (4)$$

$$\gamma_x = \gamma_{O1} = \left[C + \sum_{v=1}^n \frac{(-1)^{v-1}}{p_1 p_2 \dots p_v} + \frac{(-1)^n \alpha_n}{p_1 p_2 \dots p_n} \right] X_0, \gamma_x = \gamma_{O2} = \left[C + \sum_{v=1}^n \frac{(-1)^{v-1}}{p'_v} + \frac{(-1)^n \beta_n}{p'_1 p'_2 \dots p'_n} \right] X_0, \quad (5)$$

$$\gamma_x = \gamma_{K1} = \left[C + (1 - \xi_0) + \sum_{v=1}^n \frac{(-1)^{1+\xi_0+\xi_1+\dots+\xi_{v-1}}}{\lambda_1 \lambda_2 \dots \lambda_v} + \frac{(-1)^{1+\xi_0+\dots+\xi_n} k_n}{\lambda_1 \lambda_2 \dots \lambda_n} \right] X_0, \quad (6a)$$

$$\gamma_x = \gamma_{K2} = \left[C + (1 - \xi_0) + \sum_{v=1}^n \frac{(-1)^{1+\xi_0+\dots+\xi_{v-1}}}{\lambda'_v} + \frac{(-1)^{1+\xi_0+\dots+\xi_n} k'_n}{\lambda'_1 \lambda'_2 \dots \lambda'_n} \right] X_0, \quad (6b)$$

де д.ч. γ_x відповідної розмірності відображає істинне значення деякої фізичної величини X , що підлягає перетворенню, X_0 – розмір одиниці вказаної величини, а $g, g', p, p', \lambda, \lambda'$ – цілі числа або елементи ітераційного коду запропонованих відображень, тобто:

$$\begin{aligned} \gamma_{E1} &\equiv [X_0; g_1, g_2, g_3 \dots], \gamma_{E2} \equiv [X_0; g'_1, g'_2, g'_3 \dots], \\ \gamma_{O1} &\equiv [X_0; p_1, p_2, p_3 \dots], \gamma_{O2} \equiv [X_0; p'_1, p'_2, p'_3 \dots], \\ \gamma_{K1} &\equiv [X_0; \lambda_1, \lambda_2, \lambda_3 \dots], \gamma_{K2} \equiv [X_0; \lambda'_1, \lambda'_2, \lambda'_3 \dots]. \end{aligned} \quad (7)$$

На основі наведених виразів можуть бути розроблені методи і засоби перетворення електричних величин, наприклад розгортаючого і порозрядного врівноваження. При цьому алгоритми розгортаючого, в тому числі інтегруючого, аналого-цифрового, цифроаналогового

перетворення доцільно реалізувати відповідно до ліній 1, 2, 3 (рис.1,а), які є графічною інтерпретацією теоретико-числових алгоритмів Енгеля, Остроградського і комбінованого.

Вираз для ламаної кривої 3 (рис.1,а) за комбінованим алгоритмом має вигляд:

$$y_{k_2} = \begin{cases} (\lambda+1)x_{k_2-1} - 1, & \frac{1}{\lambda+1} < x_{k_2-1} < \frac{2}{2\lambda+1} \\ 1 - \lambda x_{k_2-1}, & \frac{2}{2\lambda+1} \leq x_{k_2-1} \leq \frac{1}{\lambda} \end{cases}, \quad (8)$$

де $x_{k_2-1} = A_{k_2-1} / X_0$; $y_{k_2} = A_{k_2} / X_0$, ($\lambda = 1, 2, 3, \dots$); A_{k_2} - елементи комбінованого алгоритму [2].

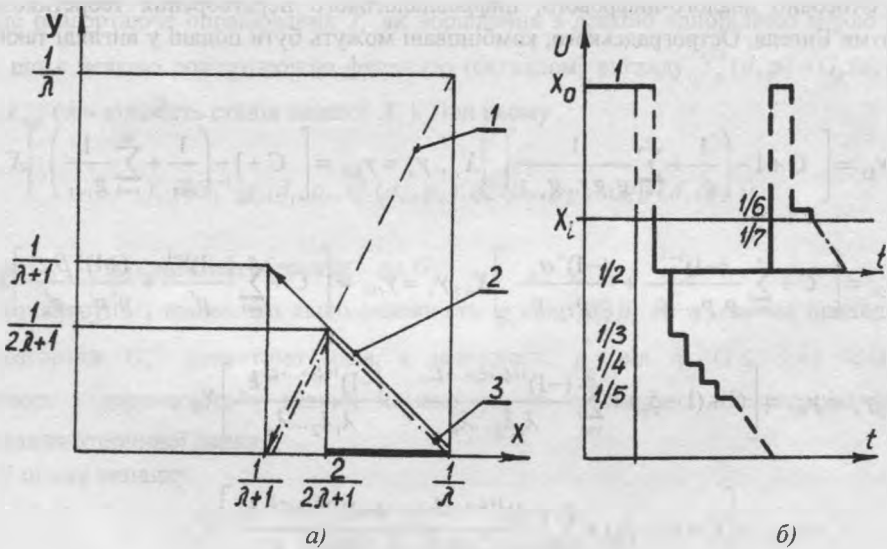


Рис.1 Графічна інтерпретація теоретико-числових алгоритмів Енгеля, Остроградського і комбінованого:

а) графічна інтерпретація алгоритмів Енгеля (1), Остроградського (2) і комбінованого (3);

б) приклад порозрядного врівноваження за другим алгоритмом Енгеля.

Розгортаюче (інтегруюче) перетворення за алгоритмами Енгеля, Остроградського, комбінованими алгоритмами при порівняно однакових витратах дає змогу збільшити завадозахищеність, контролюючи процес перетворення за такими співвідношеннями [2]:

$$g_1 \geq 2, g_v \geq g_{v-1}$$

для ряду Енгеля першого роду;

$$g'_1 \geq 2, g'_v \geq (g'_{v-1} - 1)g'_{v-1} + 1$$

для ряду Енгеля другого роду;

$$p_1 \geq 1, p_{v+1} > p_v$$

для ряду Остроградського першого роду;

- $p'_v \geq 1, p'_v \geq (p'_{v-1} + 1)p'_{v-1}$ для ряду Остроградського другого роду;
 $\lambda_1 > 2, \lambda_v > 2^v$ для комбінованого ряду першого роду,
 $\lambda'_1 > 2, \lambda'_v > 2^{2^{v-1}-1} (\lambda'_1)^{2^{v-1}}$ для комбінованого ряду другого роду.

Щодо КА (6а,6б) стосовно АЦП та ЦАП, то вони дозволяють отримати теоретично будь-яку розподільчу здатність за допомогою тільки однієї зразкової величини X_0 , а також абсолютну похибку, яка є меншою за мінімальне значення залишку порівняння. Роздільча здатність відповідних АЦП, ЦАП обмежується не кількістю і значеннями зразкових величин, як у відомих методах, а тільки порогом чутливості пристрою порівняння [8]. Останнє технічне обмеження спостерігається і у відомих методах, де під час досягнення найбільшої точності молодший ступінь квантування також близький за значенням до порогу чутливості пристрою порівняння; суттєвим мінусом є те, що абсолютна похибка вимірювального перетворення дорівнює значенню залишку порівняння.

Порозрядне врівноваження простіше реалізується другими алгоритмами Енгеля (АЕ), Остроградського (АО) та комбінованими (КА). На рис.16 показаний приклад порозрядного врівноваження відповідно до другого АЕ. Як результат, за елементами ітераційного коду $g'_1 = 1/2$ і $g'_2 = 1/7$, за X_0 і виразом (1) може бути обчислене значення X_i .

Перетворення за алгоритмами Енгеля, Остроградського та комбінованими алгоритмами при порівняльних апаратурних витратах дозволяє також підвищити швидкодію і завадостійкість, до того ж останню за рахунок усереднення і корелювання даних, отриманих у циклах порівняння. Ступінь еквівалентного усереднення випадкових завад у кілька разів більший, ніж при традиційному статистичному опрацюванні, що зумовлено знакозмінністю швидкозбіжних рядів Остроградського і комбінованих, а також тим, що елементи ітераційного коду (4) розташовані у знаменниках відповідних рядів (2) і (3). Реалізація запропонованих методів порівняно з відовими при зіставних метрологічних характеристиках може забезпечуватися елементами, які мають знижені вимоги до їх технічних характеристик. Останнє також зумовлено розміщенням елементів ітераційних кодів у знаменниках рядів (1), (2), (3). Тобто при підрахунках остаточного результату, наприклад, за комбінованим алгоритмом, здійснюється підсумовування елементів вигляду $\pm \frac{1}{\lambda_1 \lambda_2 \dots \lambda_n}$.

Стосовно до АЦП і ЦАП алгоритми Енгеля, Остроградського, комбіновані моделювались на ЕОМ. За результатами моделювання було зроблено ряд висновків. Подамо деякі з них для КА:

- абсолютна і відносна похибки зменшуються при вимірювальному перетворенні значення X_i , що міститься в деякому незмінному діапазоні вхідної величини X , за допомогою зразкової величини X_0 , значення якої збільшується відносно X ;
- для кожного незмінного діапазону значень X_i можна вчислити оптимальне значення X_0 , тобто знайти таке співвідношення X і X_0 , при якому буде здійснюватися найменша кількість ітерацій (порівнянь) і отримана найменша похибка відображення X_i ;

- відображення деякого X_i ітераційним кодом за КА дозволяє зменшити кількість його елементів ($\lambda_i, i = 1, 2, 3, \dots$) порівняно з відображенням за АЕ, АО і будь-яким кодом з постійною цілочисельною основою;
- при врівноваженні за КА порівняно з двійковим порозрядним врівноваженням [9,10] кількість порівнянь зменшується мінімум у три рази. Так, при двійковому порозрядному врівноваженні для досягнення похибок 0,1% і 0,01% треба відповідно 10 і 13 циклів порівняння:

$$0,1\% \approx 10 t_n;$$

$$0,01\% \approx 13 t_n,$$

де t_n – час циклу порівняння.

При врівноваженні за комбінованим алгоритмом

$$0,1\% \approx 2 t_n,$$

$$0,01\% \approx 3 t_n,$$

$$0,0001\% \approx 4 t_n.$$

Відзначимо, що перетворення інформації за КА дає змогу здійснювати відповідний процес практично одними й тими ж ланками АЦП-ЦАП в обох напрямках, тобто забезпечується зворотність як лінійних, так і функціональних перетворень. У цьому випадку вхід відповідного пристрою є одночасно і його виходом. Іншими словами, інформаційні потоки проходять в обох напрямках практично по тих самих елементах і ланцюгах перетворювача. Це необхідне при забезпеченні алгоритмічної і структурної уніфікації методів і засобів АЦП, ЦАП, їх однорідності та зменшення апаратних витрат, а також організації мультиплексного обміну інформацією у мережах зв'язку.

Під час опрацювання великих масивів цифрової інформації в реальному масштабі часу, у тому числі отриманої після аналого-цифрового перетворення, усуванні її надлишковості, стисненні і введенні в обчислювальні мережі пропонується використовувати ті ж теоретико-числові алгоритми (АЕ, АО, КА), а також апарат кластерного аналізу. Це дозволяє також класифікувати потік сигналів і здійснювати їх селективне опрацювання.

Як показали дослідження, для більшості завдань опрацювання сигналів з використанням вказаних алгоритмів при організації кластерів досить обмежитися порівнянням одного або двох перших елементів відповідних цифрових (ітераційних) кодів. Такий підхід значно скорочує кількість операцій при кластиризуванні сигналів відповідного масиву.

Якщо кластер утворюється, наприклад, за першим елементом ітераційного коду, то всередині кластера, утвореного за:

АЕ – найменше число має найменший другий елемент;

АО – найбільше число має найменший другий елемент;

КА – середнє за величиною число має другий елемент, який дорівнює або нулю, або значенню, найбільшому серед усіх решту.

Ці ж властивості зберігаються, якщо кластер утворюється за першими n знаками (1), (2), (3). Ступінь відмінності кодів, що входять в кластер, визначається за їх елементами, починаючи з перших нерівних, наприклад, $g \neq g', p \neq p', \lambda \neq \lambda'$, і характеризується функцією відстані (метрикою) $d(p, p'), d(g, g')$ або $d(\lambda, \lambda')$ [3].

Розв'язанням задачі кластерного аналізу є розбиття, що задовольняє деякий критерій оптимальності. Цей критерій може бути деяким функціоналом, що виражає рівні бажаності різних розбиттів і угруповань. Цей функціонал часто називають цільовою функцією. Наприклад, за цільову функцію може бути взята внутрішньогрупова сума квадратів відхилень.

Для прикладу

Не від'ємна дійснозначна функція $d(\gamma', \gamma'')$ називається функцією відстані (метрикою), якщо

- 1) $d(\gamma', \gamma'')=0$ для $\gamma'=\gamma''$;
- 2) $\gamma' \neq \gamma''$ для $0 < d(\gamma', \gamma'') \leq 1$;
- 3) $d(\gamma', \gamma'')=d(\gamma'', \gamma')$.

Нехай числам γ' і γ'' відповідають послідовності $\gamma'=(\lambda'_0, \lambda'_1, \dots, \lambda'_k)$ і $\gamma''=(\lambda''_0, \lambda''_1, \dots, \lambda''_k)$. Якщо ці послідовності однакові, то $\gamma'=\gamma''$ і $d(\gamma', \gamma'')=0$. Якщо $\lambda'_0=\lambda''_0, \lambda'_1=\lambda''_1, \dots, \lambda'_{k-1}=\lambda''_{k-1}$, а $\lambda'_k \neq \lambda''_k$, тоді можемо визначити

$$d(\gamma', \gamma'') = \frac{1}{(\lambda'_k - \lambda''_k)^2}.$$

Далі формується цільова функція, яка є критерієм стиснення інформації.

Як результат, здійснюється класифікація, стиснення інформації або селекція потрібного сигналу (коду) з кластера, усувається інформаційна надлишковість і підвищується ефективність цифрового опрацювання інформації в реальному масштабі часу.

1. Якушев В.С. Математичний підхід до розроблення нових методів і засобів перетворення та опрацювання інформації// Вісн. ДУ "ЛП".-1998, № 330.- С.263-268.
2. Якушев В.С. Використання активних алгоритмів перетворення і опрацювання інформації в інформаційних технологіях// Вісн. ДУ "ЛП".-1999.- № 383.-С.242-260.
3. Мельничук Ю.В., Раков М.А., Якушев В.С. Аналого-цифровые преобразователи с переменной значностью. - К.: Наукова думка, 1988. -128 С. 4. Левченко В.С., Пропой А.И. О математической теории систем. Математические методы в теории систем.-М.: Сб. трудов. Всесоюзный НИИ системных исследований.-1988.- Вып.6.-С.5-14. 5. Хургии Я.И., Яковлев В.П. Методы теории целых функций в радиофизике, теории связи и оптике.-М.: Физматгиз, 1962.-135 С. 6. Темников Ф.Е., Славинский В.Л. Математические развертывающие системы.- М.: Энергия, 1970.-250 С. 7. Дюран Ю., Оделл П. Кластерный анализ.-М.: Статистика, 1977.-128 С. 8. Якушев В.С. Дослідження ефективності активних алгоритмів при перетворенні і опрацюванні інформаційних потоків// Вісн. НУ "ЛП".-2001.-№ 438.- С.174-178. 9. Орнатский П.П. Автоматические измерения и приборы.-К. Вища школа, 1980.-558 С. 10. Орнатский П.П. Теоретические основы информационно-измерительной техники.- К.: Вища школа, 1983.- 456 С.