

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ДНІПРОВСЬКА ПОЛІТЕХНІКА»



В. М. Горєв, В. І. Корнієнко, О. Ю. Гусєв

**АДАПТИВНА ФІЛЬТРАЦІЯ:
ТЕОРІЯ ТА ЗАСТОСУВАННЯ**

Навчальний посібник

Дніпро
Журфонд
2025

УДК 004.021
Г 67

*Рекомендовано вченою радою НТУ «Дніпровська політехніка»
як навчальний посібник для здобувачів ступеня магістра
спеціальності 172 Електронні комунікації та радіотехніка
(протокол № 10 від 10.09.2024)*

Рецензенти:

О. Й. Соколовський – д-р фіз.-мат. наук, проф. (Дніпровський національний університет імені Олеся Гончара);

С. В. Толюпа – д-р техн. наук, проф. (Київський національний університет імені Тараса Шевченка).

Горєв В. М.

Г67 Адаптивна фільтрація: теорія та застосування : навч. посіб. /
В. М. Горєв, В. І. Корнієнко, О. Ю. Гусєв ; М-во освіти і науки України,
Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : Журфонд, 2025. –
146 с.

Викладено основи теорії адаптивної фільтрації даних із застосуванням відомих засобів – фільтра Калмана та фільтрів, побудованих на RLS- і LMS-алгоритмах, а також тих, що створені штучним інтелектом (нейронних мережах, системах нечіткого висновку тощо).

Розглянуто застосування адаптивних фільтрів до розв'язування задач спостереження, ідентифікації, фільтрації голосу, аналізу й моделювання трафіку телекомунікаційних систем та прогнозування технологічних змінних. Набуті знання дають можливість засвоїти принципи побудови та сферу застосування адаптивних фільтрів.

Посібник призначено для здобувачів ступеня магістра спеціальності 172 Електронні комунікації та радіотехніка. Також він може бути корисним студентам інших технічних або фізико-математичних спеціальностей.

УДК 004.021

ISBN 978-966-934-634-6

© В. М. Горєв, В. І. Корнієнко, О. Ю. Гусєв, 2025

© НТУ «Дніпровська політехніка», 2025

© Журфонд, 2025

Зміст

Перелік умовних скорочень	6
ВСТУП.....	7
Глава 1. Основи теорії фільтрації	9
1.1. Випадкові процеси та кореляційні функції	9
1.2. Лінійні стаціонарні фільтри	15
1.3. Фільтр Колмогорова – Вінера.....	17
1.4. Градієнтні методи пошуку мінімуму функції багатьох змінних	23
1.5. Застосування фільтра Колмогорова – Вінера до прогнозування стаціонарних процесів з важкими хвостами (з КФСРЕ)	30
1.6. Контрольні питання до глави 1	40
Глава 2. Адаптивні фільтри	42
2.1. Скалярний дискретний фільтр Калмана	42
2.2. Векторний дискретний фільтр Калмана	46
2.3. Скалярний фільтр Калмана – Бюсі.....	55
2.4. LMS-алгоритми в теорії фільтрації	60
2.4.1. LMS-алгоритм вилучення корисного сигналу, змішаного з шумом... ..	60
2.4.2. LMS-алгоритм вилучення інтерференції	63
2.4.3. LMS-алгоритм лінійного прогнозування часового ряду	66
2.5. RLS-алгоритм виділення корисного сигналу, змішаного з шумом	67
2.6. Вимоги до адаптивних фільтрів	71
2.6.1. Застосування	72
2.6.2. Структури адаптивних фільтрів.....	72
2.6.3. Алгоритми	76
2.7. Критерії функціонування адаптивних алгоритмів	78
2.8. Контрольні питання до глави 2	80
Глава 3. Адаптивні фільтри-апроксиматори в розв’язуванні задач спостереження та ідентифікації	82
3.1. Дискретизація та відновлення сигналів у розв’язуванні задач спостереження та ідентифікації.....	83

3.1.1. Дискретизація сигналів обмеженого спектра.....	86
3.1.2. Дискретизація сигналів необмеженого спектра.....	87
3.1.3. Відновлення сигналів.....	89
3.2. Адаптивні фільтри-апроксиматори.....	93
3.2.1. Лінійний адаптивний фільтр-апроксиматор.....	95
3.2.2. Нейромережевий адаптивний фільтр-апроксиматор.....	96
3.2.3. Нейронечіткий адаптивний фільтр-апроксиматор.....	97
3.2.4. Нечіткий адаптивний фільтр-апроксиматор з субтрактивною кластеризацією.....	98
3.2.5. Нечіткий адаптивний фільтр-апроксиматор з кластеризацією С- середніх.....	99
3.2.6. Нейронний вейвлет-адаптивний фільтр-апроксиматор	101
3.2.7. Вейвнет АФА з блоком, який має нескінченну імпульсну характеристику	103
3.3. Контрольні питання до глави 3	105
Глава 4. Застосування адаптивних фільтрів-апроксиматорів до розв'язування прикладних задач.....	107
4.1. Адаптивне оцінювання та ідентифікація процесів на основі вейвнетів .	107
4.1.1. Фільтрація голосу.....	107
4.1.2. Фільтрація голосу за наявності імпульсних перешкод	109
4.1.3. Апроксимація складних функцій.....	110
4.1.4. Ідентифікація нелінійного динамічного процесу за наявності нормально розподіленого шуму.....	112
4.2. Придушення шумів у сигналах мобільного зв'язку з використанням нейронечіткої ідентифікації тракту передачі.....	114
4.3. Аналіз і моделювання мережевого трафіку	117
4.4. Мультифрактальний аналіз та моделювання мережевого трафіку	125
4.5. Прогнозування вмісту кремнію в чавуні під час його виплавлення	135
4.6. Прогнозування теплового стану доменної печі та характеристик якості чавуну.....	137

4.7. Контрольні запитання до глави 4	142
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ	143

Перелік умовних скорочень

АДФ – аналіз детрендових флуктуацій
АКФ – автокореляційна функція
АТС – автоматична телефонна станція
АФА – адаптивний фільтр-апроксиматор
ДВП – дискретне вейвлет-перетворення
ЗДВП – зворотне дискретне вейвлет-перетворення
ІХ – імпульсна характеристика
КІХ – кінцева імпульсна характеристика
ЛДЛ – лінійний динамічний блок
МВМ – мультифрактальна вейвлет-модель
НВ – нейровейвлетний (фільтр)
НІХ – нескінченна імпульсна характеристика
НМ (NN) – нейронна мережа
НМПП – нейронна мережа прямого поширення
НСЛ – нелінійний статичний блок
ОК – об'єкт керування
ПСЗ – повільно спадна залежність
РБФ – радіальна базисна функція
СНС – сигнал необмеженого спектра
СОС – сигнал обмеженого спектра
ФНЧ – фільтр нижніх частот
ШСЗ – швидко спадна залежність
Anfis – адаптивна нейронна система нечіткого висновку
Genfis2 – алгоритм субтрактивної кластеризації
Genfis3 – алгоритм кластеризації C-середніх
КФСРЕ – кореляційна функція, що спадає повільніше за експоненційну

ВСТУП

Дисципліна «Теорія адаптивної фільтрації» – це вибірковий освітній компонент у підготовці здобувачів спеціальності 172 «Електронні комунікації та радіотехніка» другого (магістерського) рівня.

У посібнику викладено основи теорії фільтрації, охарактеризовано найпростіші та часто вживані адаптивні фільтри. Перш за все, це фільтр Калмана і ті засоби, робота яких базується на RLS- та LMS-алгоритмах. При цьому в посібнику відображено нещодавні наукові результати досліджень, виконаних його авторами. Зокрема, було вивчено застосування фільтра Колмогорова – Вінера до прогнозування гладких процесів з важкими хвостами, кореляційна функція яких спадає повільніше за експоненційну, які спостерігаються в телекомунікаційному трафіку. До того ж чималу увагу приділено використанню адаптивних фільтрів-апроксиматорів у розв'язуванні задач спостереження та ідентифікації. Так, у главі 4 описано багато практичних прикладів застосування таких фільтрів, які дають уявлення здобувачам про принципи побудови цих засобів.

Розділи 1.1 – 1.4 та 2.1 – 2.5 посібника написані доцентом В. М. Горєвим, глави 3, 4 – професором В. І. Корнієнком, розділи 2.6 і 2.7 – професором О. Ю. Гусєвим. Розділ 1.5 написано авторами спільно на основі результатів нещодавно проведених наукових досліджень.

Опанування матеріалу посібника дозволяє здобувачам освіти засвоїти необхідну інформацію про телекомунікації та радіотехніку в частині знання процесів адаптивної фільтрації.

Посібник видано в рамках навчально-методичної теми Ш-518 «Розробка методичного забезпечення за дисциплінами, що викладаються кафедрою фізики НТУ «Дніпровська політехніка», та наукової теми Е-345 «Інтелектуальні методи моделювання процесів в інформаційно-телекомунікаційних системах критичної інфраструктури».

Згідно з робочою програмою дисципліни «Теорія адаптивної фільтрації» визначено такі очікувані результати навчання:

- уміння використовувати фундаментальні знання в галузі теорії адаптивної фільтрації, володіти математичним апаратом;
- здатність використовувати сучасні математичні пакети для програмування адаптивних фільтрів.

Глава 1. Основи теорії фільтрації

1.1. Випадкові процеси та кореляційні функції

Випадковою величиною називається величина яка може приймати різні значення з різними ймовірностями. Функція $x(t)$ дійсної змінної t називається *випадковою функцією*, якщо при кожному значення змінної t $x(t)$ є випадковою величиною. Якщо змінна t має фізичний зміст часу, то випадкову функцію $x(t)$ називають *випадковим процесом*.

Одновимірною густиною ймовірності випадкового процесу $x(t)$ є величиною $\omega_1(x_1, t_1)$, такою що $\omega_1(x_1, t_1)dx_1$ є ймовірністю того, що випадкова величина $x(t_1)$ набуватиме значення в діапазоні від x_1 до $x_1 + dx_1$, тобто

$$\omega_1(x_1, t_1)dx_1 = P(x(t_1) \in (x_1, x_1 + dx_1)). \quad (1.1)$$

Двовимірною густиною ймовірності випадкового процесу $x(t)$ є величиною $\omega_2(x_1, t_1; x_2, t_2)$, такою що $\omega_2(x_1, t_1; x_2, t_2)dx_1dx_2$ є ймовірністю того, що $x(t_1) \in (x_1, x_1 + dx_1)$ та в той же час $x(t_2) \in (x_2, x_2 + dx_2)$:

$$\omega_2(x_1, t_1; x_2, t_2)dx_1dx_2 = P(x(t_1) \in (x_1, x_1 + dx_1), x(t_2) \in (x_2, x_2 + dx_2)). \quad (1.2)$$

З виразу (1.2) очевидно, що

$$\omega_2(x_1, t_1; x_2, t_2) = \omega_2(x_2, t_2; x_1, t_1). \quad (1.3)$$

Аналогічно, n -вимірною густиною ймовірності випадкового процесу $x(t)$ є величиною $\omega_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n)$, такою що

$$\begin{aligned} \omega_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n)dx_1dx_2\dots dx_n = \\ = P(x(t_1) \in (x_1, x_1 + dx_1), x(t_2) \in (x_2, x_2 + dx_2), \dots, x(t_n) \in (x_n, x_n + dx_n)). \end{aligned} \quad (1.4)$$

Густина ймовірності задовольняє умові нормування

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} dx_1 dx_2 \dots dx_n \omega_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = 1, \quad (1.5)$$

фізичним змістом якої є те, що сума всіх можливих ймовірностей дорівнює 1.

Взагалі кажучи, випадковий процес описується густинами ймовірностей всіх розмірностей. Однак, працювати з багатовимірними густинами ймовірностей дуже важко. Тому випадковий процес часто описується лише одновимірною та двовимірною густинами ймовірностей. Більш того, багато випадкових процесів строго описуються лише одновимірною та двовимірною густинами ймовірностей, а густини ймовірностей більших розмірностей можуть бути виражені в термінах одновимірної та двовимірної густин.

Математичне очікування (середнє значення) випадкового процесу $x(t)$ є величиною $\langle x(t) \rangle$:

$$\langle x(t) \rangle = m_x(t) = \int_{-\infty}^{+\infty} dx x \omega_1(x, t), \quad (1.6)$$

Середнє значення від функції від випадкового процесу за визначенням обчислюється як

$$\langle f(x(t)) \rangle = \int_{-\infty}^{+\infty} dx f(x) \omega_1(x, t). \quad (1.7)$$

Наприклад, середньоквадратичне значення випадкового процесу згідно (1.7) обчислюється так:

$$\langle x^2(t) \rangle = \int_{-\infty}^{+\infty} dx x^2 \omega_1(x, t). \quad (1.8)$$

Слід зауважити, що якщо значення випадкового процесу $x(t)$ взято в різні моменти часу, то середнє значення відповідного добутку виражається в термінах двовимірної густини ймовірності, а не одновірної:

$$\langle x(t_1)x(t_2) \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, t_1; x_2, t_2). \quad (1.9)$$

Дисперсія $D_x(t)$ та середньоквадратичне відхилення $\sigma_x(t)$ випадкового процесу $x(t)$ визначаються як

$$D_x(t) = \langle x^2(t) \rangle - \langle x(t) \rangle^2, \quad \sigma(t) = \sqrt{D_x(t)}. \quad (1.10)$$

В літературі поняття кореляційної функції вводиться по-різному. В ряді джерел кореляційна функція $R_x(t_1, t_2)$ випадкового процесу $x(t)$ визначається як $R_x(t_1, t_2) = \langle x(t_1)x(t_2) \rangle$, в інших – як $R_x(t_1, t_2) = \langle x(t_1)x(t_2) \rangle - \langle x(t_1) \rangle \langle x(t_2) \rangle$. Слід зауважити, що якщо середнє значення $\langle x(t) \rangle = 0$ для кожного моменту часу t , то ці визначення співпадають, проте вони не співпадають в загальному випадку. Для визначеності, визначимо *кореляційну функцію* випадкового процесу $x(t)$ як

$$R_x(t_1, t_2) = \langle x(t_1)x(t_2) \rangle. \quad (1.11)$$

Випадковий процес називається *стаціонарним*, для густин ймовірності всіх розмірностей виконується така властивість:

$$\forall \tau \quad \omega_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = \omega_n(x_1, t_1 + \tau; x_2, t_2 + \tau; \dots; x_n, t_n + \tau), \quad (1.12)$$

тобто якщо додати одну й ту ж величину до всіх часових аргументів густини ймовірності, то вона не зміниться. Такі процеси ще називають *стаціонарними у вузькому сенсі*.

Однак, як зауважено вище, для простоти випадковий процес часто описується в термінах лише одновимірної та двовимірної густин ймовірностей. Випадковий процес називається *стаціонарним в широкому сенсі*, якщо властивість (1.12) є справедливою лише для $n = 1, 2$. Будь-який стаціонарний у вузькому сенсі процес є стаціонарним і в широкому сенсі, але зворотне твердження не є вірним.

Покажемо, що *середнє значення стаціонарного випадкового процесу не залежить від часу*. Давайме розглянемо середні значення в два довільні різні моменти часу $m_x(t_1)$ та $m_x(t_2)$. Згідно визначення (1.6)

$$m_x(t_1) = \int_{-\infty}^{+\infty} dx x \omega_1(x, t_1), \quad m_x(t_2) = \int_{-\infty}^{+\infty} dx x \omega_1(x, t_2). \quad (1.13)$$

Обираючи $\tau = t_2 - t_1$ у виразі (1.12), матимемо

$$\omega_1(x_1, t_1) = \omega_1(x_1, t_1 + (t_2 - t_1)) = \omega_1(x_1, t_2), \quad (1.14)$$

на основі (1.13) та (1.14) бачимо, що

$$m_x(t_1) = m_x(t_2), \quad (1.15)$$

Тож середнє значення стаціонарного процесу однакове для двох довільних різних моментів часу, тож воно не залежить від часу.

Тепер покажемо, що *кореляційна функція* $R_x(t_1, t_2)$ *стаціонарного процесу* $x(t)$ *залежить лише від різниці часових аргументів* $t_2 - t_1$. На основі (1.11) та (1.9) матимемо

$$R_x(t_1, t_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, t_1; x_2, t_2), \quad (1.16)$$

Обираючи $\tau = -t_2$ у виразі (1.12), отримаємо

$$\omega_2(x_1, t_1; x_2, t_2) = \omega_2(x_1, t_1 - t_2; x_2, 0). \quad (1.17)$$

На основі (1.16) та (1.17) матимемо

$$R_x(t_1, t_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, t_1 - t_2; x_2, 0) = R_x(t_1 - t_2), \quad (1.18)$$

що і треба було довести. Бачимо, що *кореляційна функція випадкового стаціонарного процесу є функцією одного аргументу, а не двох*; на основі (1.18) бачимо, що

$$R_x(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, \tau; x_2, 0), \quad (1.19)$$

Обираючи $\tau = t$ в (1.12), матимемо

$$\omega_2(x_1, \tau; x_2, 0) = \omega_2(x_1, \tau + t; x_2, t), \quad (1.20)$$

звідки з урахуванням (1.9)

$$R_x(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, \tau + t; x_2, t) = \langle x(t + \tau) x(t) \rangle, \quad (1.21)$$

результат (1.21) є широкоживаним!

Покажемо, що *кореляційна функція стаціонарного випадкового процесу є парною функцією*. На основі (1.19) маємо

$$R_x(-\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, -\tau; x_2, 0). \quad (1.22)$$

На основі (1.12) маємо

$$\omega_2(x_1, -\tau; x_2, 0) = \omega_2(x_1, -\tau + \tau; x_2, 0 + \tau) = \omega_2(x_1, 0; x_2, \tau), \quad (1.23)$$

що з урахуванням (1.22) дає

$$\begin{aligned} R_x(-\tau) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, 0; x_2, \tau) = \{x_1 \leftrightarrow x_2\} = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_2 dx_1 x_2 x_1 \omega_2(x_2, 0; x_1, \tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_2, 0; x_1, \tau). \end{aligned} \quad (1.24)$$

На основі (1.24) та (1.3) отримаємо

$$R_x(-\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, \tau; x_2, 0) = R_x(\tau), \quad (1.25)$$

що і треба було довести.

Взаємною кореляційною функцією двох випадкових процесів $x(t)$ та $y(t)$ є функція $R_{xy}(t_1, t_2)$, яка визначається таким чином:

$$R_{xy}(t_1, t_2) = \langle x(t_1) y(t_2) \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy x y \omega(x, t_1; y, t_2), \quad (1.26)$$

де $\omega(x_1, t_1; y_2, t_2) dx_1 dy_2$ є ймовірністю того, що $x(t_1) \in (x_1, x_1 + dx_1)$ та одночасно $y(t_2) \in (y_2, y_2 + dy_2)$. Очевидно, що

$$\omega(x, t_1; y, t_2) = \omega(y, t_2; x, t_1). \quad (1.27)$$

Слід зазначити, що в літературі взаємна кореляційна функція не завжди визначається згідно (1.26). Визначення $R_{xy}(t_1, t_2) = \langle x(t_1) y(t_2) \rangle - \langle x(t_1) \rangle \langle y(t_2) \rangle$ також часто використовується, але для визначеності використовуватимемо визначення (1.26). Ці два визначення співпадають, якщо $x(t)$ або $y(t)$ є стаціонарним випадковим процесом з нульовим середнім значенням.

Якщо обидва процеси $x(t)$ та $y(t)$ є стаціонарними, то часто припускають, що $\omega(x, t_1; y, t_2)$ має властивість

$$\forall \tau \quad \omega(x, t_1 + \tau; y, t_2 + \tau) = \omega(x, t_1; y, t_2). \quad (1.28)$$

В такому випадку, взаємна кореляційна функція двох стаціонарних процесів має такі властивості:

$$R_{xy}(t_1, t_2) = R_{xy}(t_1 - t_2), \quad R_{xy}(\tau) = \langle x(t + \tau)y(t) \rangle, \quad R_{xy}(\tau) = R_{yx}(-\tau). \quad (1.29)$$

Доведемо їх. На основі (1.28) маємо

$$\omega(x, t_1; y, t_2) = \omega(x, t_1 - t_2; y, 0). \quad (1.30)$$

З (1.30) та (1.26) бачимо, що

$$R_{xy}(t_1, t_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy xy \omega(x, t_1 - t_2; y, 0) = R_{xy}(t_1 - t_2), \quad (1.31)$$

Тож перша властивість з (1.29) доведена. Як бачимо з (1.31), взаємна кореляційна функція двох стаціонарних процесів є функцією одного аргументу, а не двох:

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy xy \omega(x, \tau; y, 0). \quad (1.32)$$

На основі (1.28) маємо

$$\omega(x, \tau; y, 0) = \omega(x, t + \tau; y, t), \quad (1.33)$$

що з урахуванням (1.32) дає

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy xy \omega(x, t + \tau; y, t) = \langle x(t + \tau)y(t) \rangle, \quad (1.34)$$

тож щойно довели другу властивість з (1.29). На основі (1.32) бачимо, що

$$R_{yx}(-\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy xy \omega(y, -\tau; x, 0). \quad (1.35)$$

З (1.28) та (1.27) маємо

$$\omega(y, -\tau; x, 0) = \omega(y, -\tau + \tau; x, \tau) = \omega(y, 0; x, \tau) = \omega(x, \tau; y, 0), \quad (1.36)$$

тож з урахуванням (1.36), (1.35) та (1.32) отримаємо

$$R_{xy}(\tau) = R_{yx}(-\tau). \quad (1.37)$$

Тож третю властивість з (1.29) доведено.

Слід зауважити, що вище наведені визначення середніх величин є визначеннями *середнього за множиною*. Визначення *середнього за часом* від стаціонарного випадкового процесу $x(t)$ задається таким чином:

$$\langle x(t) \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} dx(t). \quad (1.38)$$

Фактично, визначення (1.38) ідентичне до відомого з математичного аналізу визначення середнього значення функції на всій числовій осі. Визначення (1.38) є коректним, якщо існує відповідний ліміт.

Для стаціонарних процесів вводять поняття часової кореляційної функції $\Re_x(\tau)$ та часової взаємної кореляційної функції $\Re_{xy}(\tau)$ на основі (1.21) та (1.34) :

$$\Re_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} dx(t)x(t+\tau), \quad \Re_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} dx(t+\tau)y(t), \quad (1.39)$$

відповідні визначення є коректними, якщо існують відповідні ліміти.

Слід зауважити, що середні за часом можуть бути різними для різних реалізацій випадкового процесу. Більше того, середнє за часом може не співпадати з середнім за множиною. Однак існує доволі широкий клас стаціонарних випадкових процесів, які називають *ергодичними* процесами. Для ергодичних процесів середнє за часом співпадає з середнім за множиною. Для ергодичних процесів часова кореляційна функція співпадає з кореляційною функцією, та часова взаємна кореляційна функція співпадає з взаємною кореляційною функцією, тобто *для ергодичних процесів*

$$\langle x(t) \rangle_t = \langle x(t) \rangle, \quad \Re_x(\tau) = R_x(\tau), \quad \Re_{xy}(\tau) = R_{xy}(\tau). \quad (1.40)$$

1.2. Лінійні стаціонарні фільтри

Під фільтром розумітимемо фізичний пристрій або програму, що виділяє корисний сигнал з суміші корисного сигналу та шуму або робить прогнозування майбутніх значень корисного сигналу.

Припустимо, що при вхідному сигналі $x_1(t)$ фільтр генерує на виході вихідний сигнал $y_1(t)$, та при вхідному сигналі $x_2(t)$ фільтр генерує вихідний сигнал $y_2(t)$. Нехай вхідний сигнал є лінійною комбінацією сигналів $x_1(t)$ та $x_2(t)$: $x(t) = Ax_1(t) + Bx_2(t)$, де A , B є константами. Тоді фільтр є *лінійним*, якщо його вихідний сигнал є лінійною комбінацією відповідних вихідних сигналів $y_1(t)$ та $y_2(t)$ з тими ж самими константами A та B : $y(t) = Ay_1(t) + By_2(t)$.

Припустимо, що маємо деяку часову затримку вхідного сигналу. Тоді фільтр є *стаціонарним*, якщо його вихідний сигнал має точно таку ж часову затримку. Іншими словами, припустимо що при вхідному сигналі $x(t)$ фільтр генерує вихідний сигнал $y(t)$. Тоді якщо фільтр є стаціонарним, то при вхідному сигналі $x(t - \Delta t)$ він генеруватиме вихідний сигнал $y(t - \Delta t)$.

Для простоти розглянемо аналогові сигнали, які задані на всій числовій осі при $t \in (-\infty, \infty)$. Позначимо вхідний та вихідний сигнали фільтра відповідно як $x(t)$ та $y(t)$. Тоді за визначенням *вагова функція* фільтра є функцією $h(t)$ такою, що

$$y(t) = \int_{-\infty}^{+\infty} d\tau h(\tau) x(t - \tau). \quad (1.41)$$

Покажемо, чому фільтр, описаний співвідношенням (1.41) є лінійним та стаціонарним. Нехай при вхідному сигналі $x_1(t)$ фільтр генерує вихідний сигнал $y_1(t)$, та при вхідному сигналі $x_2(t)$ фільтр генерує вихідний сигнал $y_2(t)$. Тоді згідно (1.41)

$$y_1(t) = \int_{-\infty}^{+\infty} d\tau h(\tau) x_1(t - \tau), \quad y_2(t) = \int_{-\infty}^{+\infty} d\tau h(\tau) x_2(t - \tau). \quad (1.42)$$

Нехай вхідним сигналом фільтра є сигнал

$$X(t) = Ax_1(t) + Bx_2(t), \quad (1.43)$$

де A, B є константами. Обчислимо відповідний вихідний сигнал фільтра $Y(t)$.

Згідно (1.41) – (1.43) бачимо, що

$$\begin{aligned} Y(t) &= \int_{-\infty}^{+\infty} d\tau h(\tau) X(t-\tau) = A \int_{-\infty}^{+\infty} d\tau h(\tau) x_1(t-\tau) + B \int_{-\infty}^{+\infty} d\tau h(\tau) x_2(t-\tau) = \\ &= Ay_1(t) + By_2(t), \end{aligned} \quad (1.44)$$

тож доведено, що такий фільтр є лінійним.

На основі (1.41) очевидно, що

$$y(t - \Delta t) = \int_{-\infty}^{+\infty} d\tau h(\tau) x(t - \Delta t - \tau), \quad (1.45)$$

тож такий фільтр є стаціонарним.

Вагова функція лінійного стаціонарного фільтра має властивість

$$h(t < 0) = 0. \quad (1.46)$$

Ця властивість є наслідком *принципу причинності*, згідно якого зміна сигналу на виході не може передувати зміні сигналу на вході. Для простоти розглянемо приклад, коли вхідний сигнал фільтра є дельта-функцією

$$x(t) = \delta(t - t_0), \quad (1.47)$$

тож вхідний сигнал ненульовий лише в момент часу $t = t_0$, та згідно принципу причинності

$$y(t < t_0) = 0. \quad (1.48)$$

На основі (1.41) та (1.47) матимемо

$$y(t) = \int_{-\infty}^{+\infty} d\tau h(\tau) \delta(t - t_0 - \tau) = h(t - t_0), \quad (1.49)$$

що згідно (1.48) дає

$$h(t - t_0) = 0 \text{ при } t < t_0, \quad (1.50)$$

що зумовлює властивість (1.46).

1.3. Фільтр Колмогорова – Вінера

Нехай вхідний сигнал $x(t)$ є сумою корисного сигналу $s(t)$ та шуму $n(t)$:

$$x(t) = n(t) + s(t), \quad (1.51)$$

$s(t)$ та $n(t)$ є стаціонарними процесами. Взаємна кореляційна функція $R_{sn}(t)$ процесів $s(t)$ та $n(t)$ вважається стаціонарною. Для простоти розглянемо аналогові сигнали, визначені для часів $t \in (-\infty, +\infty)$, всі процеси вважаються ергодичними. *Вважається, що корисний сигнал та шум є невідомими, проте взаємна кореляційна функція $R_{sx}(t)$ процесів $s(t)$ та $x(t)$, та кореляційна функція $R_x(t)$ процесу $x(t)$ є відомою.*

Фільтр Колмогорова–Вінера є лінійним стаціонарним фільтром, який може як виділяти корисний сигнал з суміші корисного сигналу та шуму, так і прогнозувати майбутні значення корисного сигналу. Метою фільтра є згенерувати вихідний сигнал $y(t)$ якомога ближчим до корисного сигналу $s(t + \eta)$, де $\eta \geq 0$ є заданим параметром. Фраза “якомога ближчим” розуміється як “має найменше середньоквадратичне відхилення”. Параметр η є часовим інтервалом, на який робиться прогнозування.

Вихідний сигнал фільтру задається виразом (1.41), стоїть задача вивести рівняння на відповідну вагову функцію $h(\tau)$. Як зазначено вище, метою фільтра є мінімізувати середньоквадратичне відхилення

$$\xi = \left\langle (s(t + \eta) - y(t))^2 \right\rangle \rightarrow \min. \quad (1.52)$$

Нашою метою наразі є вивести рівняння на вагову функцію $h(\tau)$ на основі виразів (1.52) та (1.41).

Очевидно, що

$$\xi = \left\langle (s(t + \eta) - y(t))^2 \right\rangle = \left\langle s^2(t + \eta) \right\rangle - 2\left\langle s(t + \eta)y(t) \right\rangle + \left\langle y^2(t) \right\rangle \rightarrow \min, \quad (1.53)$$

з урахуванням (1.41) матимемо

$$\xi = \xi_1 + \xi_2 + \xi_3, \quad \xi_1 = \left\langle s^2(t + \eta) \right\rangle, \quad \xi_2 = -2 \left\langle s(t + \eta) \int_{-\infty}^{+\infty} d\tau h(\tau) x(t - \tau) \right\rangle, \quad (1.54)$$

$$\xi_3 = \left\langle \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) h(\tau') x(t-\tau) x(t-\tau') \right\rangle.$$

Значення ξ_1 є середнім квадратом від стаціонарного випадкового процесу, тож не залежить від часу (це можна показати на основі визначення (1.8) аналогічно до властивості (1.15)). Тож $\xi_1 = \text{const}$. Щодо значення ξ_2 , матимемо

$$\begin{aligned} \xi_2 &= -2 \left\langle s(t+\eta) \int_{-\infty}^{+\infty} d\tau h(\tau) x(t-\tau) \right\rangle = -2 \left\langle \int_{-\infty}^{+\infty} d\tau h(\tau) s(t+\eta) x(t-\tau) \right\rangle = \\ &= -2 \int_{-\infty}^{+\infty} d\tau h(\tau) \langle s(t+\eta) x(t-\tau) \rangle, \end{aligned} \quad (1.55)$$

тут враховано, що інтеграл від середнього є середнім від інтегралу, та те, що функція $h(\tau)$ не є випадковою, тож її можна винести як множник за знак середнього. Згідно (1.26) та (1.28) матимемо

$$\begin{aligned} \langle s(t+\eta) x(t-\tau) \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds dx s x \omega(s, t+\eta; x, t-\tau) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds dx s x \omega(s, t+\eta+\tau; x, t) = \langle s(t+\eta+\tau) x(t) \rangle_t = R_{sx}(\eta+\tau), \end{aligned} \quad (1.56)$$

тож з урахуванням (1.55) отримаємо

$$\xi_2 = -2 \int_{-\infty}^{+\infty} d\tau h(\tau) R_{sx}(\tau+\eta). \quad (1.57)$$

Аналогічно матимемо

$$\begin{aligned} \xi_3 &= \left\langle \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) h(\tau') x(t-\tau) x(t-\tau') \right\rangle = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) h(\tau') \langle x(t-\tau) x(t-\tau') \rangle. \end{aligned} \quad (1.58)$$

Згідно (1.9), (1.12) та (1.21) отримаємо

$$\langle x(t-\tau) x(t-\tau') \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, t-\tau; x_2, t-\tau') = \quad (1.59)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 x_1 x_2 \omega_2(x_1, t; x_2, t - \tau' + \tau) = \langle x(t) x(t + \tau - \tau') \rangle_t = R_x(\tau - \tau'),$$

ТОЖ

$$\xi_3 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) h(\tau') R_x(\tau - \tau'). \quad (1.60)$$

Відповідно, згідно (1.53), (1.54), (1.57) та (1.60) отримаємо

$$F[h(\tau)] = -2 \int_{-\infty}^{+\infty} d\tau h(\tau) R_{sx}(\tau + \eta) + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) h(\tau') R_x(\tau - \tau') \rightarrow \min, \quad (1.61)$$

константу ξ_1 можна не виписувати, бо мова йде про мінімізацію функціоналу $F[h(\tau)]$. Надалі треба мінімізувати цей функціонал, тобто знайти функцію $h(\tau)$, для якої значення F буде найменшим. Ідея мінімізації цього функціоналу є такою. Нехай $h(\tau)$ є ваговою функцією, яка мінімізує функціонал (1.61). Якщо ми візьмемо іншу вагову функцію у вигляді $h(\tau) + \varepsilon g(\tau)$, же ε – довільне число та $g(\tau)$ – довільна функція, то має бути справедливою така нерівність:

$$\forall \varepsilon, \forall g(\tau) \quad F[h(\tau)] \leq F[h(\tau) + \varepsilon g(\tau)]. \quad (1.62)$$

На основі (1.61) матимемо

$$\begin{aligned} F[h(\tau)] - F[h(\tau) + \varepsilon g(\tau)] = \\ = -2 \int_{-\infty}^{+\infty} d\tau h(\tau) R_{sx}(\tau + \eta) + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) h(\tau') R_x(\tau - \tau') - \\ - \left(-2 \int_{-\infty}^{+\infty} d\tau (h(\tau) + \varepsilon g(\tau)) R_{sx}(\tau + \eta) + \right. \\ \left. + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' (h(\tau) + \varepsilon g(\tau)) (h(\tau') + \varepsilon g(\tau')) R_x(\tau - \tau') \right), \end{aligned} \quad (1.63)$$

що після розкриття дужок перетворюється як

$$\begin{aligned} F[h(\tau)] - F[h(\tau) + \varepsilon g(\tau)] = \\ = \varepsilon \left(-2 \int_{-\infty}^{+\infty} d\tau g(\tau) R_{sx}(\tau + \eta) + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) g(\tau') R_x(\tau - \tau') + \right. \end{aligned} \quad (1.64)$$

$$+ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) h(\tau') R_x(\tau - \tau') \Big) - \varepsilon^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) g(\tau') R_x(\tau - \tau').$$

Функція $R_x(\tau)$ є парною, тож

$$R_x(\tau - \tau') = R_x(\tau' - \tau), \quad (1.65)$$

звідки

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) g(\tau') R_x(\tau - \tau') = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) h(\tau') R_x(\tau - \tau'), \quad (1.66)$$

та

$$\begin{aligned} & F[h(\tau)] - F[h(\tau) + \varepsilon g(\tau)] = \\ & = -2\varepsilon \left(\int_{-\infty}^{+\infty} d\tau g(\tau) R_{xx}(\tau + \eta) - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) g(\tau') R_x(\tau - \tau') \right) - \\ & - \varepsilon^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) g(\tau') R_x(\tau - \tau'). \end{aligned} \quad (1.67)$$

Згідно (1.67) та (1.62) бачимо, що

$$\begin{aligned} \forall \varepsilon, \forall g(\tau) \quad & 2\varepsilon \left(\int_{-\infty}^{+\infty} d\tau g(\tau) R_{xx}(\tau + \eta) - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) g(\tau') R_x(\tau - \tau') \right) + \\ & + \varepsilon^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) g(\tau') R_x(\tau - \tau') \geq 0. \end{aligned} \quad (1.68)$$

Розглянемо член, який містить ε^2 . Згідно (1.59) отримаємо

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) g(\tau') R_x(\tau - \tau') &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' g(\tau) g(\tau') \langle x(t - \tau) x(t - \tau') \rangle = \\ &= \left\langle \int_{-\infty}^{+\infty} d\tau g(\tau) x(t - \tau) \int_{-\infty}^{+\infty} d\tau' g(\tau') x(t - \tau') \right\rangle = \left\langle \left(\int_{-\infty}^{+\infty} d\tau g(\tau) x(t - \tau) \right)^2 \right\rangle \geq 0, \end{aligned} \quad (1.69)$$

тож $\forall \varepsilon, \forall g(\tau)$ член у (1.68), що містить ε^2 , є невід'ємним.

Тож, член при ε у (1.68) теж має бути невід'ємним для будь-якого числа ε та для будь-якої функції $g(\tau)$, бо, наприклад, при $\varepsilon \rightarrow 0$ саме член при ε , а не

при ε^2 буде головним у (1.68). Так як число ε може бути як додатним, так і від'ємним, то відповідний член завжди невід'ємний лише при виконанні умови

$$\forall g(\tau) \int_{-\infty}^{+\infty} d\tau g(\tau) R_{sx}(\tau + \eta) - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\tau d\tau' h(\tau) g(\tau') R_x(\tau - \tau') = 0, \quad (1.70)$$

звідки

$$\forall g(\tau) \int_{-\infty}^{+\infty} d\tau' g(\tau') \left(R_{sx}(\tau' + \eta) - \int_{-\infty}^{+\infty} d\tau h(\tau) R_x(\tau - \tau') \right) = 0. \quad (1.71)$$

Рівність (1.71) має справджуватись для будь-якої функції $g(\tau)$, тож відповідний вираз у дужках має дорівнювати нулю:

$$R_{sx}(\tau' + \eta) - \int_{-\infty}^{+\infty} d\tau h(\tau) R_x(\tau - \tau') = 0. \quad (1.72)$$

На основі (1.46) остаточно запишемо

$$R_{sx}(\tau' + \eta) = \int_0^{+\infty} d\tau h(\tau) R_x(\tau - \tau'). \quad (1.73)$$

Таким чином вагова функція фільтра Колмогорова–Вінера має бути розв'язком рівняння (1.73), яке називається рівнянням Вінера–Хопфа.

Вихідний сигнал фільтра Колмогорова–Вінера може бути отриманим таким чином. Спершу шукаємо вагову функцію фільтра як розв'язок рівняння Вінера–Хопфа (1.73), а далі обчислюємо вихідний сигнал фільтра згідно виразу

$$y(t) = \int_0^{+\infty} d\tau h(\tau) x(t - \tau), \quad (1.74)$$

цей вираз виписано на основі (1.46) та (1.41).

Слід зауважити, що вище використано аналогові сигнали, визначені при $t \in (-\infty, +\infty)$. Якщо вхідний сигнал фільтра визначено на проміжку $t \in (0, T)$, то рівняння Вінера–Хопфа та вихідний сигнал фільтра прийматимуть вигляд

$$R_{sx}(\tau' + \eta) = \int_0^T d\tau h(\tau) R_x(\tau - \tau'), \quad y(t) = \int_0^T d\tau h(\tau) x(t - \tau). \quad (1.75)$$

Якщо ж вхідний сигнал фільтра є дискретним сигналом, визначеним для $t = 0, 1, 2, \dots, T$, то інтеграл зміниться на суму та відповідні вирази набудуть вигляду

$$R_{sx}(\tau' + \eta) = \sum_{\tau=0}^T h_{\tau} R_x(\tau - \tau'), \quad y(t) = \sum_{\tau=0}^T h_{\tau} x(t - \tau), \quad (1.76)$$

в такому разі замість вагової функції фільтра маємо набір вагових коефіцієнтів h_{τ} , $\tau = 0, 1, 2, \dots, T$, та перший вираз у (1.76) є не інтегральним рівнянням, а системою лінійних алгебраїчних рівнянь на вагові коефіцієнти фільтра.

1.4. Градієнтні методи пошуку мінімуму функції багатьох змінних

Задачі теорії фільтрації часто зводяться до пошуку мінімуму функції багатьох змінних. Відповідний числовий пошук може бути реалізованим на основі градієнтних методів.

Спочатку для простоти розглянемо ідею застосування градієнтного методу у випадку функції двох змінних $f(x, y)$. Розглянемо систему декартових координат x та y на площині. Оберемо точку з координатами (X, Y) та будемо рухатись в її малому околі. Як відомо, напрям найшвидшого зростання функції $f(x, y)$ є напрямом градієнту функції в цій точці, а напрям найшвидшого спадання протилежний до напрямку градієнту. Вектор градієнту визначається таким чином:

$$\text{grad}f(x, y)\big|_{x=X, y=Y} = \frac{\partial f(x, y)}{\partial x}\bigg|_{x=X, y=Y} \cdot \vec{e}_x + \frac{\partial f(x, y)}{\partial y}\bigg|_{x=X, y=Y} \cdot \vec{e}_y, \quad (1.77)$$

де $\vec{e}_x$ та $\vec{e}_y$ – орти відповідних осей.

Числовий пошук мінімуму функції реалізовується таким чином. Починаємо рух з довільної точки (x_0, y_0) , та в її околі рухаємось протилежно до вектору градієнту. Ми “потрапляємо” в точку (x_1, y_1) , таку що

$$\overrightarrow{(x_1 - x_0; y_1 - y_0)} = -\mu_0 \text{grad}f(x, y)\big|_{x=x_0, y=y_0}, \quad (1.78)$$

що з урахуванням (1.77) дає

$$x_1 = x_0 - \mu_0 \frac{\partial f(x, y)}{\partial x} \Big|_{x=x_0, y=y_0}, \quad y_1 = y_0 - \mu_0 \frac{\partial f(x, y)}{\partial y} \Big|_{x=x_0, y=y_0}, \quad (1.79)$$

μ_0 – додатне число, вибір якого буде описаний далі. Так як ми рухались в напрямку найшвидшого спадання, то

$$f(x_1, y_1) < f(x_0, y_0). \quad (1.80)$$

Далі, аналогічно процедурі, описаній вище, рухаємось з точки (x_1, y_1) в її околі проти напрямку градієнту, та “потрапляємо” в точку (x_2, y_2) , координати якої задаються виразами

$$\begin{aligned} \overrightarrow{(x_2 - x_1; y_2 - y_1)} &= -\mu_1 \operatorname{grad} f(x, y) \Big|_{x=x_1, y=y_1}, \quad \mu_1 > 0, \\ x_2 &= x_1 - \mu_1 \frac{\partial f(x, y)}{\partial x} \Big|_{x=x_1, y=y_1}, \quad y_2 = y_1 - \mu_1 \frac{\partial f(x, y)}{\partial y} \Big|_{x=x_1, y=y_1}, \end{aligned} \quad (1.81)$$

та маємо очевидний ланцюжок нерівностей

$$f(x_2, y_2) < f(x_1, y_1) < f(x_0, y_0), \quad (1.82)$$

і так далі. На кроці з номером n ми рухаємось з точки (x_n, y_n) в точку (x_{n+1}, y_{n+1}) :

$$\begin{aligned} \overrightarrow{(x_{n+1} - x_n; y_{n+1} - y_n)} &= -\mu_n \operatorname{grad} f(x, y) \Big|_{x=x_n, y=y_n}, \quad \mu_n > 0, \\ x_{n+1} &= x_n - \mu_n \frac{\partial f(x, y)}{\partial x} \Big|_{x=x_n, y=y_n}, \quad y_{n+1} = y_n - \mu_n \frac{\partial f(x, y)}{\partial y} \Big|_{x=x_n, y=y_n}, \end{aligned} \quad (1.83)$$

$$f(x_{n+1}, y_{n+1}) < f(x_n, y_n) < \dots < f(x_0, y_0),$$

зауважимо, що рух від точки (x_0, y_0) до точки (x_1, y_1) вважається нульовим кроком алгоритму. Алгоритм завершується при досягненні умови

$$|f(x_{n+1}, y_{n+1}) - f(x_n, y_n)| < \varepsilon, \quad (1.84)$$

де ε – наперед задана точність алгоритму. Відповідне наближено знайдене мінімальне значення функції $f_{\min} = f(x_{n+1}, y_{n+1})$.

Тепер розглянемо випадок функції N змінних $f(x^1, x^2, \dots, x^N)$, тут верхній індекс означає номер змінної, а не ступінь. Алгоритм аналогічним до випадку $N = 2$. Ми стартуємо з точки $(x_0^1, x_0^2, \dots, x_0^N)$ та на кожному кроці рухаємось з точки $(x_n^1, x_n^2, \dots, x_n^N)$ в точку $(x_{n+1}^1, x_{n+1}^2, \dots, x_{n+1}^N)$ таким чином:

$$x_{n+1}^j = x_n^j - \mu_n \left. \frac{\partial f(x^0, x^1, \dots, x^N)}{\partial x^j} \right|_{x^i = x_n^i}, \quad i, j = 1, 2, \dots, N, \quad \mu_n > 0, \quad (1.85)$$

$$f(x_{n+1}^0, x_{n+1}^1, \dots, x_{n+1}^N) < f(x_n^0, x_n^1, \dots, x_n^N) < \dots < f(x_0^0, x_0^1, \dots, x_0^N),$$

алгоритм зупиняється при досягненні умови

$$\left| f(x_{n+1}^0, x_{n+1}^1, \dots, x_{n+1}^N) - f(x_n^0, x_n^1, \dots, x_n^N) \right| < \varepsilon, \quad (1.86)$$

ε – наперед задана точність алгоритму. Відповідне наближено знайдене мінімальне значення функції $f_{\min} = f(x_{n+1}^0, x_{n+1}^1, \dots, x_{n+1}^N)$.

Тепер розглянемо вибір значень μ_n . Методом спроб та помилок ці значення намагаються підібрати так, щоб алгоритм збігався, та збігався якнайшвидше. Деякими популярними в літературі рекомендації викладені нижче.

1. *Метод з постійним кроком:* $\mu_0 = \mu_1 = \mu_2 = \dots = \text{const}$.

2. *Метод з дробовим кроком.* Методом спроб і помилок обирають константи $\alpha \in (0, 1)$ та $\delta \in (0, 1)$. На кожному n -му кроці діють таким чином. Спочатку μ_n присвоюють значення μ_0 . Надалі перевіряється, чи виконується нерівність умова

$$f\left(x_n^j - \mu_n \left. \frac{\partial f}{\partial x^j} \right|_{x^i = x_n^i}\right) \leq f(x_n^j) - \alpha \mu_n \cdot \sum_{j=1}^N \left(\left. \frac{\partial f}{\partial x^j} \right|_{x^i = x_n^i} \right)^2. \quad (1.87)$$

Якщо так, то μ_n обирають як $\mu_n = \mu_0$, і на цьому завершують крок. Якщо ні, то μ_n присвоюють значення $\mu_0 \delta$, та нерівність (1.87) перевіряють знову. Якщо вона спрацьовує, то обирають μ_n як $\mu_n = \mu_0 \delta$, та завершують крок. Якщо ні, то μ_n присвоюють значення $\mu_0 \delta^2$, і так далі. Зауважимо, що популярним в літературі вибором є $\alpha = 0,5$; $\delta = 0,95$.

3. Метод найшвидшого градієнтного спуску. Значення μ_n обирають так, щоб мінімізувати вираз

$$f\left(x_n^j - \mu_n \frac{\partial f}{\partial x^j} \Big|_{x^i=x_n^i}\right) \rightarrow \min. \quad (1.88)$$

Розглянемо тепер застосування цих методів до пошуку вагових коефіцієнтів дискретного фільтра Колмогорова–Вінера. Для простоти розглянемо лише задачу відкидання шуму. Нехай у нас є сигнал

$$x(t) = s(t) + n(t), \quad (1.89)$$

де $s(t)$ – корисний сигнал та $n(t)$ – шум. Вхідний сигнал задано в дискретні моменти часу $t = 0, 1, 2, \dots, N$. Вихідним сигналом є

$$y(t) = \sum_{\tau=0}^N h(\tau) x(t - \tau), \quad (1.90)$$

його треба знайти на основі мінімізації середньоквадратичної помилки

$$\left\langle (s(t) - y(t))^2 \right\rangle \rightarrow \min. \quad (1.91)$$

Іншими словами, стоїть задача пошуку вагових коефіцієнтів фільтра $h(j)$, $j = \overline{1, N}$, на основі виразів (1.90) та (1.91). Вважається, що кореляційні функції $R_x(t)$ та $R_{sx}(t)$ задано. Як відомо, для знаходження таких коефіцієнтів треба розв'язати систему рівнянь (1.76). Альтернативою може бути використання градієнтного методу.

На основі (1.90) та (1.91) матимемо

$$\begin{aligned} \left\langle (s(t) - y(t))^2 \right\rangle_t &= \left\langle s^2(t) \right\rangle_t - 2 \left\langle s(t) y(t) \right\rangle_t + \left\langle y^2(t) \right\rangle_t = \\ &= \left\langle s^2(t) \right\rangle - 2 \sum_{\tau=0}^N h(\tau) \left\langle s(t) x(t - \tau) \right\rangle + \sum_{\tau=0}^N \sum_{\tau'=0}^N h(\tau) h(\tau') \left\langle x(t - \tau) x(t - \tau') \right\rangle. \end{aligned} \quad (1.92)$$

На основі (1.21) матимемо

$$R_s(\tau) = \left\langle s(t + \tau) s(t) \right\rangle \Rightarrow \left\langle s^2(t) \right\rangle = R_s(\tau = 0) = \text{const}, \quad (1.93)$$

згідно (1.34) та (1.37) отримаємо

$$R_{xs}(-\tau) = \langle x(t-\tau)s(t) \rangle = R_{sx}(\tau), \quad (1.94)$$

та згідно (1.59)

$$\langle x(t-\tau)x(t-\tau') \rangle = R_x(\tau-\tau'), \quad (1.95)$$

На основі (1.92) – (1.95) бачимо, що треба мінімізувати функцію N змінних:

$$f(h^0, h^1, \dots, h^N) = -2 \sum_{\tau=0}^N h^\tau R_{sx}(\tau) + \sum_{\tau=0}^N \sum_{\tau'=0}^N h^\tau h^{\tau'} R_{xx}(\tau-\tau') \rightarrow \min, \quad (1.96)$$

$$h^j = h(j),$$

для чого може бути використаний градієнтний метод.

Для прикладу розглянемо застосування методу найшвидшого градієнтного спуску до цієї задачі. В його рамках аналітично знайдемо вирази для μ_n .

Перепишемо (1.96) у матричному вигляді:

$$f(h) = -2h^T R_{sx} + h^T R_x h, \quad (1.97)$$

де індекс T означає транспонування. Тут введено матриці

$$h = \begin{pmatrix} h^0 \\ h^1 \\ \vdots \\ h^N \end{pmatrix}, \quad R_{sx} = \begin{pmatrix} R_{sx}(0) \\ R_{sx}(1) \\ \vdots \\ R_{sx}(N) \end{pmatrix}, \quad (1.98)$$

$$R_x = \begin{pmatrix} R_x(0) & R_x(1) & R_x(2) & \dots & R_x(N) \\ R_x(1) & R_x(0) & R_x(1) & \dots & R_x(N-1) \\ R_x(2) & R_x(1) & R_x(0) & \dots & R_x(N-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_x(N) & R_x(N-1) & R_x(N-2) & \dots & R_x(0) \end{pmatrix},$$

Та при виписування матриці R_x врахована парність кореляційної функції.

Компоненти вектору градієнта задаються таким чином:

$$\text{grad}_j f(h) = \frac{\partial f(h^0, h^1, \dots, h^N)}{\partial h^j}. \quad (1.99)$$

Обчислимо їх згідно (1.96):

$$\begin{aligned}
\text{grad}_j f(h) &= -2 \sum_{\tau=0}^N \frac{\partial h^\tau}{\partial h^j} R_{sx}(\tau) + \sum_{\tau=0}^N \sum_{\tau'=0}^N \frac{\partial h^\tau h^{\tau'}}{\partial h^j} R_x(\tau - \tau') = \\
&= -2 \sum_{\tau=0}^N \delta_{\tau j} R_{sx}(\tau) + \sum_{\tau=0}^N \sum_{\tau'=0}^N (\delta_{\tau j} h^{\tau'} + \delta_{\tau' j} h^\tau) R_x(\tau - \tau') = \\
&= -2 \sum_{\tau=0}^N \delta_{\tau j} R_{sx}(\tau) + \sum_{\tau=0}^N \sum_{\tau'=0}^N \delta_{\tau j} h^{\tau'} R_x(\tau - \tau') + \sum_{\tau=0}^N \sum_{\tau'=0}^N \delta_{\tau' j} h^\tau R_x(\tau - \tau'), \\
\delta_{\tau j} &= \begin{cases} 1, \tau = j \\ 0, \tau \neq j \end{cases}.
\end{aligned} \tag{1.100}$$

З урахуванням парності кореляційної функції отримаємо

$$\sum_{\tau=0}^N \sum_{\tau'=0}^N \delta_{\tau' j} h^\tau R_x(\tau - \tau') = \sum_{\tau=0}^N \sum_{\tau'=0}^N \delta_{\tau j} h^{\tau'} R_x(\tau' - \tau) = \sum_{\tau=0}^N \sum_{\tau'=0}^N \delta_{\tau j} h^{\tau'} R_x(\tau - \tau'), \tag{1.101}$$

звідки з урахуванням (1.100) матимемо

$$\text{grad}_j f(h) = -2R_{sx}(j) + 2 \sum_{\tau=0}^N R_x(\tau - j) h^\tau, \tag{1.102}$$

що в матричній формі запишеться як

$$\text{grad} f(h) = -2R_{sx} + 2R_x \cdot h. \tag{1.103}$$

Тепер знайдемо вираз для μ_n . Умова (1.88) в матричній форму набуде вигляду

$$f(h - \mu \cdot \text{grad} f(h)) \rightarrow \min, \tag{1.104}$$

тут і надалі індекс n опущено для простоти. На основі (1.97) отримаємо

$$\begin{aligned}
f(h - \mu \cdot \text{grad} f(h)) &= -2(h - \mu \cdot \text{grad} f(h))^T R_{sx} + \\
&+ (h - \mu \cdot \text{grad} f(h))^T R_x (h - \mu \cdot \text{grad} f(h)),
\end{aligned} \tag{1.105}$$

що після розкриття дужок дає

$$\begin{aligned}
f(h - \mu \cdot \text{grad} f(h)) &= h^T R_x h - 2h^T R_{sx} + \\
&+ \mu \left(2(\text{grad} f(h))^T R_{sx} - (\text{grad} f(h))^T R_x h - h^T R_x \text{grad} f(h) \right) + \\
&+ \mu^2 \cdot (\text{grad} f(h))^T R_x \text{grad} f(h).
\end{aligned} \tag{1.106}$$

Як бачимо з (1.98), R_x є симетричною матрицею, тож

$$R_x = R_x^T, \left((\text{grad}f(h))^T R_x h \right)^T = h^T R_x^T \text{grad}f(h) = h^T R_x \text{grad}f(h). \quad (1.107)$$

Очевидно, що $h^T R_x \text{grad}f(h)$ є числом (матрицею 1×1), тож

$$(\text{grad}f(h))^T R_x h = \left((\text{grad}f(h))^T R_x h \right)^T = h^T R_x \text{grad}f(h), \quad (1.108)$$

що з урахуванням (1.106) дає

$$\begin{aligned} f(h - \mu \cdot \text{grad}f(h)) &= h^T R_x h - 2h^T R_{sx} + \\ &+ 2\mu \left((\text{grad}f(h))^T R_{sx} - h^T R_x \text{grad}f(h) \right) + \\ &+ \mu^2 \cdot (\text{grad}f(h))^T R_x \text{grad}f(h) \rightarrow \min. \end{aligned} \quad (1.109)$$

Беручи похідну по μ , після певних перетворень отримаємо

$$\frac{df(h - \mu \cdot \text{grad}f(h))}{d\mu} = 0 \Rightarrow \mu = \frac{h^T \cdot R_x \cdot \text{grad}f(h) - (\text{grad}f(h))^T \cdot R_{sx}}{(\text{grad}f(h))^T \cdot R_x \cdot \text{grad}f(h)}. \quad (1.110)$$

Слід зауважити, що у виразі (1.110) і чисельник, і знаменник є числами, а не матрицями.

Підсумовуючи вище сказане, задача пошуку вагових коефіцієнтів фільтра Колмогорова–Вінера може бути зведена, зокрема, і до задачі пошуку мінімуму функції багатьох змінних, яку можна вирішити градієнтними методами. В рамках методу найшвидшого градієнтного спуску виведено формулу для числового кроку μ_n :

$$\mu_n = \frac{h^T \cdot R_x \cdot \text{grad}f(h) - (\text{grad}f(h))^T \cdot R_{sx}}{(\text{grad}f(h))^T \cdot R_x \cdot \text{grad}f(h)} \Bigg|_{h^j = h_n^j}. \quad (1.111)$$

Наприкінці зауважимо, що градієнтні методи можуть бути використані не лише для пошуку локального мінімуму, а і для пошуку локального максимуму функції багатьох змінних. В такому разі ідея застосування методу є такою ж, але на кожному кроці треба рухатись за напрямком градієнту, а не проти нього.

1.5. Застосування фільтра Колмогорова – Вінера до прогнозування стаціонарних процесів з важкими хвостами (з КФСПЕ)

Цей параграф ґрунтується на наукових результатах, нещодавно отриманих та опублікованих авторами цього посібника.

Розглянемо застосування фільтра Колмогорова–Вінера до прогнозування стаціонарного процесу з важким хвостом. Згенерувати такий процес можна, наприклад, на основі методу симетричного ковзного середнього. Нехай маємо стаціонарний дискретний випадковий процес V_i , який є білим шумом з середнім значенням, рівним нулю, та дисперсією, рівною 1. Тоді згідно методу симетричного ковзного середнього процес X_i , побудований на основі таких виразів:

$$X_i = \sum_{j=-q}^q a_{|j|} V_{i+j+q}, \quad a_0 = \frac{\sqrt{2-2H}}{1,5-H} \gamma_0, \quad (1.112)$$
$$a_j = \frac{a_0}{2} \left((j+1)^{H+0.5} + (j-1)^{H+0.5} - 2j^{H+0.5} \right)$$

є процесом з важким хвостом, середнє значення якого дорівнює нулю (на практиці – близьке до нуля), дисперсія такого процесу дорівнює γ_0 , та коефіцієнт Херста такого процесу дорівнює H , має виконуватись умова $H \in (0,5;1)$. Іншими словами, при великих часах t кореляційна функція

$$R_x(t) \sim t^{2H-2}, \quad t \rightarrow \infty, \quad (1.113)$$

тож спадання кореляційної функції з часом є не експоненційним, а набагато більш повільним степеневим, що і обумовлює назву «процес з важким хвостом». Походження формул (1.112) детально описане в роботі [1] та у посиланнях, наведених в цій роботі. Зокрема, в [1] зазначено, що відповідний процес X_i має бути фрактальним гаусівським шумом, причому в ідеалі параметр q має прямувати до нескінченності, на практиці q має бути достатньо великим числом. У роботах [2, 3] взято значення $q = 3 \cdot 10^5$, та відповідно згенеровано 10^5 точок процесу X_i з дисперсією $\gamma_0 = 1$. В наш час телекомунікаційний трафік в системах

з пакетною передачею даних вважається випадковим процесом з важким хвостом, тож для того, щоб побудувати модельні дані трафіку у простій моделі, треба згенерувати невід’ємний випадковий процес з важким хвостом, який згенеруємо таким чином:

$$x_i = X_i + \min(X) + 10^{-3}, \quad (1.114)$$

малий доданок 10^{-3} додано, щоб уникнути при обчисленнях нескінченної абсолютної відсоткової похибки прогнозування, яка може виникнути, якщо «реальне» значення трафіку в цій точці нульове, а прогнозне – ненульове, хоч і дуже близьке до нуля.

Надалі розглядаємо відповідний центрований (тобто з нульовим середнім значенням) процес

$$xc_i = x_i - \langle x \rangle, \quad (1.115)$$

та обчислюємо його кореляційну функцію:

$$R_x(t) = \frac{1}{10^5 - \tau} \sum_{i=1}^{10^5 - \tau} xc_i xc_{i+\tau}, \quad (1.116)$$

тут враховано, що згенеровано саме 10^5 точок випадкового процесу з важким хвостом.

Надалі реалізуємо прогнозування отриманих модельних даних. Зауважимо, що вони вважаються незашумленими, тобто перед фільтром ставимо лише задачу прогнозування без відкидання шуму. Дослідимо прогнозування на z точок вперед на основі попередніх $T + 1$ точок, $z \ll T$. Тоді згідно (1.76) вагові коефіцієнти дискретного фільтра Колмогорова–Вінера обчислюються таким чином:

$$\begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \\ h_T \end{pmatrix} = \begin{pmatrix} R_x(0) & R_x(1) & R_x(2) & \cdots & R_x(T) \\ R_x(1) & R_x(0) & R_x(1) & \cdots & R_x(T-1) \\ R_x(2) & R_x(1) & R_x(0) & \cdots & R_x(T-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_x(T) & R_x(T-1) & R_x(T-2) & \cdots & R_x(0) \end{pmatrix}^{-1} \cdot \begin{pmatrix} R_x(z) \\ R_x(z+1) \\ R_x(z+2) \\ \vdots \\ R_x(z+T) \end{pmatrix}, \quad (1.117)$$

вираз (1.117) є розв’язанням системи рівнянь (1.76) в матричній формі.

На першій ітерації візьмемо перші $T + 1$ точку (точки з номерами від 1 до $T + 1$), та зробимо прогнозування для значень $xc_{T+2}, xc_{T+3}, \dots, xc_{T+z+1}$. На другій ітерації візьмемо точки з номерами від 2 до $T + 2$, та зробимо прогнозування для значень $xc_{T+3}, xc_{T+4}, \dots, xc_{T+z+2}$. І так далі вздовж всього масиву даних. На ітерації з номером i згідно (1.76) отримаємо

$$xc'_k = xc_{k+i}, k = \overline{0, T}; xc_{j+z} = \sum_{k=0}^j h_k xc'_{j-k}, j = \overline{T-z+1, T}, \quad (1.118)$$

де xc_{j+z} – прогнозне значення для xc_{j+z} . Тут слід зауважити, що так як величина xc'_{j-k} не визначена для від'ємного індексу, то верхня межа підсумовування у (1.118) взята рівною j , а не T . Але, зокрема, через те, що $z \ll T$ це, фактично, не спотворить результат прогнозування. Відповідно, прогнозне значення для величини x_j згідно (1.115) визначається як

$$x_j = xc_j + \langle x \rangle. \quad (1.119)$$

На ітерації з номером i для визначення точності прогнозування обчислюються середня абсолютна відсоткова похибка прогнозування (mean absolute percentage error, MAPE) згідно виразу

$$\text{MAPE}_i = \frac{1}{z} \sum_{j=T+1+i}^{T+z+i} \left| \frac{x_j - xc_j}{x_j} \right| \cdot 100\%. \quad (1.120)$$

Середню абсолютну відсоткову похибку прогнозування по всьому масиву обчислимо як

$$\text{MAPE} = \frac{1}{10^5 - T - z} \sum_{i=1}^{10^5 - T - z} \text{MAPE}_i. \quad (1.121)$$

Тут і надалі розглянемо числові результати для значення коефіцієнту Херста $H = 0,8$ та значення $T = 100$. Отримані результати для MAPE наведені у таблиці 1.1. Графік порівняння модельних даних та їх прогнозованих значень наведено на рис. 1.1 для випадку $z = 1$. На рис. 1.1. змодельовані дані процесу з важким хвостом зображені суцільною лінією, а прогнозовані дані – пунктирною.

Таблиця 1.1. MAPE (у %) для незгладженого процесу з важким хвостом.

z	MAPE, %
1	24,7
2	25,3
3	25,4
4	25,7
5	25,9

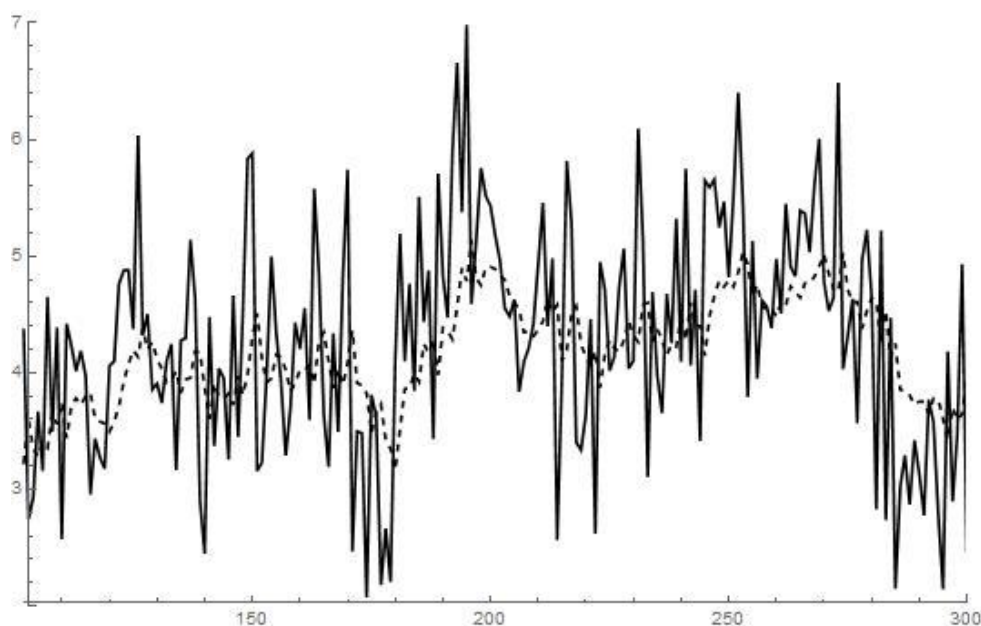


Рисунок 1.1. Порівняння модельних даних та їх прогнозованих значень для незгладженого процесу

Як бачимо, середня абсолютна відсоткова похибка прогнозування є досить великою (порядку 25%) у випадку незгладженого процесу. На рис. 1.1 бачимо, що у випадку незгладженого процесу прогнозовані значення адекватно відтворюють тенденцію процесу, але конкретні їх значення можуть дещо не співпадати з відповідними даними.

Але у випадку згладженого процесу фільтр Колмогорова–Вінера дає суттєво кращі результати прогнозування. Досліджено два випадки згладжених процесів – на основі лінійного та експоненціального згладжувань. У випадку лінійного згладжування відповідні модельні дані будуються таким чином:

$$\tilde{X}_i = \frac{1}{2l+1} \sum_{j=-l}^l X_{i+j}, \quad x_i = \tilde{X}_i + \min(\tilde{X}) + 10^{-3}, \quad (1.122)$$

а у випадку експоненціального згладжування – таким чином:

$$X'_t = \frac{1-\lambda}{1-\lambda^t} \sum_{k=0}^{t-1} \lambda^k X_{t-k}, \quad x_i = X'_i + \min(X') + 10^{-3}, \quad \lambda = 0.9. \quad (1.123)$$

Прогнозування для згладжених процесів та його помилка аналогічно обчислюються на основі формул (1.115) – (1.121). Отримані результати наведено у таблицях 1.2 та 1.3.

Таблиця 1.2. MAPE (у %) для згладженого процесу з важким хвостом, отриманому на основі лінійного згладжування,

	$z = 1$	$z = 2$	$z = 3$	$z = 4$	$z = 5$
$l = 1$	9,11	15,1	20,3	20,9	21,4
$l = 2$	6,26	10,2	13,8	17,2	21,2
$l = 3$	4,85	7,86	10,4	13,2	15,9
$l = 4$	3,92	6,13	8,07	10,0	12,0
$l = 5$	3,37	5,30	6,97	8,64	10,3

Таблиця 1.3. MAPE (у %) для згладженого процесу з важким хвостом, отриманому на основі експоненційного згладжування,

z	MAPE, %
1	4,03
2	6,09
3	7,84
4	9,20
5	10,4

Як бачимо, для згладжених процесів помилки прогнозування є значно меншими, та чим менше кількість точок, на яку треба прогнозувати, тим

точнішим є прогнозування. Відповідні графіки порівняння модельних даних та їх прогнозованих значень наведені на рис. 1.2 та рис. 1.3 (пунктир – прогноз, суцільна лінія – модельні дані).

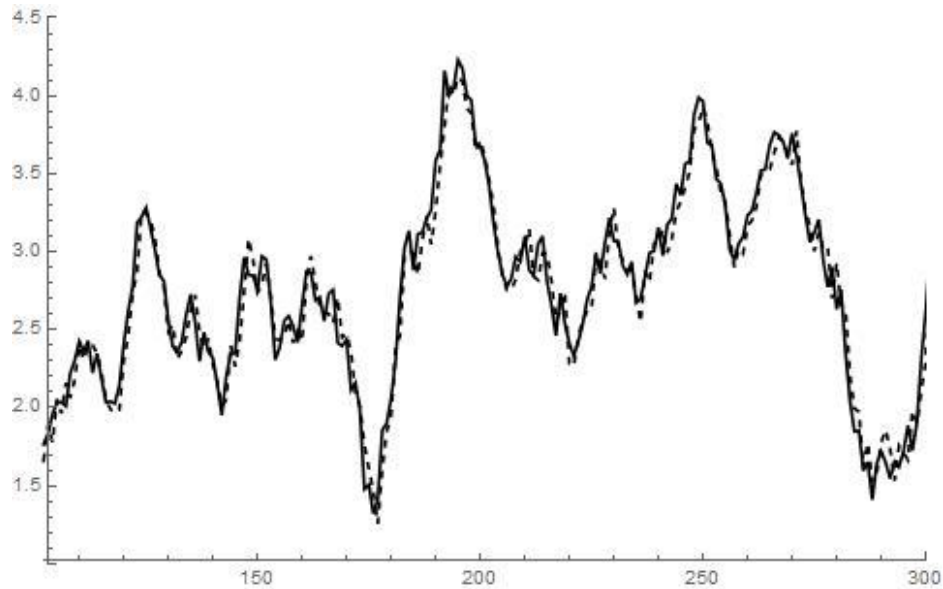


Рисунок 1.2. Порівняння модельних даних та їх прогнозованих значень для лінійно згладженого процесу, $z = 1$, $l = 3$

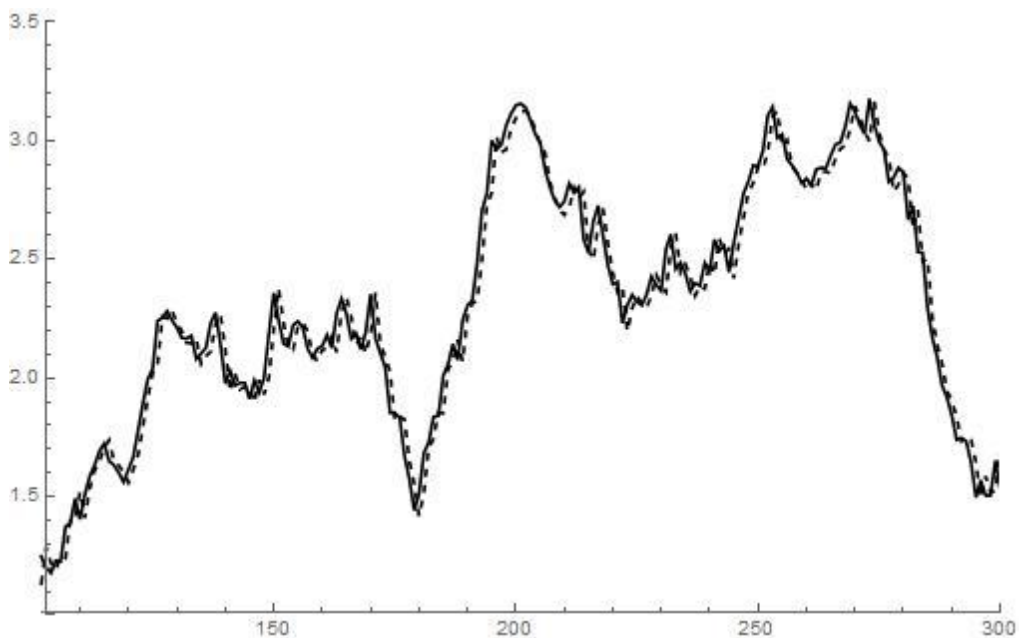


Рисунок 1.3. Порівняння модельних даних та їх прогнозованих значень для експоненційно згладженого процесу, $z = 1$

Як бачимо з рис. 1.2 та рис. 1.3, у випадку короткострокового прогнозування на 1 точку вперед фільтр Колмогорова–Вінера дає дуже хороше співпадіння модельних даних процесу з важким хвостом з їх прогнозованими значеннями.

Вище реалізовано дискретний фільтр Колмогорова–Вінера. Тепер поговоримо про неперервний фільтр. Безумовно, виміри трафіку, реально кажучи, є дискретними, та в рамках цієї задачі більш точним є використання дискретного фільтру. Але цікавим є питання дослідження застосовності неперервного фільтру Колмогорова–Вінера до цієї задачі. Припустимо, що велику кількість даних (наприклад, 10^5 точок) виміряно за невеликий проміжок часу (наприклад, за 100 секунд, тоді на 1 секунду приходить 1000 точок). Тоді такий процес можна грубо вважати неперервним, хоча його «неперервність» дещо штучна, див. рис. 1.4.

Розглянемо випадок, коли робиться прогнозування на $z = 10^{-3}$ секунди вперед на основі даних, виміряних в попередню $T = 1$ секунду (аналог дискретного випадку, де робиться прогнозування на 1 точку вперед на основі попередніх 1000 точок).

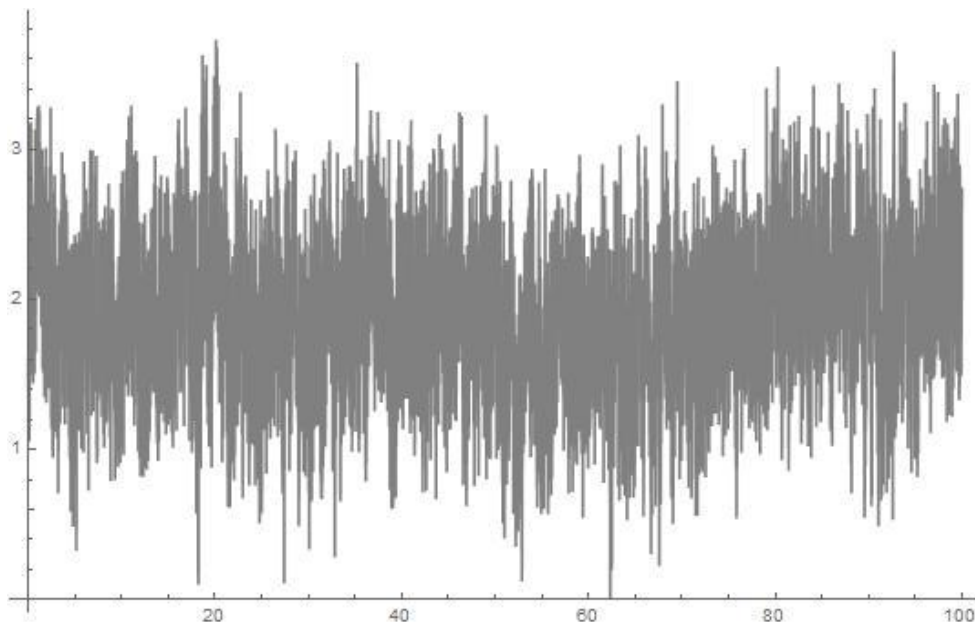


Рисунок 1.4. Експоненційно згладжений «неперервний» процес $x(t)$

Згідно (1.75) вагова функція фільтру є розв'язком інтегрального рівняння Вінера–Хопфа

$$\int_0^T h(\tau) R_x(t - \tau) d\tau = R_x(t + z), \quad (1.124)$$

яке можна, наприклад, наближено розв'язати на основі методу Галеркіна, в рамках якого невідома вагова функція шукається у вигляді обірваного розвинення за ортогональними функціями

$$h(\tau) = \sum_{s=0}^{n-1} g_s S_s(\tau), \quad (1.125)$$

де $S_s(\tau)$ – відповідні функції, ортогональні на проміжку $t \in [0, T]$ (конкретно в цьому дослідженні – функції Волша в нумерації Волша), та g_s – невідомі коефіцієнти, які треба знайти. Тоді підстановка (1.125) у (1.124) з подальшим множенням на $S_k(t)$ та додатковим інтегруванням приведе до системи рівнянь

$$\sum_{s=0}^{n-1} g_s \int_0^T \int_0^T S_s(\tau) S_k(t) R_x(t - \tau) d\tau dt = \int_0^T S_k(t) R_x(t + z) dt, \quad (1.126)$$

розв'язок якої в матричній формі має вигляд

$$\begin{pmatrix} g_0 \\ g_1 \\ \vdots \\ g_{n-1} \end{pmatrix} = \begin{pmatrix} G_{00} & G_{01} & \cdots & G_{0,n-1} \\ G_{10} & G_{11} & \cdots & G_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ G_{n-1,0} & G_{n-1,1} & \cdots & G_{n-1,n-1} \end{pmatrix}^{-1} \cdot \begin{pmatrix} B_0 \\ B_1 \\ \vdots \\ B_{n-1} \end{pmatrix}, \quad (1.127)$$

де

$$G_{ij} = \int_0^T \int_0^T S_i(\tau) S_j(t) R_x(t - \tau) dt d\tau, \quad B_i = \int_0^T S_i(t) R_x(t + z) dt, \quad (1.128)$$

величини G_{ij} називаються інтегральними дужками. Кореляційну функцію відповідного процесу апроксимовано на основі методу найменших квадратів на різних часових проміжках:

$$R_x(t) = \begin{cases} a_1 |t|^3 + b_1 |t|^2 + c_1 |t| + d_1, & |t| \in [0, 0.1] \\ a \cdot (|t| + 10^{-3})^{-b}, & |t| \in (0.1, 1] \end{cases}; \quad a = 0,0206; \quad b = 0,415; \quad (1.129)$$

$$a_1 = -369; b_1 = 79,0; c_1 = -5,89; d_1 = 0,220;$$

значення коефіцієнтів у (1.129) округлені до 3 значущих цифр.

Прогнозування робимо таким чином. Спочатку генеруємо відповідний централізований процес $xc(t)$. Далі на першій ітерації беремо дані процесу $xc(t)$ на проміжку $t \in [0, T]$, та прогнозуємо значення $xc(T + z)$. На другій ітерації беремо часовий інтервал $t \in [z, T + z]$, та прогнозуємо значення $xc(T + 2z)$, і так далі вздовж всього масиву. На ітерації з номером i матимемо

$$\begin{aligned} xc(T + iz) &= \int_0^T h(\tau) xc(T + (i-1) \cdot z - \tau) d\tau \approx \\ &\approx \sum_{j=0, 10^{-3}, 2 \cdot 10^{-3}, \dots}^{T-10^{-3}} h(j) \cdot xc(T + (i-1) \cdot z - j) \cdot 10^{-3}, \end{aligned} \quad (1.130)$$

де функція $h(\tau)$ обчислена на основі (1.125), та (1.127)–(1.129). Подальше прогнозування процесу $x(t)$ та MAPE аналогічне до розглянутого в дискретному фільтрі. Отримані числові результати наведені в таблиці 1.4.

Таблиця 1.4. MAPE (у %) у випадку використання неперервного фільтру

n	MAPE, %
32	14,9
64	11,3
128	7,94
256	5,81

Тут n – кількість функцій Волша у відповідному наближенні, див. (1.125). Як бачимо, при використанні доволі великої кількості функцій Волша можна отримати доволі хороше прогнозування (тобто помилка прогнозування буде малою). Відповідне графічне порівняння модельних даних та їх прогнозованих значень наведено на рис. 1.5 (пунктир – прогноз, суцільна лінія – модельні дані).

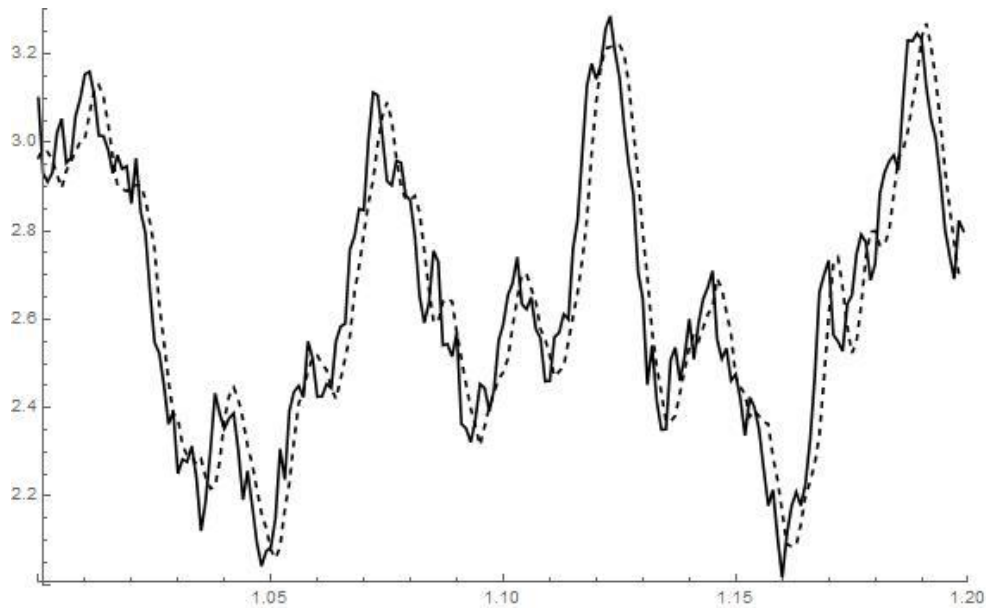


Рисунок 1.5. Порівняння модельних даних та їх прогнозованих значень у випадку використання неперервного фільтра

Як бачимо, прогнозні та модельні значення є доволі близькими, хоча і наявна певна часова затримка.

Коротким висновком до параграфу 1.5 є те, що фільтр Колмогорова–Вінера, як дискретний, так і неперервний, може бути застосовним до короткострокового прогнозування гладкого процесу з важким хвостом, що може бути застосовано на практиці для прогнозування трафіку в системах з пакетною передачею даних. Детальніше з відповідними дослідженнями можна ознайомитись у роботі [3].

Посилання, використані в параграфі 1.5:

[1] D. Koutsoyiannis, «The Hurst phenomenon and fractional Gaussian noise made easy». *Hydrological Sciences Journal*, 47, pp. 573–595, 2002.

[2] V. Gorev, A. Gusev and V. Korniienko, «The use of the Kolmogorov–Wiener filter for prediction of heavy-tail stationary processes». *CEUR Workshop Proceedings*, 3156, pp. 150–159, 2022. Available at: <http://ceur-ws.org/Vol-3156/paper9.pdf>.

[3] V. Gorev, A. Gusev, V. Korniienko and Y. Shedlovska, «On the use of the Kolmogorov–Wiener filter for heavy-tail process prediction». *Journal of Cyber Security and Mobility*, 12, Issue 3, pp. 315–338, 2023.

1.6. Контрольні питання до глави 1

1. Дайте визначення випадкової величини, випадкової функції та випадкового процесу.
2. Дайте визначення одновимірної, двовимірної та n -вимірної густини ймовірності. Вкажіть умови їх нормування та поясніть фізичний зміст цих умов.
3. Дайте визначення математичного очікування, середньоквадратичного значення, дисперсії та середньоквадратичного відхилення випадкового процесу (як середніх за множиною). Дайте визначення кореляційної функції випадкового процесу (як середнього за множиною).
4. Дайте визначення стаціонарного випадкового процесу (в широкому та у вузькому сенсі).
5. Сформулюйте та виведіть властивості математичного очікування та кореляційної функції стаціонарного процесу.
6. Сформулюйте визначення взаємної кореляційної функції двох випадкових процесів. Сформулюйте та виведіть властивості взаємної кореляційної функції, яка підпорядковується властивості (1.28).
7. Сформулюйте визначення середнього за часом значення випадкового процесу, а також часової кореляційної функції та часової взаємної кореляційної функції. Дайте визначення ергодичного процесу.
8. Дайте визначення фільтру, лінійного фільтру та стаціонарного фільтру.
9. Наведіть зв'язок вихідного та вхідного сигналів лінійного стаціонарного фільтру та доведіть, чому саме такий зв'язок описує і лінійність, і стаціонарність фільтру. Поясніть поняття вагової функції.
10. Поясніть, що таке принцип причинності, та виведіть властивість вагової функції, що базується на цьому принципі.
11. Поясніть призначення фільтру Колмогорова–Вінера та сформулюйте математичну постановку задачі цього фільтру.
12. Вкажіть ідею мінімізації функціоналу, що використовується при виведенні рівняння Вінера–Хопфа. Виведіть рівняння Вінера–Хопфа.

13. Поясніть основну ідею градієнтних методів. Вкажіть алгоритми, які реалізують градієнтні методи з дробовим та постійним кроком, а також метод найшвидшого градієнтного спуску.

14. Виведіть вираз для кроку методу найшвидшого градієнтного спуску для пошуку вагових коефіцієнтів дискретного фільтру Колмогорова–Вінера.

15. Вкажіть алгоритм методу симетричного ковзного середнього, що генерує дані фрактального гаусівського шуму. Вкажіть асимптотичну поведінку кореляційної функції фрактального гаусівського шуму на великих часах.

16. Вкажіть алгоритм дискретного фільтру Колмогорова–Вінера для прогнозування фрактального гаусівського шуму.

17. Опишіть лінійний та експоненційний способи згладжування сигналу.

18. Вкажіть характерні значення помилок прогнозування згладженого та не згладженого фрактального гаусівського шуму за допомогою як дискретного, так і неперервного фільтрів Колмогорова–Вінера.

Глава 2. Адаптивні фільтри

2.1. Скалярний дискретний фільтр Калмана

Розглянемо простий приклад. Нехай у нас є радіокерована машинка, чий рух одновимірний, тобто вона може рухатись лише вперед або назад. Припустимо, що її миттєва швидкість повністю задається джойстиком. Звісно, це дещо нефізично, до швидкостей не може змінюватись миттєво. Але поки «закриємо очі» на таку нефізичність, бо цей приклад «на пальцях» зрозуміло демонструватиме ідею скалярного фільтра Калмана.

В рамках такого прикладу, рух машинки повністю задано одним параметром – її координатою x . Нехай ця координата вимірюється в моменти часу $t_n = n \cdot \Delta t$, де Δt – малий часовий інтервал між двома «сусідніми» моментами часу. Фактично, маємо дискретний набір вимірювань $x_n = x(t_n)$.

Координата машинки, очевидно, змінюється за законом

$$x_{n+1} = x_n + v_n \Delta t, \quad (2.1)$$

де $v_n = v(t_n)$ – швидкість в момент t_n , яка керується джойстиком. Рівняння (2.1) описує зміну координати машинки в “ідеальному” випадку, коли рух повністю керується джойстиком. Але наявні певні збурення (вітер, неоднорідності дороги і т.д.), які теж впливають на рух. З їх урахуванням маємо

$$x_{n+1} = x_n + v_n \Delta t + \xi_n, \quad (2.2)$$

де величина ξ_n описує вплив вище описаних збурень, ξ_n – дискретний випадковий процес.

Нехай координата машинки вимірюється датчиком. Покази датчику z_n є такими:

$$z_n = x_n + \eta_n, \quad (2.3)$$

де η_n – дискретний випадковий процес, що описує помилку датчика (наприклад, вона може виникнути через шуми).

Фактично, ми маємо покази приладу z_n , та нам потрібен фільтр, який на виході дасть істинні значення координати x_n . Цю задачу можна розв'язати за допомогою фільтра Калмана. Наведений вище приклад був наданий для простоти. Тепер опишемо скалярний фільтр Калмана більш загально.

Нехай маємо систему, що описується одним параметром x , що змінюється дискретно, та на основі певних фізичних міркувань “ідеальний” закон зміни цього параметру може бути записаний як

$$x_{t+1} = ax_t + bu_t, \quad (2.4)$$

де u_t – відома величина, що керує еволюцією системи; a , b – деякі відомі константи (у вище розглянутому прикладі $u_t = v_t$, $a = 1$, $b = \Delta t$). Параметр x вимірюється приладом, покази якого дорівнюють z_t . Нехай η_t – випадковий стаціонарний процес, що є похибкою цього приладу. Нехай ξ_t – випадковий стаціонарний процес, що описує випадковий характер еволюції системи.

Тоді матимемо

$$x_{t+1} = ax_t + bu_t + \xi_t, \quad z_t = cx_t + \eta_t, \quad (2.5)$$

де константа c є коефіцієнтом підсилення приладу.

Вхідними даними фільтра Калмана є покази приладу z_t . Метою фільтра є згенерувати вихідну послідовність x_t^{opt} якомога ближчою до послідовності x_t . Фразу “якомога ближчою” слід розуміти як “таку, що має найменше середньоквадратичну відхилення”.

Позначимо помилку фільтра як e_t :

$$e_t = x_t - x_t^{\text{opt}}. \quad (2.6)$$

Припущеннями теорії є такі:

1. Дисперсії σ_ξ^2 та σ_η^2 процесів ξ_t та η_t , відповідно, є відомими.
2. Середні значення процесів ξ_t та η_t є нульовими:

$$\langle \xi_t \rangle = 0, \quad \langle \eta_t \rangle = 0. \quad (2.7)$$

3. Процеси ξ_t , η_t та помилка фільтру e_t – незалежні випадкові процеси, тобто

$$\begin{aligned}\forall t_1, t_2 \quad \langle \xi_{t_1} e_{t_2} \rangle &= \langle \xi_{t_1} \rangle \langle e_{t_2} \rangle = 0, \quad \langle \eta_{t_1} e_{t_2} \rangle = \langle \eta_{t_1} \rangle \langle e_{t_2} \rangle = 0, \\ \langle \xi_{t_1} \eta_{t_2} \rangle &= \langle \xi_{t_1} \rangle \langle \eta_{t_2} \rangle = 0,\end{aligned}\tag{2.8}$$

слід зауважити, що у (2.8) враховано результати (2.7).

Алгоритм фільтру Калмана є рекурентним. Припустимо, що ми обчислили x_t^{opt} та $\langle e_t^2 \rangle$. Нам треба обчислити x_{t+1}^{opt} та $\langle e_{t+1}^2 \rangle$. Шукаємо x_{t+1}^{opt} в такому вигляді:

$$x_{t+1}^{\text{opt}} = Kz_{t+1} + (1 - Kc)(ax_t^{\text{opt}} + bu_t),\tag{2.9}$$

де K – невідома константа, яка називається коефіцієнтом Калмана. Коефіцієнт Калмана може бути різним на різних кроках алгоритму. Як зазначено вище, має бути мінімізована середньоквадратична помилка фільтру

$$\langle (x_{t+1} - x_{t+1}^{\text{opt}})^2 \rangle = \langle e_{t+1}^2 \rangle \rightarrow \min.\tag{2.10}$$

Коефіцієнт Калмана шукаємо на основі умови (2.10). На основі (2.5), (2.6) та (2.9) отримаємо

$$\begin{aligned}e_{t+1} &= x_{t+1} - x_{t+1}^{\text{opt}} = ax_t + bu_t + \xi_t - Kz_{t+1} - (1 - Kc)(ax_t^{\text{opt}} + bu_t) = \\ &= ax_t + bu_t + \xi_t - K(cx_{t+1} + \eta_{t+1}) - (1 - Kc)(ax_t^{\text{opt}} + bu_t) = \\ &= ax_t + bu_t + \xi_t - K(c(ax_t + bu_t + \xi_t) + \eta_{t+1}) - (1 - Kc)(ax_t^{\text{opt}} + bu_t),\end{aligned}\tag{2.11}$$

що після розкриття дужок з урахуванням (2.6) дає

$$e_{t+1} = (1 - Kc)ae_t + (1 - Kc)\xi_t - \eta_{t+1}K,\tag{2.12}$$

звідки

$$e_{t+1}^2 = (1 - Kc)^2 a^2 e_t^2 + (1 - Kc)^2 \xi_t^2 + \eta_{t+1}^2 K^2 + G_t\tag{2.13}$$

де G_t є сумою так званих «перехресних членів»:

$$G_t = 2a(1 - Kc)^2 \xi_t e_t - 2a(1 - Kc)K e_t \eta_{t+1} - 2(1 - Kc)K \xi_t \eta_{t+1}.\tag{2.14}$$

Слід зауважити, що в цих викладках випадковими процесами є лише η_t , e_t , ξ_t , тож на основі (2.13) та (2.14) отримаємо

$$\begin{aligned}\langle G_t \rangle &= 2a(1 - Kc)^2 \langle \xi_t e_t \rangle - 2a(1 - Kc)K \langle e_t \eta_{t+1} \rangle - 2(1 - Kc)K \langle \xi_t \eta_{t+1} \rangle = 0, \\ \langle e_{t+1}^2 \rangle &= (1 - Kc)^2 a^2 \langle e_t^2 \rangle + (1 - Kc)^2 \langle \xi_t^2 \rangle + K^2 \langle \eta_{t+1}^2 \rangle.\end{aligned}\tag{2.15}$$

Звідси стає зрозумілим, чому ми шукаємо x_{t+1}^{opt} у вигляді (2.9). Вираз (2.12) є сумою трьох незалежних доданків, що зумовлює те, що «перехресні» члени не дадуть внеску до величини $\langle e_{t+1}^2 \rangle$.

На основі того, що η_t та ξ_t є стаціонарними процесами з нульовими середніми значеннями, отримаємо

$$\sigma_{\xi}^2 = \langle \xi_t^2 \rangle - \langle \xi_t \rangle^2 = \langle \xi_t^2 \rangle, \quad \sigma_{\eta}^2 = \langle \eta_t^2 \rangle - \langle \eta_t \rangle^2 = \langle \eta_t^2 \rangle = \langle \eta_{t+1}^2 \rangle, \quad (2.16)$$

що дає змогу переписати другий вираз у (2.15) як

$$\langle e_{t+1}^2 \rangle = (1 - Kc)^2 a^2 \langle e_t^2 \rangle + (1 - Kc)^2 \sigma_{\xi}^2 + K^2 \sigma_{\eta}^2. \quad (2.17)$$

Перед нами стоїть задача мінімізувати вираз $\langle e_{t+1}^2 \rangle$, тож

$$\frac{\partial \langle e_{t+1}^2 \rangle}{\partial K} = 0 \Rightarrow K = c \frac{a^2 \langle e_t^2 \rangle + \sigma_{\xi}^2}{\sigma_{\eta}^2 + c^2 a^2 \langle e_t^2 \rangle + c^2 \sigma_{\xi}^2}. \quad (2.18)$$

Вихід фільтру x_{t+1}^{opt} обчислюється на основі (2.9), де значення коефіцієнту Калмана K задається виразом (2.18).

Також нам треба обчислити значення $\langle e_{t+1}^2 \rangle$. На основі (2.17) та (2.18) прямим обчисленням після певних перетворень можна показати, що

$$\langle e_{t+1}^2 \rangle = \sigma_{\eta}^2 \frac{a^2 \langle e_t^2 \rangle + \sigma_{\xi}^2}{\sigma_{\eta}^2 + c^2 \sigma_{\xi}^2 + c^2 a^2 \langle e_t^2 \rangle}. \quad (2.19)$$

Очевидно, нам потрібні початкові значення для виходу фільтру x^{opt} та для середньоквадратичної помилки фільтру $\langle e^2 \rangle$. Їх зазвичай обирають як

$$\langle e_0^2 \rangle = \sigma_{\eta}^2, \quad x_0^{\text{opt}} = z_0. \quad (2.20)$$

Для того, щоб підсумувати вище сказане, давайте ще раз проговоримо алгоритм повністю. Спершу, задаємо початкові значення виходу фільтру x^{opt} та середньоквадратичної помилки фільтру $\langle e^2 \rangle$ (зазвичай, згідно (2.20)). На основі значень $\langle e_0^2 \rangle$ та x_0^{opt} обчислюємо $\langle e_1^2 \rangle$ та x_1^{opt} :

$$x_1^{\text{opt}} = Kz_1 + (1 - Kc)(ax_0^{\text{opt}} + bu_0), \quad K = c \frac{a^2 \langle e_0^2 \rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2 a^2 \langle e_0^2 \rangle + c^2 \sigma_\xi^2}; \quad (2.21)$$

$$\langle e_1^2 \rangle = \sigma_\eta^2 \frac{a^2 \langle e_0^2 \rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2 \sigma_\xi^2 + c^2 a^2 \langle e_0^2 \rangle}.$$

Надалі обчислюємо $\langle e_2^2 \rangle$ та x_2^{opt} на основі значень x_1^{opt} та $\langle e_1^2 \rangle$, і так далі. На n -му кроці алгоритму, $\langle e_n^2 \rangle$ та x_n^{opt} обчислюються на основі значень x_{n-1}^{opt} та $\langle e_{n-1}^2 \rangle$ на основі виразів

$$x_n^{\text{opt}} = K_n z_n + (1 - K_n c)(ax_{n-1}^{\text{opt}} + bu_{n-1}), \quad K_n = c \frac{a^2 \langle e_{n-1}^2 \rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2 a^2 \langle e_{n-1}^2 \rangle + c^2 \sigma_\xi^2}; \quad (2.22)$$

$$\langle e_n^2 \rangle = \sigma_\eta^2 \frac{a^2 \langle e_{n-1}^2 \rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2 \sigma_\xi^2 + c^2 a^2 \langle e_{n-1}^2 \rangle},$$

тут ми дописати індекс n до величини K_n з метою підкреслити, що коефіцієнт Калмана може бути різним на різних кроках алгоритму.

Як бачимо, на кожному кроці алгоритм заново обчислює («адаптує») коефіцієнт Калмана, на основі якого надалі обчислюється вихід фільтру. Тобто на кожному кроці алгоритм адаптує свої налаштування. Фільтри, які побудовані на основі таких алгоритмів, називаються *адаптивними*.

2.2. Векторний дискретний фільтр Калмана

В попередньому параграфі нами розглянуто випадок, коли система описується одним параметром. Мета векторного фільтра Калмана точно така ж, як і скалярного, але векторний фільтр Калмана застосовний у випадку, коли кількість параметрів, що описують систему, є більша за один.

Розглянемо систему, що описується n параметрами $x_1, x_2, \dots, x_n, n > 1$. Ці параметри змінюються дискретно. Вектор стану системи – це вектор-стовпець, що містить всі параметри, які описують систему:

$$x_t = \begin{pmatrix} x_{t,1} \\ x_{t,2} \\ \vdots \\ x_{t,n} \end{pmatrix}, \quad (2.23)$$

індекс t описує величину, взяту в момент часу t .

Для подальших обчислень нам знадобиться поняття *матриці коваріації*. Матрицею коваріації (позначення $\text{cov}(x)$) випадкового вектору-стовпця x є матриця, компоненти якої визначені як

$$\text{cov}(x)_{ij} = \text{cov}(x)_{ji} = \langle x_i x_j \rangle. \quad (2.24)$$

Надалі опишемо дискретний векторний фільтр Калмана на основі аналогії з дискретним скалярним фільтром, див. таблицю 2.1.

Таблиця 2.1. Аналогія між векторним та скалярним фільтрами Калмана

Скалярний фільтр	Векторний фільтр
<i>Зміна параметру з часом:</i> $x_{t+1} = Ax_t + Bu_t + \xi_t,$ x – параметр, що описує систему, ξ_t – випадковий процес, що описує випадковий характер еволюції системи, u_t – відома величина, що керує еволюцією системи; A, B – задані константи.	<i>Зміна параметрів з часом:</i> $x_{t+1} = Ax_t + Bu_t + \xi_t, \quad (2.25)$ x – вектор стану системи, фізичний зміст всіх інших величин такий же, як і в скалярному фільтрі, однак слід зауважити, що A та B – матриці $n \times n$ та x_t, u_t, ξ_t – вектори-стовпці. Рівняння (2.25) написано в матричній формі.
<i>Покази приладу:</i> $z_t = Cx_t + \eta_t$ z_t – покази приладу, η_t – похибка приладу, константа C – коефіцієнт підсилення приладу.	<i>Покази приладу:</i> $z_t = Cx_t + \eta_t, \quad (2.26)$ фізичний зміст всіх величин такий же, як і в скалярному фільтрі, однак слід зауважити, що C є матрицею $n \times n$, та x_t, z_t, η_t –

	вектори-стовпці. Рівняння (2.26) написано в матричній формі.
<i>Помилка фільтра та його мета:</i> $e_t = x_t - x_t^{\text{opt}}, \quad \forall t \quad \langle e_t^2 \rangle \rightarrow \min.$	<i>Помилка фільтра та його мета:</i> $e_t = x_t - x_t^{\text{opt}}, \quad \forall t \quad \sum_{j=1}^n \langle e_{t,j}^2 \rangle \rightarrow \min, \quad (2.27)$ e_t – вектор-стовпець.
<i>Припущення теорії:</i> $\sigma_{\xi}^2 \text{ та } \sigma_{\eta}^2 \text{ є відомими,}$ $\langle \xi_t \rangle = 0, \quad \langle \eta_t \rangle = 0;$ $\forall t_1, t_2 \quad \langle \xi_{t_1} e_{t_2} \rangle = \langle \xi_{t_1} \rangle \langle e_{t_2} \rangle = 0,$ $\langle \eta_{t_1} e_{t_2} \rangle = \langle \eta_{t_1} \rangle \langle e_{t_2} \rangle = 0,$ $\langle \xi_{t_1} \eta_{t_2} \rangle = \langle \xi_{t_1} \rangle \langle \eta_{t_2} \rangle = 0.$	<i>Припущення теорії:</i> $\text{cov}(\xi_t) \text{ та } \text{cov}(\eta_t) \text{ є відомими}$ $\forall i = 1, 2, \dots, n \quad \langle \xi_{t,i} \rangle = 0 \quad \langle \eta_{t,i} \rangle = 0; \quad (2.28)$ $\forall i, j = 1, 2, \dots, n \quad \forall t_1, t_2$ $\langle \xi_{t_1,i} e_{t_2,j} \rangle = \langle \xi_{t_1,i} \rangle \langle e_{t_2,j} \rangle = 0,$ $\langle \eta_{t_1,i} e_{t_2,j} \rangle = \langle \eta_{t_1,i} \rangle \langle e_{t_2,j} \rangle = 0, \quad (2.29)$ $\langle \xi_{t_1,i} \eta_{t_2,j} \rangle = \langle \xi_{t_1,i} \rangle \langle \eta_{t_2,j} \rangle = 0.$
<i>Рекурентне співвідношення для виходу фільтра:</i> $x_{t+1}^{\text{opt}} = Kz_{t+1} + (1 - KC)(ax_t^{\text{opt}} + bu_t)$, коефіцієнт Калмана K шукаємо згідно умови $\langle e_{t+1}^2 \rangle \rightarrow \min$	<i>Рекурентне співвідношення для виходу фільтра:</i> $x_{t+1}^{\text{opt}} = Kz_{t+1} + (I - KC)(Ax_t^{\text{opt}} + Bu_t) \quad (2.30)$, I – одинична матриця, коефіцієнт Калмана K є матрицею $n \times n$, та має бути знайдений на основі умови $\sum_{j=1}^n \langle e_{t+1,j}^2 \rangle \rightarrow \min. \quad (2.31)$ Рекурентне співвідношення (2.31) виписане в матричній формі.

Слід зауважити, що в загальному випадку матриці A , B , C можуть залежати від часу, але для простоти розглянемо поширений випадок, де вони є постійними.

Розглянемо величину $e_{t+1}^T e_{t+1}$; тут та в подальшому індекс T означає транспонування. Маємо

$$e_{t+1}^T e_{t+1} = \begin{pmatrix} e_{t+1,1} & e_{t+1,2} & \dots & e_{t+1,n} \end{pmatrix} \begin{pmatrix} e_{t+1,1} \\ e_{t+1,2} \\ \vdots \\ e_{t+1,n} \end{pmatrix} = e_{t+1,1}^2 + e_{t+1,2}^2 + \dots + e_{t+1,n}^2 = \sum_{j=1}^n e_{t+1,j}^2, \quad (2.32)$$

тож бачимо, що умова (2.31) може бути переписана як

$$\langle e_{t+1}^T e_{t+1} \rangle \rightarrow \min. \quad (2.33)$$

Аналогічно до виведення формули (2.12), на основі виразів (2.25)–(2.27) та (2.30) отримаємо

$$e_{t+1} = (A - KCA)e_t + (I - KC)\xi_t - K\eta_{t+1}, \quad (2.34)$$

під час виведення цього виразу та в подальшому слід не забувати, що у матричних добутках не можна переставляти матриці місцями!

На основі відомих властивостей

$$(AB)^T = B^T A^T, \quad (A + B)^T = A^T + B^T \quad (2.35)$$

та формули (2.34) отримаємо

$$e_{t+1}^T = e_t^T (A - KCA)^T + \xi_t^T (I - KC)^T - \eta_{t+1}^T K^T. \quad (2.36)$$

На основі (2.36) та (2.34) виразимо

$$\begin{aligned} e_{t+1}^T e_{t+1} &= e_t^T (A - KCA)^T (A - KCA)e_t + \\ &+ \xi_t^T (I - KC)^T (I - KC)\xi_t + \eta_{t+1}^T K^T K\eta_{t+1} + G_t, \end{aligned} \quad (2.37)$$

Де G_t є сумою «перехресних» членів

$$\begin{aligned} G_t &= e_t^T (A - KCA)^T (I - KC)\xi_t - e_t^T (A - KCA)^T K\eta_{t+1} + \\ &+ \xi_t^T (I - KC)^T (A - KCA)e_t - \xi_t^T (I - KC)^T K\eta_{t+1} - \\ &- \eta_{t+1}^T K^T (A - KCA)e_t - \eta_{t+1}^T K^T (I - KC)\xi_t. \end{aligned} \quad (2.38)$$

Перш за все, зауважимо що G_t – це число, а не матриця. Розглянемо перший член у (2.38). e_t^T є матрицею $1 \times n$, $(A - KCA)^T$ є матрицею $n \times n$, $(I - KC)$ є

матрицею $n \times n$ та ξ_t є матрицею $n \times 1$, тож їхній добуток має структуру $(1 \times n) \cdot (n \times n) \cdot (n \times n) \cdot (n \times 1) = 1 \times 1$. Аналогічно можна показати, що кожен член у (2.38) є числом. Більше того, аналогічно можна показати, що число є кожен член у (2.37).

Розглянемо середнє значення першого члену у правій частині (2.38). Згідно відомих правил підсумовування Ейнштейна маємо

$$e_i^T (A - KCA)^T (I - KC) \xi_t = e_{t,i}^T (A - KCA)_{ij}^T (I - KC)_{jk} \xi_{t,k}, \quad (2.39)$$

тут і надалі за індексами, що повторюються, проводиться підсумовування від 1 до n . На основі (2.39) отримаємо

$$\begin{aligned} \left\langle e_i^T (A - KCA)^T (I - KC) \xi_t \right\rangle &= \left\langle e_{t,i}^T (A - KCA)_{ij}^T (I - KC)_{jk} \xi_{t,k} \right\rangle = \\ &= (A - KCA)_{ij}^T (I - KC)_{jk} \left\langle e_{t,i}^T \xi_{t,k} \right\rangle = 0, \end{aligned} \quad (2.40)$$

тут враховано (2.29). Аналогічно можна показати, що середні значення кожного доданку в G_t є нульовими, тож згідно (2.37) та (2.33)

$$\begin{aligned} \left\langle e_{t+1}^T e_{t+1} \right\rangle &= f(K) = \left\langle e_t^T (A - KCA)^T (A - KCA) e_t \right\rangle + \\ &+ \left\langle \xi_t^T (I - KC)^T (I - KC) \xi_t \right\rangle + \left\langle \eta_{t+1}^T K^T K \eta_{t+1} \right\rangle \rightarrow \min. \end{aligned} \quad (2.41)$$

Результат (2.41) обґрунтовує те, що x_{t+1}^{opt} шукаємо у вигляді (2.30), бо після усереднення «зникають» всі «перехресні» члени.

Залишилось мінімізувати функцію від матриці $f(K)$. Використаємо ідею, аналогічну до тої, на основі якої мінімізувався функціонал (1.61). Нехай K – матриця, що мінімізує функцію $f(K)$. Тоді якщо ми візьмемо іншу матрицю у вигляді $K + \varepsilon k$, де ε – довільне число та k – довільна матриця $n \times n$, то має виконатись нерівність

$$\forall \varepsilon, \forall k \quad f(K) \leq f(K + \varepsilon k), \quad (2.42)$$

Згідно (2.41) отримаємо

$$f(K + \varepsilon k) - f(K) = \left\langle e_t^T (A - (K + \varepsilon k)CA)^T (A - (K + \varepsilon k)CA) e_t \right\rangle + \quad (2.43)$$

$$\begin{aligned}
& + \left\langle \xi_t^T (I - (K + \varepsilon k)C)^T (I - (K + \varepsilon k)C) \xi_t \right\rangle + \\
& + \left\langle \eta_{t+1}^T (K + \varepsilon k)^T (K + \varepsilon k) \eta_{t+1} \right\rangle - \left\langle e_t^T (A - KCA)^T (A - KCA) e_t \right\rangle - \\
& - \left\langle \xi_t^T (I - KC)^T (I - KC) \xi_t \right\rangle - \left\langle \eta_{t+1}^T K^T K \eta_{t+1} \right\rangle,
\end{aligned}$$

та після розкриття дужок

$$\begin{aligned}
\forall \varepsilon, \forall k \quad f(K + \varepsilon k) - f(K) &= \varepsilon \left(- \left\langle e_t^T (kCA)^T (A - KCA) e_t \right\rangle - \right. \\
& - \left\langle e_t^T (A - KCA)^T kCA e_t \right\rangle - \left\langle \xi_t^T (I - KC)^T kC \xi_t \right\rangle - \\
& - \left\langle \xi_t^T (kC)^T (I - KC) \xi_t \right\rangle + \left\langle \eta_{t+1}^T k^T K \eta_{t+1} \right\rangle + \left\langle \eta_{t+1}^T K^T k \eta_{t+1} \right\rangle \Big) + \\
& + \varepsilon^2 \left(\left\langle (kCA e_t)^T (kCA e_t) \right\rangle + \left\langle (kC \xi_t)^T (kC \xi_t) \right\rangle + \left\langle (k \eta_{t+1})^T (k \eta_{t+1}) \right\rangle \right) \geq 0.
\end{aligned} \tag{2.44}$$

Спершу, розглянемо члени при ε^2 . k , C та A – матриці $n \times n$, e_t – матриця $n \times 1$, тож $kCAe_t$ є вектором-стовпцем (матрицею $n \times 1$):

$$kCAe_t = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \Rightarrow (kCAe_t)^T kCAe_t = \sum_{j=1}^n b_j^2 \geq 0. \tag{2.45}$$

Аналогічно можна довести невід’ємність інших членів при ε^2 .

Розглянемо члени при ε в (2.44). Вони також мають бути невід’ємними, тому що нерівність (2.44) має виконуватись для довільних ε та k , а, наприклад, коли ε – малий параметр, то саме члени при ε , а не при ε^2 , є головними у (2.44). Можна показати, що кожен член при ε числом, а не матрицею. Позначимо

$$\begin{aligned}
a_1(k, K) &= - \left\langle e_t^T (kCA)^T (A - KCA) e_t \right\rangle, \\
a_2(k, K) &= - \left\langle e_t^T (A - KCA)^T kCA e_t \right\rangle, \\
a_3(k, K) &= - \left\langle \xi_t^T (I - KC)^T kC \xi_t \right\rangle, \quad a_4(k, K) = - \left\langle \xi_t^T (kC)^T (I - KC) \xi_t \right\rangle, \\
a_5(k, K) &= \left\langle \eta_{t+1}^T k^T K \eta_{t+1} \right\rangle, \quad a_6(k, K) = \left\langle \eta_{t+1}^T K^T k \eta_{t+1} \right\rangle.
\end{aligned} \tag{2.46}$$

Тоді

$$\forall \varepsilon, \forall k \quad \varepsilon \sum_{\alpha=1}^6 a_{\alpha}(k, K) \geq 0. \quad (2.47)$$

Очевидно, що

$$a_1(k, K) = a_1^T(k, K), \quad (2.48)$$

бо $a_1(k, K)$ – число. В той же час, згідно (2.46) отримаємо

$$a_1^T(k, K) = -\left\langle e_t^T (A - KCA)^T kCAe_t \right\rangle = a_2(k, K), \quad (2.49)$$

тож

$$a_1(k, K) = a_2(k, K). \quad (2.50)$$

Аналогічно можна показати, що

$$a_3(k, K) = a_4(k, K), \quad a_5(k, K) = a_6(k, K). \quad (2.51)$$

Тоді згідно (2.47), (2.50), (2.51)

$$\forall \varepsilon, \forall k \quad \varepsilon (a_1(k, K) + a_3(k, K) + a_5(k, K)) \geq 0. \quad (2.52)$$

Число ε може бути як додатним, так і від'ємним, тож умова (2.52) може бути виконана лише, коли сума у дужках дорівнює нулю:

$$-\left\langle e_t^T (kCA)^T (A - KCA)e_t \right\rangle - \left\langle \xi_t^T (I - KC)^T kC\xi_t \right\rangle + \left\langle \eta_{t+1}^T k^T K\eta_{t+1} \right\rangle = 0, \quad (2.53)$$

тут використано явні вирази (2.46).

Розглянемо вираз $\left\langle \eta_{t+1}^T Q \eta_{t+1} \right\rangle$, де Q – матриця $n \times n$, компоненти якої не є випадковими величинами. На основі правила підсумовування Ейнштейна отримаємо

$$\begin{aligned} \left\langle \eta_{t+1}^T Q \eta_{t+1} \right\rangle &= \left\langle \eta_{t+1,i}^T Q_{ij} \eta_{t+1,j} \right\rangle = Q_{ij} \left\langle \eta_{t+1,i}^T \eta_{t+1,j} \right\rangle = Q_{ij} \left\langle \eta_{t+1,j} \eta_{t+1,i} \right\rangle = \\ &= Q_{ij} \text{cov}(\eta_{t+1})_{ji} = \text{Tr}(Q \cdot \text{cov}(\eta_{t+1})), \end{aligned} \quad (2.54)$$

тут і надалі $\text{Tr}A$ позначає слід матриці A . Аналогічно

$$\left\langle \xi_t^T Q \xi_t \right\rangle = \text{Tr}(Q \cdot \text{cov}(\xi_t)), \quad \left\langle e_t^T Q e_t \right\rangle = \text{Tr}(Q \cdot \text{cov}(e_t)). \quad (2.55)$$

Згідно (2.54) та (2.55) перепишемо (2.53):

$$\begin{aligned} \forall k \quad & -\text{Tr}\left((kCA)^T (A - KCA) \text{cov}(e_t)\right) - \text{Tr}\left((kC)^T (I - KC) \text{cov}(\xi_t)\right) + \\ & + \text{Tr}\left(k^T K \text{cov}(\eta_{t+1})\right) = 0. \end{aligned} \quad (2.56)$$

На основі відомих властивостей

$$\text{Tr}(AB) = \text{Tr}(BA), \quad \text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B) \quad (2.57)$$

вираз (2.56) може бути переписаний як

$$\begin{aligned} \forall k \quad & \text{Tr}\left[\left(-(A - KCA) \text{cov}(e_t)(CA)^T - \right. \right. \\ & \left. \left. -(I - KC) \text{cov}(\xi_t)C^T + K \text{cov}(\eta_{t+1})\right)k^T\right] = 0, \end{aligned} \quad (2.58)$$

що виконується лише при

$$-(A - KCA) \text{cov}(e_t)(CA)^T - (I - KC) \text{cov}(\xi_t)C^T + K \text{cov}(\eta_{t+1}) = 0. \quad (2.59)$$

На основі (2.59) прямим обчисленням можна показати, що

$$K = P_t C^T (C P_t C^T + \text{cov}(\eta_{t+1}))^{-1}, \quad P_t = A \text{cov}(e_t) A^T + \text{cov}(\xi_t). \quad (2.60)$$

Також дуже важливим є обчислити $\text{cov}(e_{t+1})$, тому що згідно (2.60) та (2.30), на наступному кроці x_{t+2}^{opt} буде залежати від $\text{cov}(e_{t+1})$.

На основі (2.34) матимемо

$$e_{t+1,i} = (A - KCA)_{ij} e_{t,j} + (I - KC)_{ij} \xi_{t,j} - K_{ij} \eta_{t+1,j}, \quad (2.61)$$

звідки з урахуванням (2.24) та (2.29)

$$\begin{aligned} \text{cov}(e_{t+1})_{il} &= \langle e_{t+1,i} e_{t+1,l} \rangle = (A - KCA)_{ij} \text{cov}(e_t)_{jk} (A - KCA)_{kl}^T + \\ &+ (I - KC)_{ij} \text{cov}(\xi_t)_{jk} (I - KC)_{kl}^T + K_{ij} \text{cov}(\eta_{t+1})_{jk} K_{kl}^T \Rightarrow \\ &\Rightarrow \text{cov}(e_{t+1}) = (A - KCA) \text{cov}(e_t) (A - KCA)^T + \\ &+ (I - KC) \text{cov}(\xi_t) (I - KC)^T + K \text{cov}(\eta_{t+1}) K^T. \end{aligned} \quad (2.62)$$

Розкриємо вираз $(I - KC)^T$ та відповідні дужки:

$$\text{cov}(e_{t+1}) = (I - KC)P_t - P_t C^T K^T + K C P_t C^T K^T + K \text{cov}(\eta_{t+1}) K^T. \quad (2.63)$$

Перезберемо (2.63):

$$\text{cov}(e_{t+1}) = (I - KC)P_t - P_t C^T K^T + K(CP_t C^T + \text{cov}(\eta_{t+1}))K^T, \quad (2.64)$$

та підставивши вираз для K в останній доданок у (2.64), остаточно отримаємо

$$\text{cov}(e_{t+1}) = (I - KC)P_t. \quad (2.65)$$

Нам також потрібні початкові значення для $\text{cov}(e)$ та x^{opt} . Зазвичай їх обирають таким чином:

$$\text{cov}(e_0) = \text{cov}(\eta_0), \quad x_0^{\text{opt}} = z_0. \quad (2.66)$$

Для того, щоб підсумувати вище сказане, давайте ще раз проговоримо алгоритм повністю. Спершу задаємо початкові значення виходу фільтру x^{opt} та матриці коваріації помилки фільтру $\text{cov}(e_0)$ (зазвичай, згідно (2.66)). Надалі на кожному кроці вихід фільтру та матриця коваріації фільтру обчислюються за рекурентними формулами

$$\begin{aligned} x_{t+1}^{\text{opt}} &= K_{t+1} z_{t+1} + (I - K_{t+1} C) (A x_t^{\text{opt}} + B u_t), \\ K_{t+1} &= P_t C^T (C P_t C^T + \text{cov}(\eta_{t+1}))^{-1}, \quad P_t = A \text{cov}(e_t) A^T + \text{cov}(\xi_t), \\ \text{cov}(e_{t+1}) &= (I - K_{t+1} C) P_t, \end{aligned} \quad (2.67)$$

тут ми дописали індекс $t + 1$ до величини K_{t+1} , щоб підкреслити, що коефіцієнт Калмана може бути різним на різних кроках алгоритму. Зауважимо, що якщо матриці A , B , C залежать від часу, то виведення формул є точно таким же, та результат матиме вигляд

$$\begin{aligned} x_{t+1}^{\text{opt}} &= K_{t+1} z_{t+1} + (I - K_{t+1} C_t) (A_t x_t^{\text{opt}} + B_t u_t), \\ K_{t+1} &= P_t C_t^T (C_t P_t C_t^T + \text{cov}(\eta_{t+1}))^{-1}, \quad P_t = A_t \text{cov}(e_t) A_t^T + \text{cov}(\xi_t), \\ \text{cov}(e_{t+1}) &= (I - K_{t+1} C_t) P_t. \end{aligned} \quad (2.68)$$

2.3. Скалярний фільтр Калмана – Бюсі

За своїм фізичним змістом скалярний фільтр Калмана–Бюсі є аналогічним до скалярного фільтра Калмана, але на відміну від нього, є застосовним до неперервних, а не до дискретних систем.

Нехай система описується одним параметром x , який змінюється неперервно, та “ідеальний” закон зміни цього параметру є таким:

$$\frac{dx(t)}{dt} = a(t)x(t) + b(t)u(t), \quad (2.69)$$

де $u(t)$ – задана величина, яка керує еволюцією системи; $a(t)$ та $b(t)$ є відомим функціями. Параметр $x(t)$ вимірюється приладом, покази якого дорівнюють $z(t)$. Нехай $\eta(t)$ – випадковий стаціонарний процес, що є похибкою приладу та $\xi(t)$ – випадковий стаціонарний процес, що описує випадковий характер еволюції системи:

$$\frac{dx(t)}{dt} = a(t)x(t) + b(t)u(t) + \xi(t), \quad z(t) = c(t)x(t) + \eta(t), \quad (2.70)$$

$c(t)$ – відома функція, що є коефіцієнтом підсилення приладу. Входом фільтра є функція $z(t)$, мета фільтра – згенерувати на виході функцію $x^{\text{opt}}(t)$, найближчу до $x(t)$. Слово “найближчу” слід розуміти як “таку, що має найменше середньоквадратичне відхилення”.

Помилка фільтра вводиться як

$$e(t) = x(t) - x^{\text{opt}}(t), \quad (2.71)$$

метою фільтра є мінімізувати вираз

$$\forall t \quad \langle e^2(t) \rangle \rightarrow \min. \quad (2.72)$$

Припущення теорії є такими:

1. Кореляційні функції процесів $\eta(t)$ та $\xi(t)$ мають вигляд

$$R_{\xi}(t, \tau) = Q(t)\delta(t - \tau), \quad R_{\eta}(t, \tau) = R(t)\delta(t - \tau), \quad (2.73)$$

де $Q(t)$ та $R(t)$ – відомі функції. Фактично, робиться припущення, що процеси $\eta(t)$ та $\xi(t)$ є білими шумами.

2. Середні значення процесів $\eta(t)$ та $\xi(t)$ дорівнюють нулю:

$$\forall t \quad \langle \xi(t) \rangle = 0, \langle \eta(t) \rangle = 0. \quad (2.74)$$

3. Процеси $\eta(t)$, $\xi(t)$ та $e(t)$ є незалежними:

$$\begin{aligned} \forall t_1, t_2 \quad \langle \xi(t_1)\eta(t_2) \rangle &= \langle \xi(t_1) \rangle \langle \eta(t_2) \rangle = 0, \langle \xi(t_1)e(t_2) \rangle = \langle \xi(t_1) \rangle \langle e(t_2) \rangle = 0, \\ \langle e(t_1)\eta(t_2) \rangle &= \langle e(t_1) \rangle \langle \eta(t_2) \rangle = 0. \end{aligned} \quad (2.75)$$

Часову похідну від $x^{\text{opt}}(t)$ шукаємо у вигляді

$$\frac{dx^{\text{opt}}(t)}{dt} = K(t) \{z(t) - c(t)x^{\text{opt}}(t)\} + a(t)x^{\text{opt}}(t) + b(t)u(t), \quad (2.76)$$

де невідома функція $K(t)$ є коефіцієнтом Калмана, яку шукатимемо на основі умови (2.72). На основі (2.70), (2.71) та (2.76) отримаємо

$$\frac{de(t)}{dt} = \frac{dx(t)}{dt} - \frac{dx^{\text{opt}}(t)}{dt} = \{a(t) - K(t)c(t)\}e(t) + \xi(t) - K(t)\eta(t). \quad (2.77)$$

Проінтегруємо обидві частини у (2.77):

$$\int_t^{t+dt} d\tau \frac{de(\tau)}{d\tau} = \int_t^{t+dt} d\tau \{a(\tau) - K(\tau)c(\tau)\}e(\tau) + \int_t^{t+dt} d\tau \xi(\tau) - \int_t^{t+dt} d\tau K(\tau)\eta(\tau), \quad (2.78)$$

тут $dt > 0$ є дуже малим приростом часу. Очевидно, що

$$\int_t^{t+dt} d\tau \frac{de(\tau)}{d\tau} = e(t+dt) - e(t). \quad (2.79)$$

Розглянемо вираз $\int_t^{t+dt} d\tau \varphi(\tau)$. Функція $\varphi(\tau)$ може бути виражена в термінах поліному Тейлора в околі точки t :

$$\varphi(\tau) = \varphi(t) + \left. \frac{d\varphi(\tau)}{d\tau} \right|_{\tau=t} (t - \tau) + o(t - \tau). \quad (2.80)$$

Згідно (2.80) матимемо

$$\begin{aligned}
& \int_t^{t+dt} \varphi(\tau) d\tau = \varphi(t) \int_t^{t+dt} d\tau + \left. \frac{d\varphi(\tau)}{d\tau} \right|_{\tau=t} \int_t^{t+dt} d\tau(t-\tau) + \int_t^{t+dt} d\tau o(t-\tau) = \\
& = \left\{ \begin{array}{l} \tau' = \tau - t, d\tau' = d\tau, \\ \tau = t \rightarrow \tau' = 0, \\ \tau = t + dt \rightarrow \tau' = dt \end{array} \right\} = \varphi(t) dt - \left. \frac{d\varphi(\tau)}{d\tau} \right|_{\tau=t} \int_0^{dt} d\tau' \tau' + o(dt^2) = \\
& = \varphi(t) dt - \left. \frac{d\varphi(\tau)}{d\tau} \right|_{\tau=t} \frac{dt^2}{2} + o(dt^2) = \varphi(t) dt + O(dt^2),
\end{aligned} \tag{2.81}$$

звідки

$$\int_t^{t+dt} d\tau \{a(\tau) - K(\tau)c(\tau)\} e(\tau) = \{a(t) - K(t)c(t)\} e(t) dt + O(dt^2). \tag{2.82}$$

Зауважимо, що функції $a(\tau)$, $K(\tau)$, $c(\tau)$, $e(\tau)$ вважаються хоча би 1 раз диференційованими. Функції $\xi(\tau)$ та $\eta(\tau)$ можуть бути не диференційованими через свою поведінку аналогічну до білого шуму, тож до них не можна застосувати викладки (2.81). Згідно (2.78), (2.79) та (2.82) отримаємо

$$\begin{aligned}
e(t+dt) = & \left[1 + \{a(t) - K(t)c(t)\} dt \right] e(t) + \int_t^{t+dt} d\tau \xi(\tau) - \\
& - \int_t^{t+dt} d\tau K(\tau) \eta(\tau) + O(dt^2).
\end{aligned} \tag{2.83}$$

Зауважимо, що перші три члени в правій частині виразу (2.83) є незалежними, бо в цьому виразі випадковими процесами є лише $e(t)$, $\xi(\tau)$ та $\eta(\tau)$, а ці процеси є незалежними згідно вище наведеного припущення теорії. Аналогічно до викладок в минулих параграфах, після возведення (2.83) у квадрат та усереднення «перехресні» члени «знищуються», та отримаємо

$$\begin{aligned}
\langle e^2(t+dt) \rangle = & \left[1 + \{a(t) - K(t)c(t)\} dt \right]^2 \langle e^2(t) \rangle + \left\langle \left(\int_t^{t+dt} d\tau \xi(\tau) \right)^2 \right\rangle + \\
& + \left\langle \left(\int_t^{t+dt} d\tau K(\tau) \eta(\tau) \right)^2 \right\rangle + O(dt^2).
\end{aligned} \tag{2.84}$$

Згідно (1.11), (2.73) та (2.81) матимемо

$$\begin{aligned} \left\langle \left(\int_t^{t+dt} d\tau \xi(\tau) \right)^2 \right\rangle &= \int_t^{t+dt} d\tau \int_t^{t+dt} d\tau' \langle \xi(\tau) \xi(\tau') \rangle = \int_t^{t+dt} d\tau \int_t^{t+dt} d\tau' R_\xi(\tau, \tau') = \\ &= \int_t^{t+dt} d\tau \int_t^{t+dt} d\tau' Q(\tau) \delta(\tau - \tau') = \int_t^{t+dt} d\tau Q(\tau) = Q(t)dt + O(dt^2). \end{aligned} \quad (2.85)$$

Аналогічно обчислимо

$$\left\langle \left(\int_t^{t+dt} d\tau K(\tau) \eta(\tau) \right)^2 \right\rangle = K^2(t)R(t)dt + O(dt^2), \quad (2.86)$$

вважається, що $Q(t)$ та $R(t)$ є диференційованими хоча б один раз.

На основі (2.84)–(2.86) отримаємо

$$\begin{aligned} \langle e^2(t+dt) \rangle &= f[K(t)] = [1 + 2\{a(t) - K(t)c(t)\}dt] \langle e^2(t) \rangle + Q(t)dt + \\ &+ K^2(t)R(t)dt + O(dt^2). \end{aligned} \quad (2.87)$$

Метою фільтра є мінімізувати $\langle e^2(t+dt) \rangle$, тобто

$$f[K(t)] \rightarrow \min. \quad (2.88)$$

Використаємо ідею, аналогічну до тої, яка використовувалась для мінімізації функціоналу (1.61). Нехай функція $K(t)$ мінімізує функціонал $f[K(t)]$. Тоді якщо взяти якусь іншу функцію у вигляді $K(t) + \varepsilon g(t)$, де ε – довільне число та $g(t)$ – довільна функція, то має виконатись така нерівність:

$$\forall \varepsilon, \forall g(t) \quad f[K(t)] \leq f[K(t) + \varepsilon g(t)]. \quad (2.89)$$

Надалі через малість параметру dt знехтуємо величиною $O(dt^2)$, та згідно (2.87)

отримаємо

$$\begin{aligned} \forall \varepsilon, \forall g(t) \quad f[K(t) + \varepsilon g(t)] - f[K(t)] &= \\ &= -2\varepsilon g(t)dt [c(t)\langle e^2(t) \rangle - K(t)R(t)] + \varepsilon^2 g^2(t)R(t)dt \geq 0. \end{aligned} \quad (2.90)$$

Згідно (2.73) та (1.11) отримаємо

$$R_\eta(t, t) = \langle \eta(t) \eta(t) \rangle = \langle \eta^2(t) \rangle = R(t) \delta(0), \quad (2.91)$$

так як $\langle \eta^2(t) \rangle \geq 0$ та $\delta(0) = +\infty > 0$, то функція $R(t)$ є невід'ємною, тож член при ε^2 у (2.90) є невід'ємним. Аналогічно до схожих міркувань з минулого параграфу, член при ε також має бути невід'ємним $\forall \varepsilon$ та $\forall g(t)$, що зумовлює вимогу

$$c(t) \langle e^2(t) \rangle - K(t) R(t) = 0, \quad (2.92)$$

звідки

$$K(t) = \frac{c(t) \langle e^2(t) \rangle}{R(t)}, \quad (2.93)$$

тож диференціальне рівняння для $x^{\text{opt}}(t)$ набуває вигляду

$$\frac{dx^{\text{opt}}(t)}{dt} = \frac{c(t) \langle e^2(t) \rangle}{R(t)} \{z(t) - c(t) x^{\text{opt}}(t)\} + a(t) x^{\text{opt}}(t) + b(t) u(t). \quad (2.94)$$

Нам також потрібно вивести диференціальне рівняння для $\langle e^2(t) \rangle$. На основі (2.87) та (2.93) матимемо

$$\langle e^2(t + \Delta t) \rangle = \langle e^2(t) \rangle + \left[2a(t) \langle e^2(t) \rangle - \frac{c^2(t) \langle e^2(t) \rangle^2}{R(t)} + Q(t) \right] \Delta t + O(\Delta t^2), \quad (2.95)$$

що призводить до

$$\begin{aligned} \frac{d \langle e^2(t) \rangle}{dt} &= \lim_{\Delta t \rightarrow 0} \frac{\langle e^2(t + \Delta t) \rangle - \langle e^2(t) \rangle}{\Delta t} = 2a(t) \langle e^2(t) \rangle - \\ &\quad - \frac{c^2(t) \langle e^2(t) \rangle^2}{R(t)} + Q(t). \end{aligned} \quad (2.96)$$

Тож щойно виведене диференціальне рівняння для $\langle e^2(t) \rangle$:

$$\frac{d \langle e^2(t) \rangle}{dt} = - \frac{c^2(t)}{R(t)} \langle e^2(t) \rangle^2 + 2a(t) \langle e^2(t) \rangle + Q(t). \quad (2.97)$$

Диференціальне рівняння (2.97) є диференціальним рівнянням Ріккаті, яке в деяких випадках є аналітично розв'язним.

Підсумуємо вище сказане. Спершу, шукаємо функцію $\langle e^2(t) \rangle$ як розв'язок диференційного рівняння (2.97), попередньо задавши початкове значення $\langle e^2(0) \rangle$. Надалі, задавши початкове значення $x^{\text{opt}}(0)$ шукаємо функцію $x^{\text{opt}}(t)$ як розв'язок диференційного рівняння (2.94), яке містить щойно обчислену функцію $\langle e^2(t) \rangle$.

2.4. LMS-алгоритми в теорії фільтрації

2.4.1. LMS-алгоритм вилучення корисного сигналу, змішаного з шумом

Розглянемо LMS-алгоритм (least mean squares), що виділяє корисний сигнал з суміші корисного сигналу та шуму. Такий алгоритм (фільтр), на відміну від фільтра Колмогорова–Вінера, використовується тоді, коли кореляційні функції є невідомими.

Нехай маємо сигнал d_t , що є сумою стаціонарного корисного сигналу s_t та шуму n_t :

$$d_t = s_t + n_t. \quad (2.98)$$

Назвемо сигнал d_t «еталонним» сигналом. Для простоти вважатимемо, що корисний сигнал та шум є некорельованими. Нехай x_t – сигнал, який певним чином корельований з шумом і не корельований з корисним сигналом. LMS-фільтр базується на мінімізації середньоквадратичної помилки фільтра. Він є лінійним та стаціонарним, вихід якого обчислюється за формулою

$$y_t = \sum_{\tau=0}^p h_{\tau} x_{t-\tau} \quad (2.99)$$

по $p+1$ попереднім відлікам сигналу x_t , вагові коефіцієнти h_{τ} випадковими не є. Тоді покажемо, що помилка такого фільтра

$$e_t = d_t - y_t \quad (2.100)$$

і є очищеним від шуму корисним сигналом. Дійсно, мінімізація середньоквадратичної помилки згідно (2.98)–(2.100) дає

$$\begin{aligned}
\langle e_t^2 \rangle &= \langle (d_t - y_t)^2 \rangle = \langle (s_t + n_t - y_t)^2 \rangle = \langle s_t^2 \rangle + 2\langle s_t(n_t - y_t) \rangle + \langle (n_t - y_t)^2 \rangle = \\
&= \langle s_t^2 \rangle + 2\langle s_t n_t \rangle - 2\sum_{\tau=0}^p h_\tau \langle s_t x_{t-\tau} \rangle + \langle (n_t - y_t)^2 \rangle \rightarrow \min.
\end{aligned} \tag{2.101}$$

Через некорельованість сигналів x_t та s_t і сигналів s_t та n_t маємо

$$\langle s_t n_t \rangle = 0, \quad \langle s_t x_{t-\tau} \rangle = 0, \tag{2.102}$$

і так як $\langle s_t^2 \rangle$ – середньоквадратичне значення корисного сигналу – є константою, то мінімізація (2.101) еквівалентна до $\langle (n_t - y_t)^2 \rangle \rightarrow \min$, тож вихід фільтру становиться «якомога ближчим» до шуму n_t . Відповідно, через те, що згідно (2.98) та (2.100)

$$e_t - s_t = n_t - y_t, \tag{2.103}$$

та процес y_t «якомога ближчий» до n_t , то випадковий процес e_t становиться «якомога ближчим» до корисного сигналу s_t , що і треба було показати.

Так як взаємна кореляційна функція процесів n_t та x_t невідома, то вимога $\langle e_t^2 \rangle \rightarrow \min$ замінюється на вимогу $e_t^2 \rightarrow \min$, і на кожному кроці набір вагових коефіцієнтів фільтра «адаптується» на основі градієнтного методу.

Згідно (2.100) розпишемо

$$e_t^2 = y_t^2 - 2y_t d_t + d_t^2 \rightarrow \min, \tag{2.104}$$

що з урахуванням (2.99) дає

$$e_t^2 = \sum_{\tau, \tau'=0}^p h_\tau h_{\tau'} x_{t-\tau} x_{t-\tau'} - 2\sum_{\tau=0}^p h_\tau x_{t-\tau} d_t + d_t^2 \rightarrow \min. \tag{2.105}$$

Обчислимо похідну $\partial e_t^2 / \partial h_\alpha$, іншими словами, α -ту компоненту градієнту функції $e_t^2(h_0, h_1, \dots, h_p)$:

$$\begin{aligned}
\frac{\partial e_t^2}{\partial h_\alpha} &= \sum_{\tau, \tau'=0}^p \frac{\partial h_\tau h_{\tau'}}{\partial h_\alpha} x_{t-\tau} x_{t-\tau'} - 2\sum_{\tau=0}^p \frac{\partial h_\tau}{\partial h_\alpha} x_{t-\tau} d_t = \\
&= \sum_{\tau, \tau'=0}^p \left(\frac{\partial h_\tau}{\partial h_\alpha} h_{\tau'} + h_\tau \frac{\partial h_{\tau'}}{\partial h_\alpha} \right) x_{t-\tau} x_{t-\tau'} - 2\sum_{\tau=0}^p \frac{\partial h_\tau}{\partial h_\alpha} x_{t-\tau} d_t =
\end{aligned} \tag{2.106}$$

$$\begin{aligned}
&= \sum_{\tau, \tau'=0}^p (\delta_{\alpha\tau} h_{\tau'} + h_{\tau} \delta_{\alpha\tau'}) x_{t-\tau} x_{t-\tau'} - 2 \sum_{\tau=0}^p \delta_{\alpha\tau} x_{t-\tau} d_t = \\
&= \sum_{\tau'=0}^p h_{\tau'} x_{t-\tau'} x_{t-\alpha} + \sum_{\tau=0}^p h_{\tau} x_{t-\tau} x_{t-\alpha} - 2 x_{t-\alpha} d_t,
\end{aligned}$$

так як перші два доданки однакові, то остаточно отримаємо

$$\frac{\partial e_t^2}{\partial h_{\alpha}} = 2 x_{t-\alpha} \left(\sum_{\tau=0}^p h_{\tau} x_{t-\tau} - d_t \right) = 2 x_{t-\alpha} (y_t - d_t) = -2 e_t x_{t-\alpha}. \quad (2.107)$$

Відповідно, на кожному кроці величини h_{α} змінюються (адаптуються) таким чином, щоб зробивши крок проти напрямку градієнта прагнути мінімізувати величину e_t^2 :

$$h_{\alpha} := h_{\alpha} - \mu \frac{\partial e_t^2}{\partial h_{\alpha}}, \quad (2.108)$$

що згідно (2.107) еквівалентно до

$$h_{\alpha} := h_{\alpha} + 2 \mu e_t x_{t-\alpha}, \quad (2.109)$$

тут і надалі операція $:=$ означає присвоювання.

Тож остаточно LMS–алгоритм (least mean squares), що виділяє корисний сигнал з суміші корисного сигналу та шуму, є таким. Спочатку визначається кількість попередніх точок, за якими надалі обчислюється вихідне значення згідно (2.99) (задається параметр p , відповідна кількість точок дорівнює $p+1$). Також задається крок градієнтного методу μ та послідовність (сигнал) x_t , що є корельованим з шумом та не корельованим з корисним сигналом. Надалі задаються стартові опції алгоритму:

$$t := p; \forall \alpha \in \overline{0, p} \quad h_{\alpha} := \beta; y_t := \sum_{\tau=0}^p h_{\tau} x_{t-\tau}; e_t := d_t - y_t. \quad (2.110)$$

де β – задана константа (часто обирають $\beta = 0$).

І далі на кожному кроці робляться такі обчислення:

$$\forall \alpha \in \overline{0, p} \quad h_{\alpha} := h_{\alpha} + 2 \mu e_t x_{t-\alpha}; t := t + 1; y_t := \sum_{\tau=0}^p h_{\tau} x_{t-\tau}; e_t := d_t - y_t. \quad (2.111)$$

Очищеним від шуму корисним сигналом, який видає LMS–алгоритм, є e_t . Слід зауважити, що алгоритм починає давати адекватні (тобто доволі близькі до істинних) значення корисного сигналу e_t не одразу, тобто не на перших ітераціях, а після проходження певного часу (певної кількості ітерацій).

2.4.2. LMS-алгоритм вилучення інтерференції

Розглянемо випадок, коли шум є синусоїдальною («інтерференційною») завадою. Нехай маємо «еталонний» сигнал згідно виразу (2.98), де s_t є корисним сигналом, а шум є так інтерференційною завадою

$$n_t = A \sin(\omega_0 t + \varphi), \quad (2.112)$$

частота якої відома, а амплітуда та фаза – ні. Вважається, що ця синусоїда некорельована з корисним сигналом s_t . Задачею алгоритму є виокремити корисний сигнал s_t з його суміші з даною синусоїдою.

Алгоритм генерує сигнал у вигляді

$$y_t = a \sin(\omega_0 t) + b \cos(\omega_0 t), \quad (2.113)$$

та адаптує його коефіцієнти a, b так, щоб сигнал y_t став «якомога ближчим» до сигналу

$$n_t = A \cos \varphi \cdot \sin(\omega_0 t) + A \sin \varphi \cdot \cos(\omega_0 t), \quad (2.114)$$

тож з часом a має прямувати до $A \cos \varphi$, а b до $A \sin \varphi$. Виходом фільтру є сигнал e_t , що задається виразом (2.100).

Задача мінімізації величини $e_t^2 \rightarrow \min$ розв'язується згідно градієнтного методу, ідея якого аналогічна до описаної в пункті 2.4.1. Обчислимо похідну $\partial e_t^2 / \partial a$ на основі (2.100) та (2.113):

$$\frac{\partial e_t^2}{\partial a} = \frac{\partial}{\partial a} \left((d_t - a \sin(\omega_0 t) - b \cos(\omega_0 t))^2 \right) = -2e_t \sin(\omega_0 t), \quad (2.115)$$

аналогічно

$$\frac{\partial e_t^2}{\partial b} = -2e_t \cos(\omega_0 t), \quad (2.116)$$

тож на кожному кроці алгоритму коефіцієнти a, b адаптуються згідно виразів

$$a := a + 2\mu e_t \sin(\omega_0 t), \quad b := b + 2\mu e_t \cos(\omega_0 t). \quad (2.117)$$

Позначимо a_t – значення величини a на кроці t . Тоді на основі (2.117), (2.113), (2.98) та (2.114) отримаємо

$$\begin{aligned} a_{t+1} &= a_t + 2\mu e_t \sin(\omega_0 t) = a_t + 2\mu(d_t - a_t \sin(\omega_0 t) - b_t \cos(\omega_0 t)) \sin(\omega_0 t) = \\ &= a_t + 2\mu(s_t + n_t - a_t \sin(\omega_0 t) - b_t \cos(\omega_0 t)) \sin(\omega_0 t) = \\ &= a_t + 2\mu(s_t + (A \cos \varphi - a_t) \cdot \sin(\omega_0 t) + (A \sin \varphi - b_t) \cdot \cos(\omega_0 t)) \sin(\omega_0 t), \end{aligned} \quad (2.118)$$

звідки

$$\begin{aligned} \langle a_{t+1} \rangle &= \langle a_t \rangle + 2\mu \langle s_t \sin(\omega_0 t) \rangle + 2\mu \langle (A \cos \varphi - a_t) \cdot \sin^2(\omega_0 t) \rangle + \\ &+ 2\mu \langle (A \sin \varphi - b_t) \cdot \cos(\omega_0 t) \sin(\omega_0 t) \rangle. \end{aligned} \quad (2.119)$$

Через те, що інтерференційна завада вважається некорельованою з шумом, маємо

$$\langle s_t \sin(\omega_0 t) \rangle = 0. \quad (2.120)$$

Для простоти припустимо, що величини a_t та b_t є незалежними з величинами $\cos(\omega_0 t)$ та $\sin(\omega_0 t)$. Тоді

$$\begin{aligned} \langle (A \cos \varphi - a_t) \sin^2(\omega_0 t) \rangle &= \langle A \cos \varphi - a_t \rangle \langle \sin^2(\omega_0 t) \rangle = \langle A \cos \varphi - a_t \rangle \cdot \frac{1}{2}, \\ \langle (A \sin \varphi - b_t) \cos(\omega_0 t) \sin(\omega_0 t) \rangle &= \langle A \sin \varphi - b_t \rangle \langle \cos(\omega_0 t) \sin(\omega_0 t) \rangle = 0, \end{aligned} \quad (2.121)$$

та на основі (2.119)–(2.121) отримаємо

$$\langle a_{t+1} \rangle = \langle a_t \rangle + \mu \langle A \cos \varphi - a_t \rangle = \mu A \cos \varphi + (1 - \mu) \langle a_t \rangle. \quad (2.122)$$

Тоді на основі (2.122) можна отримати такий ланцюжок рівностей:

$$\begin{aligned} \langle a_{t+1} \rangle &= \mu A \cos \varphi + (1 - \mu) \langle a_t \rangle = \\ &= \mu A \cos \varphi + (1 - \mu) (\mu A \cos \varphi + (1 - \mu) \langle a_{t-1} \rangle) = \\ &= (1 + (1 - \mu)) \mu A \cos \varphi + (1 - \mu)^2 \langle a_{t-1} \rangle = \\ &= (1 + (1 - \mu)) \mu A \cos \varphi + (1 - \mu)^2 (\mu A \cos \varphi + (1 - \mu) \langle a_{t-2} \rangle) = \end{aligned} \quad (2.123)$$

$$\begin{aligned}
&= \left(1 + (1 - \mu) + (1 - \mu)^2\right) \mu A \cos \varphi + (1 - \mu)^3 \langle a_{t-2} \rangle = \dots = \\
&= \left(1 + (1 - \mu) + (1 - \mu)^2 + \dots + (1 - \mu)^t\right) \mu A \cos \varphi + (1 - \mu)^{t+1} \langle a_0 \rangle.
\end{aligned}$$

Згідно відомого виразу для суми геометричної прогресії

$$1 + (1 - \mu) + (1 - \mu)^2 + \dots + (1 - \mu)^t = \frac{1 - (1 - \mu)^{t+1}}{1 - (1 - \mu)} = \frac{1 - (1 - \mu)^{t+1}}{\mu}, \quad (2.124)$$

тож

$$\langle a_{t+1} \rangle = \left(1 - (1 - \mu)^{t+1}\right) A \cos \varphi + (1 - \mu)^{t+1} \langle a_0 \rangle, \quad (2.125)$$

відповідно,

$$\langle a_t \rangle = \left(1 - (1 - \mu)^t\right) A \cos \varphi + (1 - \mu)^t \langle a_0 \rangle, \quad (2.126)$$

що дає

$$\lim_{t \rightarrow \infty} \langle a_t \rangle = \langle a_0 \rangle \lim_{t \rightarrow \infty} (1 - \mu)^t + A \cos \varphi \left(1 - \lim_{t \rightarrow \infty} (1 - \mu)^t\right). \quad (2.127)$$

При $|1 - \mu| < 1 \Leftrightarrow \mu \in (0, 2)$ маємо

$$\lim_{t \rightarrow \infty} \langle a_t \rangle = A \cos \varphi. \quad (2.128)$$

Аналогічно можна довести, що при $\mu \in (0, 2)$

$$\lim_{t \rightarrow \infty} \langle b_t \rangle = A \sin \varphi. \quad (2.129)$$

Тож при $\mu \in (0, 2)$ маємо адаптацію коефіцієнтів, яка призводить до *збіжності в середньому*. Іншими словами, навіть при вдалому підборі кроку алгоритму μ графіки a_t та b_t починаючи з певного часу можуть слабко осилювати навколо значень $A \cos \varphi$ та $A \sin \varphi$, відповідно; проте не в точності співпадати з цими значеннями. Відповідно, починаючи з певного часу значення сигналу e_t , який очищеним від синусоїди (тобто від інтерференційної завади) корисним сигналом, можуть слабко осилювати навколо значень s_t .

2.4.3. LMS-алгоритм лінійного прогнозування часового ряду

Нехай d_t – заданий числовий ряд, значення якого треба спрогнозувати. Прогнозовані значені часового ряду шукатимемо як лінійну комбінацію попередніх даних:

$$x_t = \sum_{j=1}^p h_j d_{t-j}, \quad (2.130)$$

p – кількість попередніх точок, за якими робиться прогнозування наступної (так звана «глибина» прогнозуючого фільтру), вагові коефіцієнти h_j адаптуються на кожному кроці згідно мінімізації квадрату відхилення реальних та прогнозованих даних:

$$e_t = d_t - x_t; e_t^2 \rightarrow \min. \quad (2.131)$$

Аналогічно до виведення результату (2.107) на основі (2.130) та (2.131) отримаємо, що

$$\frac{\partial e_t^2}{\partial h_\alpha} = -2e_t d_{t-\alpha}, \quad (2.132)$$

звідки на основі ідеї градієнтного методу можна отримати, що на кожному кроці вагові коефіцієнти модифікуються на основі їх попередніх значень як

$$h_\alpha := h_\alpha + 2\mu e_t d_{t-\alpha}, \quad (2.133)$$

або

$$h_\alpha := h_\alpha + 2\mu \left(d_t - \sum_{j=1}^p h_j d_{t-j} \right) d_{t-\alpha}, \quad (2.134)$$

де μ – заданий крок градієнтного методу.

На перших кроках такий алгоритм може не давати хорошого прогнозування, проте після проходження певної кількості ітерацій вагові коефіцієнти стають достатньо адаптованими для того, щоб досягнути непоганого співпадіння «реальних» та прогнозованих даних. Відповідно, бажано обрати значення p набагато меншим за задану кількість точок процесу d_t , та спершу покроково робити «прогнозування» для вже заданих точок, щоб перед тим, як алгоритм

почне прогнозувати майбутні точки, його вагові коефіцієнти вже встигли адекватно адаптуватися.

2.5. RLS-алгоритм виділення корисного сигналу, змішаного з шумом

Розглянемо RLS-алгоритм (recursive least squares) для виділення корисного сигналу з його суміші з шумом. Нехай маємо сигнал d_t , що є сумою стаціонарного корисного сигналу s_t та шуму n_t згідно виразу (2.98). Нехай x_t – сигнал, який певним чином корельований з шумом і не корельований з корисним сигналом. Вище показано, що тоді «очищеним від шуму» сигналом є сигнал e_t , що описується виразом (2.100); величина y_t задається виразом (2.99).

В рамках RLS-алгоритму цільова функція (функція, яку треба мінімізувати) задається таким чином: якщо поточний час T , то

$$c = \sum_{t=0}^T \lambda^{T-t} (y_t - d_t)^2 \rightarrow \min, \quad (2.135)$$

де $\lambda \in (0,1]$ – так званий фактор забування, врахування якого свідчить про те, що чим давнішим є часовий відлік, тим меншу вагу він має (при $\lambda = 1$ кожен відлік «важить» однаково). Найчастіше обирають значення фактору забування від 0,98 до 1.

Згідно (2.99) на основі (2.135) отримаємо

$$c = \sum_{t=0}^T \lambda^{T-t} \left(\sum_{\tau, \tau'=0}^p h_{\tau} h_{\tau'} x_{t-\tau} x_{t-\tau'} - 2d_t \sum_{\tau=0}^p h_{\tau} x_{t-\tau} + d_t^2 \right) \rightarrow \min. \quad (2.136)$$

Аналогічно до виведення виразу (2.107) отримаємо, що

$$\frac{\partial c}{\partial h_j} = -2 \sum_{t=0}^T \lambda^{T-t} x_{t-j} e_t. \quad (2.137)$$

На відміну від ідеї градієнтного методу, розвинутій в LMS-алгоритмі, тут прирівнюватимемо частинні похідні (2.137) до нуля, бо шукаємо мінімум функції (2.136). Відповідно, теж буде отримано рекурентну формулу для вагових коефіцієнтів h_j , але на основі іншої ідеї. Згідно (2.137), (2.99) та (2.100) отримаємо

$$\sum_{t=0}^T \lambda^{T-t} x_{t-j} \left(d_t - \sum_{\tau=0}^p h_{\tau} x_{t-\tau} \right) = 0 \Rightarrow r_j(T) = \sum_{\tau=0}^p R_{j\tau}(T) h_{\tau}, \quad (2.138)$$

де

$$r_j(T) = \sum_{t=0}^T \lambda^{T-t} x_{t-j} d_t, \quad R_{j\tau}(T) = \sum_{t=0}^T \lambda^{T-t} x_{t-j} x_{t-\tau}. \quad (2.139)$$

Відповідно, у матричному вигляді

$$\mathbf{r}(T) = \mathbf{R}(T) \mathbf{h}(T), \quad (2.140)$$

де $\mathbf{r}(T)$ – вектор-стовпець, компонентами якого є величини $r_j(T)$, $\mathbf{h}(T)$ – вектор-стовпець, компонентами якого є вагові коефіцієнти h_j та $\mathbf{R}(T)$ – матриця, компонентами якої є величини $R_{j\tau}(T)$. $\mathbf{r}(T)$, $\mathbf{h}(T)$ – стовпці, що містять $p+1$ компоненту, $\mathbf{R}(T)$ – матриця розмірності $(p+1) \times (p+1)$.

Розпишемо вираз для $r_j(T)$:

$$\begin{aligned} r_j(T) &= \sum_{t=0}^T \lambda^{T-t} x_{t-j} d_t = \sum_{t=0}^{T-1} \lambda^{T-t} x_{t-j} d_t + x_{T-j} d_T = \\ &= \lambda \sum_{t=0}^{T-1} \lambda^{T-1-t} x_{t-j} d_t + x_{T-j} d_T = \lambda r_j(T-1) + x_{T-j} d_T, \end{aligned} \quad (2.141)$$

звідки

$$\mathbf{r}(T) = \lambda \mathbf{r}(T-1) + d_T \mathbf{X}(T), \quad (2.142)$$

$\mathbf{X}(T)$ – вектор-стовпець з компонентами $X_j = x_{T-j}$.

Розпишемо вираз для $R_{j\tau}(T)$:

$$\begin{aligned} R_{j\tau}(T) &= \sum_{t=0}^T \lambda^{T-t} x_{t-j} x_{t-\tau} = \lambda \sum_{t=0}^{T-1} \lambda^{T-1-t} x_{t-j} x_{t-\tau} + x_{T-j} x_{T-\tau} = \\ &= \lambda R_{j\tau}(T-1) + x_{T-j} x_{T-\tau}, \end{aligned} \quad (2.143)$$

звідки

$$\mathbf{R}(T) = \lambda \mathbf{R}(T-1) + \mathbf{X}(T) \mathbf{X}^T(T), \quad (2.144)$$

такий запис є коректним, бо $\mathbf{X}(T)$ – стовпчик $(p+1) \times 1$, $\mathbf{X}^T(T)$ – строка $1 \times (p+1)$, їх добуток – саме матриця $(p+1) \times (p+1)$;

$$\left(\mathbf{X}\mathbf{X}^T\right)_{ij} = \mathbf{X}_{ik}\mathbf{X}_{kj}^T = \mathbf{X}_{i1}\mathbf{X}_{1j}^T = X_iX_j = x_{T-i}x_{T-j}. \quad (2.145)$$

Увага – не плутати момент часу T та позначення транспонування матриці!

Згідно (2.140) маємо

$$\mathbf{h}(T) = \mathbf{R}^{-1}(T)\mathbf{r}(T). \quad (2.146)$$

Як відомо, матрична тотожність Вудбері має такий вигляд:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \quad (2.147)$$

На основі (2.144) маємо

$$\mathbf{R}^{-1}(T) = \left(\lambda\mathbf{R}(T-1) + \mathbf{X}(T)\mathbf{I}\mathbf{X}^T(T)\right)^{-1}, \quad (2.148)$$

$\mathbf{I}$ – одинична матриця розміром 1 на 1. Тоді згідно (2.147) та (2.148) отримаємо

$$\begin{aligned} \mathbf{R}^{-1}(T) &= \left\{\mathbf{I}^{-1} = \mathbf{I} = 1 - \text{число}\right\} = \lambda^{-1}\mathbf{R}^{-1}(T-1) - \\ &- \lambda^{-1}\mathbf{R}^{-1}(T-1)\mathbf{X}(T)\underbrace{\left(1 + \mathbf{X}^T(T)\lambda^{-1}\mathbf{R}^{-1}(T-1)\mathbf{X}(T)\right)^{-1}}_{\text{число}}\mathbf{X}^T(T)\lambda^{-1}\mathbf{R}^{-1}(T-1) \end{aligned} \quad (2.149)$$

Введемо вектор-стовпець

$$\mathbf{g}(T) = \lambda^{-1}\mathbf{R}^{-1}(T-1)\underbrace{\mathbf{X}(T)}_{\text{стовпець}}\underbrace{\left(1 + \mathbf{X}^T(T)\lambda^{-1}\mathbf{R}^{-1}(T-1)\mathbf{X}(T)\right)^{-1}}_{\text{число}}. \quad (2.150)$$

Тоді

$$\mathbf{R}^{-1}(T) = \lambda^{-1}\mathbf{R}^{-1}(T-1) - \lambda^{-1}\mathbf{g}(T)\mathbf{X}^T(T)\mathbf{R}^{-1}(T-1). \quad (2.151)$$

Перетворимо стовпець $\mathbf{g}(T)$:

$$\begin{aligned} \mathbf{g}(T) &= \lambda^{-1}\mathbf{R}^{-1}(T-1)\mathbf{X}(T) \cdot \frac{1}{1 + \mathbf{X}^T(T)\lambda^{-1}\mathbf{R}^{-1}(T-1)\mathbf{X}(T)} = \\ &= \lambda^{-1}\mathbf{R}^{-1}(T-1)\mathbf{X}(T) \cdot \frac{1}{\lambda^{-1}(\lambda + \mathbf{X}^T(T)\mathbf{R}^{-1}(T-1)\mathbf{X}(T))} = \\ &= \mathbf{R}^{-1}(T-1)\mathbf{X}(T) \cdot \frac{1}{\lambda + \mathbf{X}^T(T)\mathbf{R}^{-1}(T-1)\mathbf{X}(T)}, \end{aligned} \quad (2.152)$$

так можна робити, бо $\lambda + \mathbf{X}^T(T)\mathbf{R}^{-1}(T-1)\mathbf{X}(T)$ – число, і λ – число. Тож

$$\mathbf{g}(T) \cdot (\lambda + \mathbf{X}^T(T)\mathbf{R}^{-1}(T-1)\mathbf{X}(T)) = \mathbf{R}^{-1}(T-1)\mathbf{X}(T), \quad (2.153)$$

поділимо це вираз на λ :

$$\mathbf{g}(T) \cdot (1 + \lambda^{-1} \mathbf{X}^T(T) \mathbf{R}^{-1}(T-1) \mathbf{X}(T)) = \lambda^{-1} \mathbf{R}^{-1}(T-1) \mathbf{X}(T), \quad (2.154)$$

розкриємо дужки та перезберемо члени:

$$\mathbf{g}(T) = (\lambda^{-1} \mathbf{R}^{-1}(T-1) - \lambda^{-1} \mathbf{g}(T) \mathbf{X}^T(T) \mathbf{R}^{-1}(T-1)) \mathbf{X}(T), \quad (2.155)$$

що з урахуванням (2.151) дає

$$\mathbf{g}(T) = \mathbf{R}^{-1}(T) \mathbf{X}(T). \quad (2.156)$$

На основі (2.146) та (2.142) отримаємо

$$\begin{aligned} \mathbf{h}(T) &= \mathbf{R}^{-1}(T) (\lambda \mathbf{r}(T-1) + d_T \mathbf{X}(T)) = \\ &= \lambda \mathbf{R}^{-1}(T) \mathbf{r}(T-1) + d_T \mathbf{R}^{-1}(T) \mathbf{X}(T), \end{aligned} \quad (2.157)$$

з урахуванням (2.156) матимемо

$$\mathbf{h}(T) = \lambda \mathbf{R}^{-1}(T) \mathbf{r}(T-1) + d_T \mathbf{g}(T). \quad (2.158)$$

Підставимо (2.151) у (2.158):

$$\begin{aligned} \mathbf{h}(T) &= \lambda (\lambda^{-1} \mathbf{R}^{-1}(T-1) - \lambda^{-1} \mathbf{g}(T) \mathbf{X}^T(T) \mathbf{R}^{-1}(T-1)) \mathbf{r}(T-1) + d_T \mathbf{g}(T) = \\ &= \underbrace{\mathbf{R}^{-1}(T-1) \mathbf{r}(T-1)}_{=\mathbf{h}(T-1)} - \underbrace{\mathbf{g}(T) \mathbf{X}^T(T) \mathbf{R}^{-1}(T-1) \mathbf{r}(T-1)}_{=\mathbf{h}(T-1)} + \underbrace{d_T \mathbf{g}(T)}_{\text{число}} = \\ &= \mathbf{h}(T-1) + \mathbf{g}(T) (d_T - \mathbf{X}^T(T) \mathbf{h}(T-1)). \end{aligned} \quad (2.159)$$

Позначимо

$$\alpha(T) = d_T - \mathbf{X}^T(T) \mathbf{h}(T-1), \quad \mathbf{P}(T) = \mathbf{R}^{-1}(T). \quad (2.160)$$

зауважимо, що $\alpha(T)$ – число (не матриця).

Тоді, резюмуючи вище сказане, запишемо рекурентні співвідношення, на яких базується робота RLS-алгоритму:

$$\begin{aligned} \alpha(T) &= d_T - \mathbf{X}^T(T) \mathbf{h}(T-1), \quad \mathbf{g}(T) = \frac{\mathbf{P}(T-1) \mathbf{X}(T)}{\lambda + \mathbf{X}^T(T) \mathbf{P}(T-1) \mathbf{X}(T)}, \\ \mathbf{P}(T) &= \lambda^{-1} \mathbf{P}(T-1) - \lambda^{-1} \mathbf{g}(T) \mathbf{X}^T(T) \mathbf{P}(T-1), \\ \mathbf{h}(T) &= \mathbf{h}(T-1) + \mathbf{g}(T) \alpha(T), \end{aligned} \quad (2.161)$$

$$e_T = d_T - y_T = d_T - \sum_{\tau=0}^p h_{\tau}(T) x_{T-\tau} = d_T - \mathbf{h}^T(T) \mathbf{X}(T),$$

$$\mathbf{h}(T) = (h_0(T) \quad h_1(T) \quad \dots \quad h_p(T))^T,$$

$$\mathbf{X}(T) = (x_T \quad x_{T-1} \quad \dots \quad x_{T-p})^T.$$

e_T є «очищеним від шуму» значенням корисного сигналу в момент часу T . Стартові значення вагових коефіцієнтів зазвичай обирають нулями: $\mathbf{h}(0)$ – стовпчик з нулів. Стартове значення матриці $\mathbf{P}$ зазвичай обирають у вигляді

$$\mathbf{P}(0) = \beta^{-1} \mathbf{I}, \quad (2.162)$$

де $\mathbf{I}$ – одинична матриця розмірності $(p+1) \times (p+1)$ та $\beta > 0$ – мала константа при великому відношенні сигнал-шум та велика константа при малому відношенні сигнал-шум.

2.6. Вимоги до адаптивних фільтрів

В загальному випадку одноканальний адаптивний фільтр є пристроєм, схема роботи якого зображена на рис. 2.1.

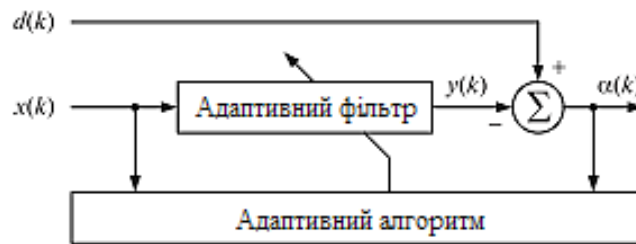


Рисунок 2.1. Схема адаптивного фільтра

Тут $x(k)$ – вхідний сигнал, $d(k)$ – сигнал, що вимагається (позначення « d » від англ. «desired»), $y(k)$ – вихідний сигнал, $\alpha(k) = d(k) - y(k)$ – сигнал помилки, k – індекс дискретного часу або номер відліку сигналу, що опрацьовується. Ці відліки зазвичай рівномірно розподілені в часі: $t = kT_S = k/F_S$, де T_S – період дискретизації, а F_S – частота дискретизації сигналів, що оброблюються. Тривалість однієї ітерації адаптивного алгоритму, тобто час, впродовж якого

обчислюються вагові коефіцієнти, необхідні для обчислення вихідного сигналу фільтру на наступній ітерації, в більшості випадків дорівнює одному періоду дискретизації. Якщо розрахунок вагових коефіцієнтів відбувається повільно, тобто впродовж декількох періодів дискретизації, то формування сигналу $y(k)$ все одно треба робити на кожному періоді дискретизації для того, щоб задовольнити вимоги теореми Котельникова.

Для роботи адаптивного фільтру, як правило, недостатньо інформації, що отримується лише з вхідного сигналу. Джерелом такої додаткової необхідної інформації часто є сигнал, що вимагається $d(k)$, фізична природа якого задається конкретним застосуванням адаптивного фільтру. Застосування є одним з факторів, що задають вимоги до адаптивного фільтру на етапі його проектування.

Основні вимоги до адаптивного фільтру задаються такими складовими, як застосування, структура та алгоритм.

2.6.1. Застосування

Тип застосування задається сигналами, що надходять з навколишнього середовища в якості вхідного сигналу $x(k)$ та сигналу, що вимагається $d(k)$. Найбільш відомими та широко застосовними застосуваннями адаптивних фільтрів є фільтри-апроксиматори в задачах спостереження та ідентифікації, ехо-компенатори, еквалайзери, адаптивні ґратки та низка інших пристроїв. Ця книга не ставить метою розглянути всі можливі застосування адаптивних фільтрів, проте низка таких застосувань в книзі розглянута.

2.6.2. Структури адаптивних фільтрів

Існують два основні структури адаптивних фільтрів – з кінцевою імпульсною характеристикою (KIX) та нескінченною імпульсною характеристикою (НІХ).

Структура одноканального KIX-фільтру наведена на рис. 2.2. Вихідний сигнал адаптивного фільтру $y(k)$ формується як лінійна комбінація затриманих

відліків відного сигналу $x(k - n + 1)$, взятих з ваговими коефіцієнтами $h_n(k - 1)$, що обчислюються на минулих $k - 1$ ітерації по відношенню до поточної ітерації k :

$$y(k) = \sum_{n=1}^N h_n^*(k-1)x(k-n+1) = \mathbf{h}_N^H(k-1)\mathbf{x}_N(k), \quad (2.163)$$

де N – число вагових коефіцієнтів фільтра, $\mathbf{h}_N(k-1)=[h_1(k-1), h_2(k-1), \dots, h_n(k-1), \dots, h_{N-1}(k-1), h_N(k-1)]^T$ – вектор вагових коефіцієнтів та $\mathbf{x}_N(k)=(x(k), x(k-1), \dots, x(k-n+1), \dots, x(k-N+2), x(k-N+1)]^T$ – вектор сигналів у фільтрі. Порядок фільтра визначається як $N - 1$. Вектори $\mathbf{h}_N(k)$ та $\mathbf{x}_N(k)$ в загальному випадку можуть містити комплексні величини.

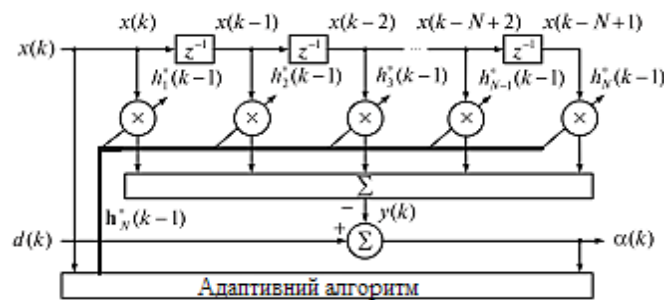


Рисунок 2.2. Одноканальний адаптивний КІХ-фільтр

Тут і надалі верхній індекс Т означає транспонування, а верхній індекс Н – ермітове спряження, тобто транспонування, а потім комплексне спряження. Зірочкою * позначається комплексне спряження відповідної величини. Перший нижній індекс N означає або число елементів в заданому векторі, або число рядків або стовпців у відповідній квадратній матриці розмірності $N \times N$.

На основі (2.163) можна обчислити сигнал апіорної помилки моделювання сигналу $d(k)$, що вимагається, таким чином:

$$\alpha(k) = d(k) - y(k) = d(k) - \mathbf{h}_N^H(k-1)\mathbf{x}_N(k), \quad (2.164)$$

Сигнал апостеріорної помилки визначається при вагових коефіцієнтах, рівних $\mathbf{h}_N(k)$, таким чином:

$$e(k) = d(k) - \hat{y}(k) = d(k) - \mathbf{h}_N^H(k-1)\mathbf{x}_N(k). \quad (2.165)$$

Як бачимо, терміни «апостеріорний» та «апостеріорний» пов'язані зі значеннями вагових коефіцієнтів адаптивного фільтра, що обчислюються, відповідно, на минулій та поточній ітераціях алгоритму адаптивної фільтрації.

На практиці при роботі адаптивного фільтра на його виході спостерігаються апостеріорні помилки (2.164), бо поточне значення вихідного сигналу фільтра $y(k)$ формується по значенням вагових коефіцієнтів, обчислених на минулій ітерації. Апостеріорні помилки (2.165) зазвичай формуються в алгоритмах обчислення вагових коефіцієнтів та при формуванні цільової функції роботи адаптивного фільтра, наприклад середньоквадратичної помилки (Mean Square Error, MSE) $F = E\{e(k)e^*(k)\} = E\{|e(k)|^2\}$, де $E\{\bullet\}$ – операція усереднення по ансамблю реалізацій. У випадку КІХ-фільтрів така функція є унімодальною квадратичною функцією в просторі дійсних або комплексних вагових коефіцієнтів, тобто характеризується єдиним мінімумом.

В загальному випадку адаптивні КІХ-фільтри можуть бути багатоканальними, та при цьому містити неоднакове число вагових коефіцієнтів в каналах (див. рис. 2.3). Вектор вагових коефіцієнтів такого M -канального адаптивного фільтру

$$\mathbf{h}_N(k-1) = [\mathbf{h}_{N_1}^T(k-1), \mathbf{h}_{N_2}^T(k-1), \dots, \mathbf{h}_{N_m}^T(k-1), \dots, \mathbf{h}_{N_{M-1}}^T(k-1)]^T$$

утворюється з послідовності векторів вагових коефіцієнтів каналів

$$\mathbf{h}_{N_m}(k) = [h_{1m}(k-1), h_{2m}(k-1), \dots, h_{nm}(k-1), \dots, h_{N_{m,m}}(k-1)]^T,$$

а вектор сигналів $\mathbf{x}_N(k) = [\mathbf{x}_{N_1}^T(k), \mathbf{x}_{N_2}^T(k), \dots, \mathbf{x}_{N_m}^T(k), \dots, \mathbf{x}_{N_{M-1}}^T(k), \mathbf{x}_{NM}(k)]^T$ – з послідовності векторів сигналів каналів

$$\mathbf{x}_{N_m}(k) = [x_m(k), x_m(k-1), \dots, x_m(k-n_m+1), \dots, x_m(k-N_m+2), \dots, x_m(k-N_m+1)]^T.$$

Сумарне число вагових коефіцієнтів багатоканального фільтру задається як $\sum_{m=1}^M N_m$.

Неоднакове число вагових коефіцієнтів в каналах багатоканального адаптивного фільтра, з одного боку, часто диктується фізичною природою задачі, а з іншого боку обумовлено існуючими обмеженнями на обчислювальну

складність реалізації фільтрів. Так як в адаптивних фільтрах складність обчислення є функцією від повного числа вагових коефіцієнтів N , то це число не слід збільшувати, якщо на це немає суттєвих причин, не дивлячись навіть на те, що алгоритми, наприклад для багатоканальних адаптивних фільтрів з однаковим числом вагових коефіцієнтів в каналах, математично є набагато простішими за алгоритми для фільтрів з неоднаковим числом вагових коефіцієнтів в каналах.

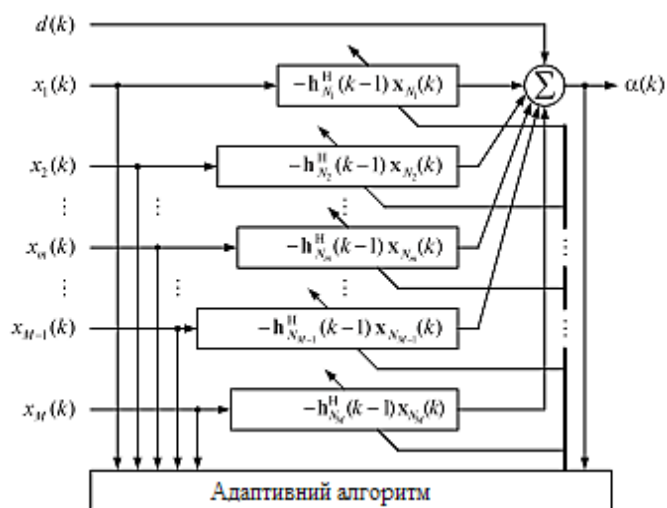


Рисунок 2.3. Багатоканальний адаптивний КІХ-фільтр

Під обчислювальною складністю алгоритмів цифрової обробки сигналів, до яких належать також і алгоритми адаптивної фільтрації сигналів, розуміється число арифметичних операцій, що вимагаються для виконання однієї ітерації алгоритму.

Слід відзначити два частинні випадки структури адаптивного фільтра. Це багатоканальний фільтр з одним ваговим коефіцієнтом ($N_m = 1$, $m = 1, 2, \dots, M$) в кожному з каналів, та багатоканальний фільтр з однаковим числом вагових коефіцієнтів в каналах. Структура багатоканального фільтра загального вигляду (див. рис. 2.3) використовується, зокрема, при реалізації нелінійних поліноміальних адаптивних фільтрів, в яких нелінійні ядра є наборами багатоканальних фільтрів з числом вагових коефіцієнтів в каналах, що змінюється від $N_1 = 1$ до $N_M = N$ з кроком, рівним одному ваговому коефіцієнту.

Зазначимо, що компенсатор сигналів ближнього та дальнього еха в модемах для провідних каналів зв'язку або еквалайзер зі зворотним зв'язком також можна розглядати як двоканальний адаптивний фільтр з різним числом вагових коефіцієнтів в каналах.

На сьогодні адаптивні НІХ-фільтри ще не знайшли широкого застосування на практиці, бо окрім проблем зі стійкістю, їм притаманна також проблема багатоекстремальності (тобто неунімодалності) цільової функції, що в загальному випадку не дозволяє гарантувати збіжність процесу обчислення вагових коефіцієнтів до глобального (тобто найкращого) розв'язку.

2.6.3. Алгоритми

Адаптивний алгоритм – процедура обчислення вагових коефіцієнтів, що забезпечують мінімізацію цільової функції адаптивного фільтра, тобто виконання критерію його роботи. Алгоритм характеризується виглядом використаної цільової функції, методом пошуку оптимального рішення та природою сигналів помилок. В адаптивних алгоритмах, як правило, використовуються арифметичні операції як над апіорними, так і над апостеріорними помилками. Процедури обчислення вихідних сигналів адаптивного фільтра $y(k) = \mathbf{h}_M^H(k-1)\mathbf{x}_M(k)$ та $\check{y}(k) = \mathbf{h}_M^H(k)\mathbf{x}_N(k)$ в рівняннях (2.164) та (2.165), тобто безпосередньо фільтрація сигналів, також часто вважаються складовими частинами адаптивного алгоритму.

Більшість обчислень в адаптивних фільтрах являють собою ітераційні процедури. Ці ітерації, як правило, співпадають за тривалістю з періодом дискретизації сигналів, що оброблюються. При цьому обчислювальні змінні оновлюються на кожній ітерації в результаті додавання певних величин до минулих значень або віднімання певних величин з минулих значень.

Наприклад, обчислення вагових коефіцієнтів в адаптивному фільтрі по критерію найменшого середнього квадрату виконується таким чином:

$$\mathbf{h}_N(k) = \mathbf{h}_N(k-1) + \mu \mathbf{x}_N(k) \alpha^*(k), \quad (2.166)$$

тобто поточне значення вектору вагових коефіцієнтів дорівнює минулому значенню вектору вагових коефіцієнтів плюс крок збіжності, помножений на вектор вхідних сигналів та на помилку.

В рекурсивних алгоритмах по критерію найменших квадратів (Recursive Least Squares, RLS) обчислення вагових коефіцієнтів проводиться таким чином:

$$\mathbf{h}_N(k) = \mathbf{h}_N(k-1) + \mathbf{g}_N(k) \alpha^*(k). \quad (2.167)$$

В низці інших алгоритмів, наприклад, в алгоритмах на основі QR-розвинення та в драбинних алгоритмах, також використовуються рекурсивні обчислення, які пов'язані зі зміною порядку фільтра на одиницю на кожній ітерації по k .

Обрання адаптивного алгоритму зазвичай зумовлене низкою таких вимог, як тривалість перехідного процесу; характеризуюча швидкість збіжності; значення залишкових похибок в усталеному режимі, що характеризує точність знаходження оптимального розв'язку; обчислювальна складність алгоритму; складність апаратної або програмної реалізації алгоритму, та ін.

До цих вимог слід додати поняття алгоритмічної складності, під якою розуміється складність математичного представлення алгоритмів, що зумовлена числом та різноманіттям математичних виразів (формул), що і є, власне, алгоритмом. Алгоритмічна та арифметична складності врешті решт впливають на складність реалізації алгоритмів.

З точки зору складності алгоритми адаптивної обробки сигналів можна умовно розділити на декілька груп. Існують обчислювально прості алгоритми, тобто з малим числом арифметичних операцій на одну ітерацію. Це – алгоритми з лінійною обчислювальною складністю $O(N)$ та малою алгоритмічною складністю. Такі алгоритми зазвичай описуються лише декількома простими математичними виразами. Прикладами простих алгоритмів є LMS-алгоритм та інші адаптивні алгоритми на основі стратегій градієнтного пошуку вагових коефіцієнтів. Алгоритм по критерію найменших квадратів (Least Squares, LS)

характеризується більшою обчислювальною складністю $O(N^3)$ та малою алгоритмічною складністю. RLS-алгоритми характеризуються середньою обчислювальною складністю $O(N^2)$ та середньою алгоритмічною складністю. Швидкі, тобто обчислювально-ефективні або з малим числом арифметичних операцій на одну ітерацію, RLS-алгоритми характеризуються низькою обчислювальною складністю $O(N)$ та більшою алгоритмічною складністю, бо задаються доволі великим числом різноманітних математичних формул.

Слід зауважити, що сучасна елементна база дозволяє реалізовувати як прості, так і доволі складні алгоритми адаптивної фільтрації.

2.7. Критерії функціонування адаптивних алгоритмів

В процесі роботи адаптивного алгоритму шукаються такі вагові коефіцієнти фільтру $\mathbf{h}_N(k)$, при яких значення заданої цільової функції є мінімальним. Зазвичай це функція від вхідного, вихідного сигналів, а також від сигналу, що вимагається $F=f\{x(k), y(k), d(k)\}$. Цільову функцію можна розглядати і як функцію помилок, тобто $F=f\{e(k)\}$. Цільова функція має бути дійсною та невід'ємною. При роботі адаптивного фільтра $\mathbf{h}_N(k) \rightarrow \mathbf{h}_{N,0}$, де $\mathbf{h}_{N,0}$ – вектор координат мінімум багатовимірної цільової функції в просторі вагових коефіцієнтів. Досягнення мінімуму цієї функції є критерієм роботи адаптивного фільтра.

Існують різні способи визначення цільової функції адаптивного фільтра, що врешті решт впливають на складність алгоритмів її мінімізації. Цільовими функціями, які найчастіше використовуються в адаптивній фільтрації, є

- 1) $F = f\{e(k)\} = E\{|e(k)|^2\}$ – середньоквадратична помилка (MSE);
- 2) $F = f\{e(k)\} = \sum_{i=1}^k |e(i)|^2$ – найменші квадрати помилок (Least Squares, LS);
- 3) $F = f\{e(k)\} = \sum_{i=1}^k \lambda^{k-i} |e(i)|^2$ – зважені найменші квадрати помилок (Weighted Least Squares, WLS);
- 4) $F = f\{e(k)\} = |e(k)|^2$ – квадрат миттєвої помилки.

На практиці використовують і інші критерії, наприклад критерій постійності модулі інформаційних символів (Constant Modulus, CM), що часто застосовується при адаптивній обробці сигналів в системах цифрового зв'язку. В основі цього критерію лежить мінімізація функції помилок між значенням модуля (комплексної обгортки) вихідного сигналу адаптивного фільтру та відомим значенням модуля (комплексної обгортки) інформаційних символів, возведених в певний ступінь. Ця мінімізація забезпечує прямування комплексної обгортки вихідного сигналу адаптивного фільтру до відомого постійного значення комплексної обгортки інформаційних символів. Тому цей критерій отримав свою назву не від назви цільової функції, а від назви остаточного результату, який вимагається, тобто значення обгортки вихідного сигналу адаптивного фільтру, що дозволяє коректно розпізнавати в цьому сигналі інформаційні символи, що приймаються.

Вибір квадратичних цільових функцій в адаптивних фільтрах багато в чому зумовлений тим, що в результаті їх використання можна отримувати алгоритми адаптивної фільтрації у вигляді рекурентних обчислювальних процедур, які не містять логічних операцій, що відрізняє такі алгоритми від обчислювальних процедур оптимізації в загальному вигляді. Крім цього, ці функції є унімодальними, що гарантує збіжність алгоритмів адаптивної фільтрації на їх основі до єдиного оптимального розв'язку.

Перша з вище перелічених, MSE-функція, є лише зручним математичним поняттям, бо для її обчислення вимагається наявність нескінченного об'єму даних, бо отримання цієї функції полягає в усередненні по ансамблю реалізацій сигналів, що спостерігаються. На практиці вона використовується доволі рідко. Вона використовується, зокрема, в оптимальній Вінерівській фільтрації, що являє в основному теоретичний інтерес.

Цільові функції з другої по четверту характеризуються як складністю реалізації алгоритмів на їх основі, так і характеристиками збіжності та залишковими помилками цих алгоритмів в усталеному режимі. Зокрема, квадрат миттєвої помилки є найпростішою з точки зору реалізації алгоритмів цільовою

функцією. Однак, алгоритми на її основі характеризуються повільною збіжністю, бо така цільова функція є надто спрощеною. Цільова LS-функція зазвичай використовується при обробці стаціонарних сигналів, в WLS – при обробці сигналів, що повільно змінюються.

2.8. Контрольні питання до глави 2

1. Сформулюйте математичну постановку задачі скалярного дискретного фільтра Калмана та основні припущення відповідної теорії.
2. Виведіть рекурентні формули, що описують покрокову роботу скалярного дискретного фільтра Калмана. Вкажіть початкові умови алгоритму.
3. Проведіть аналогію між скалярним та векторним дискретними фільтрами Калмана. Сформулюйте математичну постановку задачі векторного дискретного фільтра Калмана та основні припущення відповідної теорії.
4. Виведіть рекурентні формули, що описують покрокову роботу векторного дискретного фільтра Калмана. Вкажіть початкові умови алгоритму. Вкажіть основну ідею мінімізації функції від матриці, що застосована при даному виведенні.
5. Сформулюйте математичну постановку задачі скалярного фільтра Калмана–Бюсі та основні припущення відповідної теорії.
6. Виведіть формули, що описують роботу скалярного фільтра Калмана–Бюсі. Вкажіть основну ідею мінімізації функціонала, що застосована при даному виведенні. Вкажіть тип виведеного диференційного рівняння для середньоквадратичної помилки фільтру.
7. Сформулюйте математичну постановку задачі LMS–алгоритму, що вилучає корисний сигнал з його суміші з шумом.
8. Виведіть формули, що описують покрокову роботу LMS–алгоритму, що вилучає корисний сигнал з його суміші з шумом.
9. Сформулюйте математичну постановку задачі LMS–алгоритму для вилучення інтерференції.

10. Виведіть формули, що описують покрокову роботу LMS–алгоритму для вилучення інтерференції.
11. Виведіть умови збіжності в середньому LMS–алгоритму для вилучення інтерференції.
12. Сформулюйте математичну постановку задачі та виведіть формулу покрокової роботи LMS–алгоритму для лінійного прогнозування часового ряду.
13. Сформулюйте математичну постановку задачі RLS–алгоритму, що вилучає корисний сигнал з його суміші з шумом.
14. Виведіть формули, що описують покрокову роботу RLS –алгоритму, що вилучає корисний сигнал з його суміші з шумом.
15. Сформулюйте та коротко опишіть основні вимоги до адаптивного фільтру.
16. Опишіть типи цільових функцій, які застосовуються в теорії адаптивної фільтрації.

Глава 3. Адаптивні фільтри-апроксиматори в розв'язуванні задач спостереження та ідентифікації

З точки зору теорії керування проблема оперативності спостереження в рамках системи керування може бути вирішена шляхом прогнозування (на необхідний інтервал випередження) значень змінних об'єкта керування (ОК) за їх вимірюваним у минулому значеннями, що здійснюється за допомогою прогнозуючих фільтрів.

У класичній постановці Вінера-Колмогорова здійснюється оптимальне (у сенсі мінімуму похибки) виділення корисного сигналу $s(t)$ із вхідного сигналу $y(t)$ на тлі шуму $n(t)$, тобто вхідний сигнал має вид:

$$y(t) = s(t) + n(t).$$

Метою прогнозування є одержання оцінки майбутнього значення сигналу

$$\hat{y}(t) = s(t + \Delta t),$$

де Δt – інтервал випередження.

Фільтрами називаються пристрої (алгоритми), що реалізують наведені вище вирази при $\Delta t=0$. При $\Delta t>0$ і $n(t)\equiv 0$ пристрої називаються випереджувачами, а при $\Delta t>0$ і $n(t)\neq 0$ – прогнозуючими фільтрами.

Фільтр Вінера-Колмогорова є оптимальним для оцінки стаціонарного сигналу. Калманом і Б'юсі у 1960 р. був застосований метод диференціальних рівнянь, для якого умова стаціонарності сигналу і завади не є обов'язковою. Після цього Д. Габор у 1962 р. запропонував алгоритми по створенню адаптивного самоналагоджувального фільтра. Якщо процес ергодичний, то в ході навчання прогнозуючий фільтр відтворює всі статистичні характеристики процесу.

Рішення задачі прогнозування полягає в інтерполяції часового ряду (за допомогою апроксимуючих функцій) і екстраполяції значень ряду на майбутнє за його попередніми значеннями функцій (наприклад, для стаціонарного процесу екстраполяція повинна враховувати сталість статистичних моментів часового

ряду, а для нестационарних – розвиток його тренда в часі) з метою забезпечення обраного критерію якості.

Розвиток прогностики як науки привело до створення багатьох методів і прийомів прогнозування. Зараз налічується понад сто методів прогнозування, у зв'язку з чим виникає завдання вибору методів, які давали б адекватні прогнози для процесів і систем.

Необхідно вказати на три принципові проблеми, що виникають при прогнозуванні. Перша проблема – це визначення необхідних і достатніх параметрів для оцінки стану ОК.

Друга проблема полягає в так званому «проклятті розмірності». Бажання врахувати в моделі якомога більше показників і критеріїв оцінки може привести до її надмірної громіздкості.

Третя проблема – наявність феномена «надсистемності». Взаємодіючі системи утворюють систему більш високого рівня, яка має властивості, що роблять принципово недосяжною можливість реалізації надсистемного відображення і цільових функцій з погляду систем, що входять до складу надсистеми.

Для подолання перерахованих проблем робляться спроби застосування таких розділів сучасної математики, як нейрокомп'ютери, теорія хаосу, теорія ризиків, теорія катастроф і теорія самоорганізуючих систем. Ці методи дозволяють збільшити глибину прогнозу за рахунок виявлення прихованих закономірностей і взаємозв'язків там, де погано формалізуються моделі звичайними методами.

3.1. Дискретизація та відновлення сигналів у розв'язуванні задач спостереження та ідентифікації

В задачах спостереження й ідентифікації метою дискретизації є забезпечення мінімальної похибки апроксимації неперервних сигналів дискретними формами.

Дискретизація і відновлення сигналів визначаються теоремами Котельникова-Шеннона, які стверджують, що, якщо спектр безперервної функції $x(t)$ обмежений круговою частотою $\omega_{\max}$ (частотою Найквіста $F_{\max}$), то:

- частота дискретизації функції $x(t)$ повинна бути $F_{\partial} \geq \omega_{\max} / \pi = 2F_{\max}$,

що дозволяє уникнути перекручування спектра функції, яку дискретизували, за рахунок його повторення (*теорема дискретизації Шеннона*);

- безперервна функція $x(t)$ однозначно відновлюється (визначається) за її дискретними відліками $x(kT)$, узятими рівномірно з періодом $T = F_{\partial}^{-1} \leq \pi / \omega_{\max} = (2F_{\max})^{-1}$:

$$x(t) = \sum_k x(kT) \cdot \text{sinc}(t/T - k), \quad (3.1)$$

де $\text{sinc}(t) = \sin(\pi t) / (\pi t)$; k – номери відліків (*теорема відновлення (відліків) Котельникова-Шеннона*).

Область визначення цих теорем, крім сигналів з обмеженим спектром (СОС), розповсюджується також на смугові і багатосмугові сигнали.

Коли в сигналі є частоти більші $F_{\max}$ (сигнали з необмеженим спектром – СНС), то для одержання сигналів СОС необхідно використовувати формуючий передфільтр нижніх частот (ФНЧ) для запобігання ефекту накладання, а для відновлення безперервних сигналів за їх відліками використовується формуючий постфільтр ФНЧ (рис. 3.1).

Дискретизовані відліки (при $T = 1$) визначаються як:

$$c_{\delta}(k) = c(k) \cdot \delta(t - k), \quad (3.2)$$

а відновлений сигнал дорівнює:

$$\hat{x}(t) = \sum_k c_{\delta}(k) \cdot \varphi(t - k), \quad (3.3)$$

де, відповідно до (11.12): $c(k) = x(kT)$; $h(t) = \varphi(t) = \text{sinc}(t)$ – ідеальні ФНЧ; δ – функція Дірака.

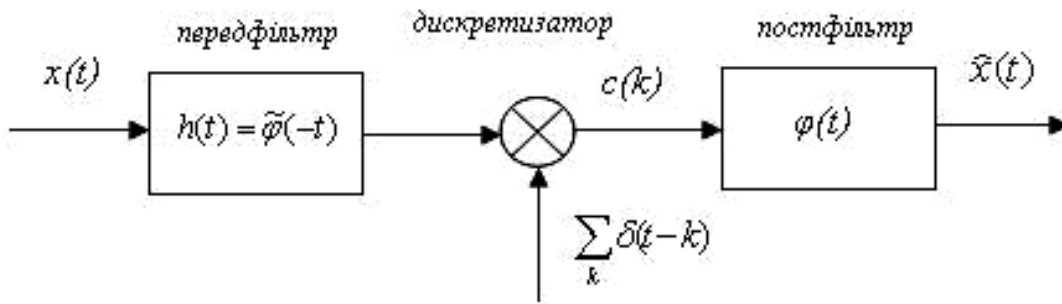


Рисунок 3.1. Схема дискретизації безперервної функції (сигналу)

Використання моделі апроксимації сигналу (3.1) - (3.3) не завжди коректне через неможливість реалізації ідеальних ФНЧ, а область використання *sinc*-функцій обмежена через їх низьку роздільну здатність.

У сучасних підходах фільтри ФНЧ для сигналів СНС можуть обиратися не ідеальними, але вони повинні бути біортогональними:

$$\langle \varphi(t-k), \tilde{\varphi}(t-m) \rangle = \int_{-\infty}^{\infty} \varphi(t-k) \cdot \tilde{\varphi}(t-m) dt = \delta_{k-m}.$$

Тоді сигнали в L_2 -мірному просторі визначаються як:

$$\left\{ s(t) = \sum_k c(k) \varphi(t-k) : c \in L_2 \right\}, \quad (3.4)$$

звідки випливає, що будь-яка безперервна функція (сигнал СНС) визначається послідовністю коефіцієнтів $c(k)$ (що не обов'язково є відліками цього сигналу) і сімейством лінійно незалежних базисних функцій $\varphi_k = \varphi(t-k)$, що задовольняють умовам ортонормальності й обмеженості Рісса (Riesz):

$$\sum_k \varphi_k = 1; \quad 0 < \sum |\Phi(f+k)|^2 < +\infty,)$$

де $\Phi(f)$ – перетворення Фур'є базисної функції $\varphi(t)$.

Модель (3.4) описує сигнали при їх вейвлет-розкладанні, багатомасштабних розкладаннях, відновленні В-сплайнами тощо.

3.1.1. Дискретизація сигналів обмеженого спектра

Нескінченні сигнали СОС $x(t)$ у (3.1) фізично спостерігаються на кінцевому інтервалі T_H , тобто вони фінітні:

$$x_T(t) = x(t) \cdot \text{rect}_{T_H}(t),$$

де прямокутна функція $\text{rect}_{T_H}(t) = \begin{cases} 1, t \in [0, T_H]; \\ 0, t \notin [0, T_H] \end{cases}$.

Тоді перетворення Фур'є (ПФ) сигналу $x_T(t)$ визначається як:

$$X_T(f) = \text{ПФ}[x_T(t)] = X(f) * [T_H \cdot \text{sinc}(fT_H)] , \quad (3.5)$$

де $X(f)$ – ПФ сигналу $x(t)$; $*$ – знак згортки; f – поточна частота.

З виразу (3.5) через необмеженість функції $\text{sinc}(fT_H)$ випливає, що спектр $X_T(f)$ фінітного сигналу також необмежений. Для сигналу $x_T(t)$ середньоквадратична похибка апроксимації ε_{FT} , внесена обмеженням його спектра в межах $[-F_{\max}, F_{\max}]$, дорівнює:

$$\varepsilon_{FT} = |X_T(f)|_{f \notin [-F_{\max}, F_{\max}]} \approx (F_{\max} \cdot T_H)^{-1/2}. \quad (3.6)$$

Якщо прийняти припустимим значенням $\varepsilon_{FT} = 0,01 = 1\%$, то інтервал спостереження повинен бути $T_H = 10^4 / F_{\max}$.

Частотна характеристика ідеального ФНЧ має вигляд прямокутної функції $H(f) = \text{rect}_{F_{\max}}(f)$, що не може бути забезпечене фізичними фільтрами. При цьому, якщо задатися коефіцієнтом запасу за частотою дискретизації $\alpha \geq 1$ ($F_{\partial} = \alpha \cdot 2F_{\max}$), то похибка апроксимації ідеального ФНЧ ε_H задовольняє умові:

$$\varepsilon_H \geq |H[(2\alpha - 1) \cdot F_{\max}]|.$$

Обмежуючи величину похибки ε_H , можна обрати тип фільтра H (Баттерворта, Чебишева чи ін.) і коефіцієнт α .

У виразі (3.2) дискретизація здійснюється за допомогою δ -функцій, що мають нескінченно малу тривалість. У реальності ж тривалість імпульсів дискретизації (час виміру відліків безперервної функції) $\theta > \delta$, що призводить до зміни спектра дискретизованого сигналу:

$$X_{\theta}(f) = X(f) \cdot \text{sinc}(f\theta) \cdot \exp(-j\pi f\theta),$$

тобто модуль спектра змінюється в $\text{sinc}(f\theta)$ разів, а зсув фази дорівнює $\pi f\theta$.

Виразимо тривалість імпульсу в частках періоду дискретизації $\theta = \gamma T$, тоді одержимо, що похибка апроксимації ε_{θ} , яка обумовлена не нульовою тривалістю імпульсів дискретизації, буде дорівнювати (для $f = F_{\max}$):

$$\varepsilon_{\theta} \leq 1 - \text{sinc}(\gamma/2\alpha),$$

а зсув фаз – $\gamma/2\alpha$. При припустимій похибці $\varepsilon_{\theta} = 0,01 = 1\%$ одержимо, що $\text{sinc}(\gamma/2\alpha) \geq 0,99$, звідки $\gamma/2\alpha \leq 0,08$. Тоді при $\alpha = 1$ тривалість імпульсу дискретизації повинна бути $\theta \leq 0,16T$, а зсув фаз складе $\pi \cdot 0,08 = 0,25$ радіан чи 14° . При $\alpha = 5$ одержимо $\theta \leq 0,8T$, а зсув фаз той же.

3.1.2. Дискретизація сигналів необмеженого спектра

На практиці передфільтри обирають, виходячи з апіорних знань про сигнал $x(t)$. У загальному випадку передфільтр є не тільки не ідеальним, але і не обов'язково оптимальним, тоді коефіцієнти розкладання рівні $x_1(k) = \langle x, \varphi_1(t-k) \rangle$, де базисною функцією є імпульсна перехідна функція передфільтра $\varphi_1(t) = h(-t)$.

Для компенсації неідеальних характеристик передфільтра можливе використання цифрового корегуючого фільтра з імпульсною перехідною функцією $\beta(t)$ (рис. 3.2).

Тут апроксимація сигналу $x(t)$ визначається як

$$\hat{x}(t) = \sum_k \underbrace{[x_1(k) * \beta(t)]}_{c(k)} \cdot \varphi_2(t-k), \quad (3.7)$$

а передатна функція цифрового корегуючого фільтра дорівнює

$$B(z) = [\sum_k a_{12}(k) \cdot z^{-k}]^{-1},$$

де $a_{12}(k) = \langle \varphi_1(t-k), \varphi_2(t) \rangle$ – коефіцієнти взаємної кореляції перед- і постфільтрів.

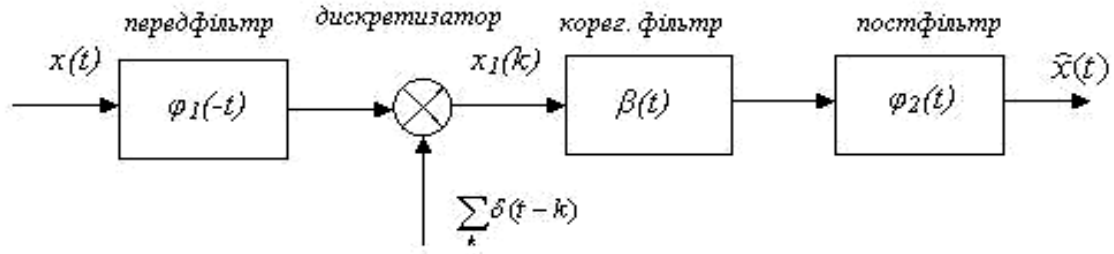


Рисунок 3.2. Схема дискретизації з цифровим корегуючим фільтром

Вираз (3.7) дозволяє відновлювати сигнали СНС при використанні не ідеальних ФНЧ. Стійкість відновлення при цьому залежить від оборотності коефіцієнтів $a_{12}(k)$.

Похибка апроксимації сигналів СНС при класичному підході (3.1) визначається енергією сигналу поза розглянутої смуги частот (вище частоти Найквіста згідно (3.6)), однак при більш загальних підходах (наприклад, при часо-частотних розкладаннях (3.4)) такий підхід не є конструктивним.

Більш ефективним є використання аналізу спектрів Фур'є функціоналів похибок. При цьому ключовим параметром для керування точністю апроксимації є період дискретизації T .

Нехай простір сигналів згідно (3.4) має вид:

$$\left\{ s_T(t) = \sum_k c_T(k) \varphi(t/T - k) : c_T \in L_2 \right\},$$

тоді змінюючи значення T можна одержати значення $T_{\max}$, при якому похибка апроксимації ε_T сигналу $s(t)$ за допомогою коефіцієнтів $c_T(k)$ буде менше деякого припустимого значення $\varepsilon_T^{\text{don}}$: $\varepsilon_T \leq \varepsilon_T^{\text{don}} \Big|_{T=T_{\max}}$.

Оператор апроксимації сигналу $x(t)$ з кроком T дорівнює (див. рис. 3.2):

$$B_T(x) = \sum_k \underbrace{\left[\int x(\tau) \tilde{\varphi}(\tau/T - k) d\tau / T \right]}_{\tilde{c}_T(k)} \cdot \varphi(t/T - k), \quad (3.8)$$

де $\tilde{\varphi}(t)$ – базисні функції, що задовольняють умовам Рісса. Тоді похибка апроксимації визначається як

$$\varepsilon_T = C \cdot T^L \cdot \|x^L\|, \quad (3.9)$$

де C – константа, а порядок похибки апроксимації ε_T дорівнює $O(T^L)$, де L – параметр порядку (для вейвлет-перетворень L – це кількість рівнів розкладання, а для відновлення – це ступінь В-сплайна $l = L - 1$). Чим вище параметр порядку (вище ступінь сплайна і порядок вейвлет), тим менше похибка ε_T .

Вирази (3.8), (3.9) дозволяють обирати період дискретизації T для сигналів СНС, виходячи з умови обмеження похибки апроксимації припустимою величиною $\varepsilon_T \leq \varepsilon_T^{\text{доп}}$.

3.1.3. Відновлення сигналів

У виразі (3.1) відновлення сигналів СОС здійснюється за допомогою *sinc*-функцій, що мають низьку роздільну здатність. Більш простою і широко використовуваною є лінійна інтерполяція сигналу $x(t)$ в дискретних відліках з наступною низькочастотною фільтрацією. При цьому частота дискретизації залежить від припустимої похибки відновлення сигналу $\varepsilon_{\text{відн}}$ і його форми.

Наприклад, для гармонійних сигналів частота дискретизації обирається з умови:

$$F_{\partial} = T^{-1} \geq 2,2 F_{\text{max}} \cdot (\varepsilon_{\text{відн}})^{-1/2}.$$

Якщо прийняти припустимим значення похибки $\varepsilon_{\text{відн}} = 0,01 = 1\%$, то частота дискретизації повинна бути $F_{\partial} \geq 22 F_{\text{max}}$.

В умовах перевищення періоду технологічних випробувань керованого процесу $T_{\text{опр}}$ над необхідним періодом дискретизації T ($T_{\text{опр}} > T$) актуальним є відновлення пропущених даних у часових послідовностях (сигналах СНС).

Нехай ϵ дискретна послідовність $x_2(k)$ сигналу СНС із періодом дискретизації T у вигляді моделі

$$x_2(k) = \sum_m c_2(m) \cdot \varphi_2(k - m) , \quad (3.10)$$

яку необхідно відновити (інтерполювати) у дискретну послідовність з періодом дискретизації T/M :

$$x_2(k/M) = \sum_m c_2(m) \cdot \varphi_2(k/M - m) . \quad (3.11)$$

Схема відновлення дискретної послідовності наведена на рис. 3.3.

Вхідна послідовність $x_2(k)$ спочатку проходить через передфільтр із передатною функцією $\Phi_2^{-1}(z)$, зворотною z -перетворенню базисної функції $\varphi_2(k/M)$, далі використовується розширювач відліків у M разів ($\uparrow M$) і фільтр-інтерполятор $\Phi_2(z)$. Як базисні функції використовуються l -разів диференційовані гладкі функції (наприклад, В-сплайни ступеня $l+1$).

Таким чином, інтерполяція $x_2(k)$ у $x_2(k/M)$ виконується цілком у цифровому вигляді.

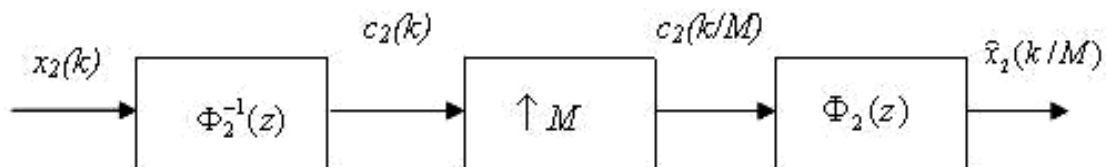


Рисунок 3.3. Цифрова інтерполяція дискретної послідовності

Наступною проблемою, що виникає при розв'язанні задач спостереження й ідентифікації, є відновлення нерівномірно дискретизованої (у відомі моменти часу) послідовності (сигналу СНС).

Якщо базисні функції $\varphi_2(k)$ в (3.9) компактні (наприклад, вейвлет-функції чи сплайни), то відновлення відліків $c(k_m)$ (вимірюваних нерівномірно в

моменти часу $\{k_m\}, m = \overline{1, N}$ у рівномірні відліки $c(k)$ з періодом T можна здійснити за допомогою послідовних перетворень (рис. 3.4):

$$x_2(k) = \sum_m c_2(k_m) \cdot \phi_2(k_m - m) ; \quad (3.12)$$

$$c_2(k) = \sum_m x_2(m) \cdot \gamma(k - m) , \quad (3.13)$$

де $\gamma(k)$ – базисна функція, що зворотна функції $\phi_2(k)$ і має z -перетворення $\Phi_2^{-1}(z)$. Тут передфільтр $\Phi_2(z)$ має структуру фільтра з кінцевою імпульсною характеристикою (КІХ), а фільтр-інтерполятор $\Phi_2^{-1}(z)$ – з нескінченною імпульсною характеристикою (НІХ).

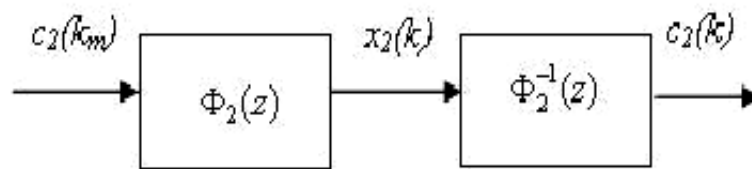


Рис. 3.4. Відновлення нерівномірно дискретизованої послідовності

Такий підхід дозволяє виключити необхідність використання зверхдискретизації.

Похибки апроксимації при інтерполяції (3.10)-(3.11) і відновленні (3.12)-(3.13) сигналів СНС визначаються строгістю виконання рівності $\Phi_2^{-1}(z) = [\Phi_2(z)]^{-1}$ (реалізацією фільтра $\Phi_2^{-1}(z)$) із похибкою:

$$\varepsilon_\Phi = \left\| \Phi_2^{-1}(z) - [\Phi_2(z)]^{-1} \right\|.$$

Більш складною є задача відновлення сигналів за дискретними відліками у нерівномірні і невідомі моменти часу (є задачею комбінаторної оптимізації).

Нехай необхідно відновити K відліків дискретного сигналу СОС (зі спектром довжиною L) $x = (x_0, \dots, x_{K-1})^T$ по N нерегулярним і невідомим у часі відлікам вимірюного сигналу $s = x(K_n) = (x_{k_1}, x_{k_2}, \dots, x_{k_n})^T$, де $K_n = \{k_i\}; i = \overline{1, N}$

– нерегулярно розташована підмножина відліків від множини $K = \{0,1,\dots,K-1\}$;
 $L \leq N < K$.

Перетворення Фур'є шуканого сигналу дорівнює:

$$X(L) = ДПФ_K \cdot x(K), \quad (3.14)$$

де $ДПФ_K$ – матриця коефіцієнтів дискретного ПФ: $W_K = \exp(-j2\pi/K)$;
 $W_K^k = W_K^{k \bmod K}$.

З виразу (3.14) отримаємо:

$$x(K) = K^{-1} \cdot F(K, \Lambda) \cdot X(\Lambda), \quad (3.15)$$

де $F(K, \Lambda) = (ДПФ_K)^{-1} = K^{-1} \cdot ДПФ_K^*$; $\Lambda = \{0,1,\dots,L-1\}$.

Аналогічно вимірний сигнал дорівнює:

$$s = K^{-1} \cdot F(K_n, \Lambda) \cdot X(\Lambda), \quad (3.16)$$

де $F(K_n, \Lambda) = (W^{-k_1 l}, W^{-k_2 l}, \dots, W^{-k_n l})^T$; $l \in \Lambda$.

У виразі (3.16) відомими є s, Λ, K, N , а невідомими – x, K_n . Підставивши рівняння (3.16) у (3.15), одержимо залежність шуканого сигналу від вимірюваного:

$$x(K) = F(K, \Lambda) \cdot FF(K_n, \Lambda) \cdot s, \quad (3.17)$$

де $FF(K_n, \Lambda) = [F(K_n, \Lambda)^* \cdot F(K_n, \Lambda)]^{-1} \cdot F(K_n, \Lambda)^*$.

Рішення (3.17) регуляризується шляхом виконання рівності відновленого й вимірюваного сигналів у визначені моменти часу $x(M_n) = s$. Тоді з урахуванням (3.17) похибка розв'язання повинна дорівнювати:

$$e(M_n) = \{F(M_n, \Lambda) \cdot FF(M_n, \Lambda) - I\} \cdot s = 0,$$

де I – одинична матриця.

Пошук рішень системи нелінійних рівнянь (3.17) аналітичним шляхом неможливий, тому для визначення сигналу $x(K)$ використовуються методи комбінаторної оптимізації, метою якої є мінімізація норми похибки $e(M_n)$ (менше припустимого значення):

$$\varepsilon_{peg} = \|e(M_n)\| \rightarrow \min \leq \varepsilon_{peg}^{don}. \quad (3.18)$$

Легко бачити, що глобальним мінімумом (3.18) є $\varepsilon_{\text{рег}} = 0$.

3.2. Адаптивні фільтри-апроксиматори

У теорії керування спостереження за деяким заданим сигналом з можливо меншою похибкою зводиться до задачі фільтрації.

Для реалізації систем оцінювання і ідентифікації об'єктів керування (ОК), зазвичай, використовуються адаптивні фільтри-апроксиматори (АФА), які мають рекурсивні і нерекурсивні зв'язки, а також виконують перетворення сигналів. У них процес адаптації включає оцінювання шуканого виходу фільтру і корегування його параметрів за значенням вихідної похибки.

Структурні схеми використання АФА як прогнозуючий фільтр сигналів і апроксиматор (еталонна модель) ОК наведені на рис. 3.5.

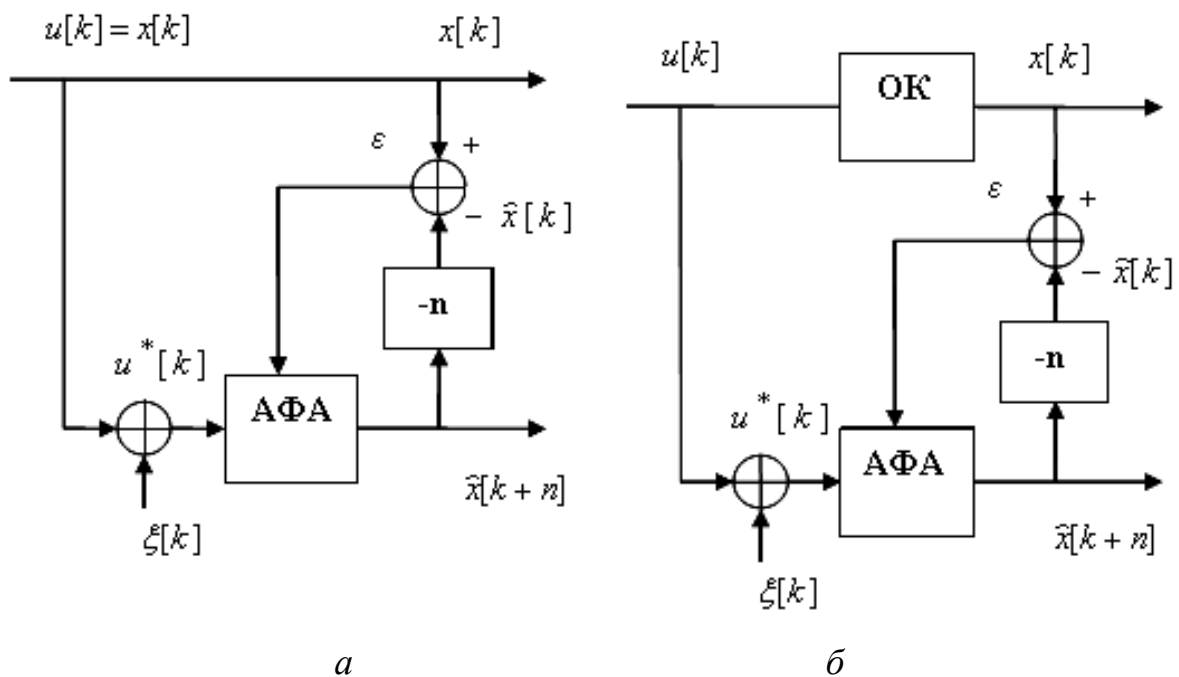


Рисунок 3.5. Прогнозування сигналу (*a*) та ідентифікація ОК (*б*) за допомогою АФА

Тут затримка і прогноз на n тактів позначені, як $-n$ і $+n$, а виміряні значення входів рівні:

$$u^*[k] = u[k] + \xi[k], \quad (3.19)$$

де $\xi[k]$ – похибка (шум) вимірювання.

В процесі роботи АФА на кожному такті по величині похибки $\varepsilon[k] = x[k] + \hat{x}[k]$ між фактичним $x[k]$ і прогнозованим $\hat{x}[k]$ значеннями сигналу (виходу ОК) здійснюється адаптація параметрів АФА. При цьому, як структуру АФА зазвичай використовують лінійні поліноміальні моделі, які для складних динамічних ОК не є адекватними.

Жорсткі вимоги до знання статистичних властивостей часових рядів обмежують можливості методів математичної статистики, теорії розпізнавання образів, теорії випадкових процесів тощо. Багато реальних процесів не можуть адекватно бути описані за допомогою традиційних статистичних моделей, оскільки є істотно нелінійними і мають або хаотичну, або квазіперіодичну, або змішану основу.

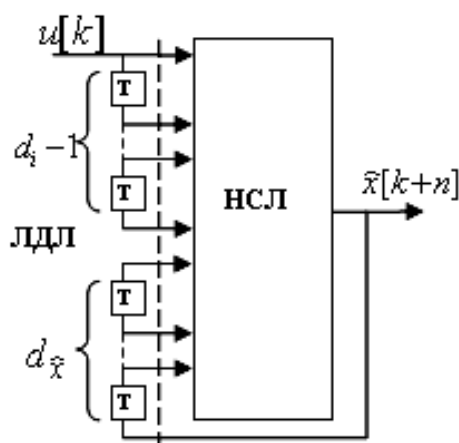


Рисунок 3.6. Структура Вінера-Гаммерштейна прогнозуючої моделі нелінійного динамічного ОК

Для побудови і реалізації структури динамічної прогнозуючої моделі ОК можуть використовуватися різні підходи. При цьому відомо, що нелінійна динамічна система (модель ОК) може бути представлена шляхом композиції лінійного динамічного (ЛДЛ) і нелінійного статичного (НСЛ) ланок, наприклад, у вигляді моделі Вінера-Гаммерштейна (Wiener-Hammerstein), яка наведена на рис. 3.6. Тут ЛДЛ є лініями затримки, величини яких (глибина пам'яті) визначаються розмірністю входних $d_i - 1$ і вихідних $d_{\hat{x}}$ змінних. А як НСЛ

можуть використовуватися як традиційні засоби (поліноми Лежандра, Вольтерра, Колмогорова-Габора тощо), так і інтелектуальні (НМ, гібридні НМ з нечіткою логікою тощо).

Переважними тут є засоби інтелектуальної обробки інформації, оскільки вони легко настроюються (адаптуються) під властивості ОК, що змінюються, і, відповідно, є ефективними засобами моделювання складних систем.

3.2.1. Лінійний адаптивний фільтр-апроксиматор

Сигнал на виході лінійного АФА визначається згорткою вхідного сигналу з імпульсною перехідною характеристикою фільтру:

$$x[k] = \sum_{m=-\infty}^{\infty} u[m] \cdot h[k-m] = \sum_{m=-\infty}^{\infty} h[m] \cdot u[k-m], \quad (3.20)$$

де $u[k]$ – вхідний сигнал, $x[k]$ – вихідний сигнал, $h[k]$ – імпульсна перехідна характеристика.

Різницеве рівняння прогнозуючого лінійного АФА з кінцевою імпульсною характеристикою (КІХ) має вигляд:

$$\hat{x}[k+n] = \sum_{r=0}^R b_r \cdot u[k-r], \quad (3.21)$$

де b_r , R – коефіцієнти фільтру і його порядок, відповідно.

Параметрами настройки (навчання) цього АФА є $\{b_r\} \subset a$.

АФА з КІХ мають асимптотичну стійкість і лінійну фазочастотну характеристику.

Різницеве рівняння лінійного АФА з нескінченною імпульсною характеристикою (НІХ) має вигляд:

$$\hat{x}[k+n] = \sum_{m=1}^M a_m \cdot x[k-m] + \sum_{r=0}^R b_r \cdot u[k-r], \quad (3.22)$$

де a_m , M – коефіцієнти і порядок зворотного зв'язку фільтру, відповідно.

Параметрами настройки (навчання) цього АФА є $\{a_m, b_r\} \subset a$.

Ця рекурентна формула використовується для послідовного обчислення поточного значення вихідного сигналу АФА за його минулими значеннями. Фільтри з НІХ мають більшу гнучкість при реалізації, ніж з КІХ.

У більшості алгоритмів адаптації лінійних АФА мінімізується середньоквадратична похибка шляхом апроксимації градієнта цільової функції і ітеративного підстроювання коефіцієнтів фільтру.

3.2.2. Нейромережевий адаптивний фільтр-апроксиматор

Для прогнозуючих систем на базі НМ якнайкращі якості показує гетерогенна мережа, яка складається з прихованих шарів з нелінійною функцією активації нейронів і вихідного лінійного нейрона.

Рівняння ОК, що здійснює прогноз на n тактів, на основі НМ прямого поширення (НМПП) з прихованим шаром представляється у вигляді рівняння згортки:

$$\hat{x}[k+n] = \sum_{\tau \in P} F_{\hat{x}} \left\{ \sum_{l \in Q} v_l[\tau] \cdot F_l \left(\sum_{m \in Q} v_{l,m}[\tau] \cdot u_m[k-\tau] \right) \right\}, \quad (3.23)$$

де P – множина глибин пам'яті відповідних входів; $F_{\hat{x}}$ – активаційна функція вихідного шару НМ; Q – множина входів нейронів; l – порядковий номер входу вихідного шару НМ; v_l – вагові коефіцієнти вихідного шару; F_l – активаційна функція нейронів прихованого шару; m – порядковий номер входу НМ; $v_{l,m}$ – вагові коефіцієнти зв'язку m -го входу і l -го нейрона; u_m – вхід НМ.

Параметрами настройки (навчання) цієї НМ є $\{v_l, v_{l,m}\} \subset a$.

Прогнозування за допомогою НМ з радіальними базисними функціями (РБФ) виконується згідно рівняння:

$$\hat{x}[k+n] = \sum_{\tau \in P} F_{\hat{x}} \left\{ \sum_{l,m \in Q} v_l \cdot F_l(\vartheta_l, \|u_m[k-\tau] - v_l\|) \right\}, \quad (3.24)$$

де ϑ_l, v_l – параметри РБФ l -го нейрона прихованого шару.

Параметрами настройки НМ (3.23) є $\{v_l, \vartheta_l, v_l\} \subset a$.

Ідентифікація параметрів (настройка – навчання) НМ прямого розповсюдження (3.23) здійснюється, як правило, за допомогою градієнтних алгоритмів, наприклад, алгоритму зворотного розповсюдження помилки в просторі параметрів $\{v_l, v_{l,m}\} \subset a$ при заданих структурі моделі ОК (архітектурі НМ) і структурних функціях F з метою мінімізації похибки прогнозу:

$$\bar{\varepsilon} = E\{\|x[k+n] - \hat{x}[k+n]\|\} \rightarrow \min, \quad (3.25)$$

де E – математичне очікування.

При навчанні НМ з РБФ (3.24) спочатку визначаються центри і відхилення для радіальних елементів, після цього оптимізуються параметри лінійного вихідного шару: $\{v_l, \vartheta_l, v_l\} \subset a$.

3.2.3. Нейронечіткий адаптивний фільтр-апроксиматор

Теоретично, системи з нечіткою логікою і НМ еквівалентні один одному, проте на практиці у них є свої переваги і недоліки. У зв'язку з цим отримали розвиток гібридні НМ, в яких виводи робляться на основі апарату нечіткої логіки, а функції належності підстроюються з використанням алгоритмів навчання НМ. Такі системи не тільки використовують апріорну інформацію, але можуть набувати нових знань і для користувача є логічно прозорими.

Гібридна мережа – це НМ з чіткими сигналами, вагами і активаційною функцією, але з об'єднанням сигналів і ваг мережі з використанням t -норми, t -конорми або деяких інших безперервних операцій. Входи, виходи і ваги гібридної НМ – дійсні числа, які належать відріzk [0, 1].

Прикладом гібридної мережі є Anfis (адаптивна нейронна система нечіткого висновку структури Сугено), рівняння ОК на основі якої має вид:

$$\hat{x}[k+n] = \sum_{\tau \in P} \sum_{m \in Q} \beta_m[\tau] \cdot \alpha_m[k-\tau], \quad (3.26)$$

де $\beta_m[\tau] = U_m^{-1}(\alpha_m[\tau] / \sum_m \alpha_m[\tau])$; $\alpha_m[k-\tau] = Tn_{l,m}\{L_{l,m}(u_m[k-\tau])\}$; $U = U(a_U)$;

$L = L(a_L)$.

Тут U_m^{-1} – функція, зворотна функції належності проміжного виходу m мережі з параметрами a_U ; α_m – значення проміжного виходу; Tn – довільна t -норма моделювання логічної операції «І»; $L_{l,m}$ – функція належності нечіткого правила l входу m з параметрами a_L .

Параметрами настройки НМ (3.26) є $\{a_U, a_L\} \subset a$, її навчання виконується аналогічно НМ (3.24).

Перевагою алгоритмів параметричного навчання НМ є простота і швидкодія, а недоліком – їх локальність (висока вірогідність застрягання в локальному екстремуму).

3.2.4. Нечіткий адаптивний фільтр-апроксиматор з субтрактивною кластеризацією

Нечітка АФА за допомогою субтрактивної кластеризації генерує нечітку логічну систему типу Сугено із вхідних даних. Витяг правил з даних відбувається у два етапи. По-перше, визначається кількість правил і повноважень термінових наборів вхідних змінних. Потім, використовуючи метод найменших квадратів, визначається "тоді" – частина кожного правила. В результаті виходить система нечіткої логічної дедукції з базою правил, що охоплюють всю предметну область.

Алгоритм з субтрактивною кластеризацією (Genfis2) включає:

1. Розрахунок потенціалу кожної точки (міри просторової близькості між нею та іншими із M):

$$E(x_m) = \frac{1}{M} \sum_{i=1}^M e^{-\frac{\|x_m - x_i\|^2}{(R_c/2)^2}}. \quad (3.27)$$

2. Встановлення кількості кластерів $k_c = 0$.

3. Визначення точки з найвищим потенціалом $E(x_p)$, x_p :

$$p = \arg \max_{i=1}^M E(x_i). \quad (3.28)$$

4. Ініціалізація центра j -го кластера:

$$k_{c_j} = x_p. \quad (3.29)$$

5. Зниження потенціалу всіх точок:

$$E(x_i) = E(x_i) - E(k_{c_j}) e^{-\frac{\|k_{c_j} - x_i\|}{(r/2)^2}}, \quad r = [1, 1.5] R_c. \quad (3.30)$$

6. Перевірка значення потенціалу відносно встановленого порогу thr :

$$\max_{i=1}^M E(x_i) < thr. \quad (3.31)$$

Якщо умова (3.31) виконується, то «кінець», інакше повинен бути перехід до пункту 3.

Таким чином, рівняння АФА Genfis2 можна представити у вигляді:

$$\hat{x}[k+n] = FF_j(FV(a_{FV}, \sum_{\tau \in P} \sum_{h=1}^{k_c} PV_h(D_{mh}))), \quad (3.32)$$

де $D_{mh} = \|z_h - y_m[k-\tau]\|$; $PV_h = \sum_{m=1}^M \exp(-\alpha \cdot D_{mh})$; $j = \overline{1, 3}$; $k_c = \overline{1, n}$.

Налаштуванням цього АФА є $\{\alpha, R_c\} \subset a_{FV}$.

3.2.5. Нечіткий адаптивний фільтр-апроксиматор з кластеризацією С-середніх

В основі алгоритму кластеризації С-середніх (Genfis3) лежить метод невизначених множників Лагранжа, що дозволяє задачі знаходження умовного екстремуму цільової функції на множині допустимих значень перетворитись на задачу безумовної оптимізації функції.

Алгоритм Genfis3 - це ітеративна процедура, в якій виконуються наступні кроки:

1. Встановлення нечітких кластерів у матрицю розділів

$$M_D = [\mu_{\theta i}]; \mu_{\theta i} \in [0, 1]; \theta = \overline{1, \Theta}; i = \overline{1, k_c}. \quad (3.33)$$

де

$$\sum_{i=1}^{k_c} \mu_{\theta i} = 1; \quad 0 < \sum_{\theta=1}^{\Theta} \mu_{\theta i} < \Theta. \quad (3.34)$$

2. Встановлення початкових значень параметрів: кількість кластерів k_c ; експоненціальна вага $\varpi \in [1, \infty]$; параметр зупинки алгоритму ε .

3. Створення випадковим чином матрицю нечіткої секції з урахуванням умов (3.34).

4. Розрахунок центрів кластерів Ω_i :

$$\Omega_i = \sum_{\theta=1}^{\Theta} \mu_{\theta i}^{\varpi} * |X_{\theta}| / \sum_{\theta=1}^{\Theta} \mu_{\theta i}^{\varpi}; i = \overline{1, k_c}. \quad (3.35)$$

5. Розрахунок відстані між об'єктами від матриці спостереження X і центрами кластерів:

$$D_{\theta i} = \sqrt{\|X_{\theta} - \Omega_i\|^2}. \quad (3.36)$$

6. Перерахунок елементів матриці розділів:

- якщо $D_{\theta i} > 0$, то

$$\mu_{\theta i} = 1 / \left(D_{j\theta}^2 * \sum_{j=1}^{k_c} \frac{1}{D_{j\theta}^2} \right)^{1/(\varpi-1)}, \quad (3.37)$$

- якщо $D_{\theta i} = 0$, то

$$\mu_{\theta i} = \begin{cases} 1, j = i \\ 0, j \neq i \end{cases}; j = \overline{1, k_c}. \quad (3.38)$$

7. Перевірка умови:

$$\|M_D - M_D^*\| < \varepsilon. \quad (3.39)$$

Якщо «ні», то перейдіть до кроку 4, якщо "так", то «закінчіть».

Тоді рівняння АФА Genfis3 можна представити як рівняння згортки:

$$\hat{Y}[k+n] = FF_j(FD(a_{FV}, \sum_{\tau \in P} \sum_{h=1}^Q V_h(D_{mh}(z_n, y_m[k-\tau])))), \quad (3.40)$$

де $D_{mh} = \|V_h - y_m[k-\tau]\|$; $V_h = \sum_{m=1}^M (\mu_{mh})^{\alpha} \cdot y_m[k-\tau] / \sum_{k=1}^M (\mu_{mh})^{\alpha}$; $h = \overline{1, Q}$; $m = \overline{1, M}$;

$F = \{\mu_{mh}\}$; $\mu_{mh} \in [0, 1]$; $\sum_{h=1}^Q \mu_{mh} = 1$; $0 < \sum_{m=1}^M \mu_{mh} < M$.

Налаштуванням цього АФА є $\{Q, M, L, \varepsilon\} \subset a_{FV}$.

3.2.6. Нейронний вейвлет-адаптивний фільтр-апроксиматор

У фільтрах з перетворенням, наприклад, в частотній області в порівнянні зі згорткою в часовій області, значно скорочується об'єм обчислень і поліпшуються властивості збіжності алгоритмів адаптації. При цьому, зважаючи на здібність до представлення нестационарних сигналів часто вибирають не частотне, а час-частотне (вейлет) перетворення, причому, дискретне, із-за властивостей ортонормованості, компактності, відновлення без втрат і малого об'єму обчислень.

Прикладом такого фільтра є нейронний вейвлет (НВ) АФА (рис. 3.7):

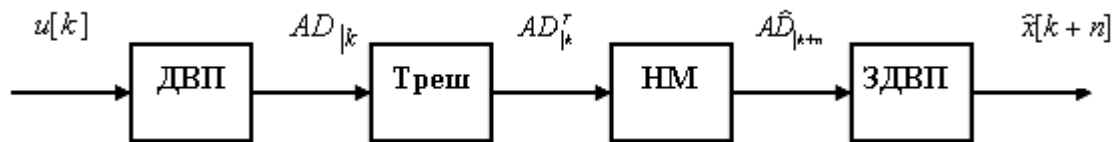


Рисунок 3.7. Структура прогнозуючого нейронного вейвлет АФА

У ньому в основу фільтрації покладена процедура прямого дискретного вейвлет перетворення (ДВП). Для зменшення впливу шуму у фільтрі здійснюється порогове обмеження коефіцієнтів вейвлет розкладання – трешолдинг (Треш), а НМ застосовується для прогнозування значень коефіцієнтів, за якими за допомогою зворотного ДВП (ЗДВП) визначається прогнозований сигнал.

У НВ АФА для вхідного дискретного сигналу $u[k]$ тривалістю K ($k = \overline{1, K}$) його ДВП є набір вейвлет коефіцієнтів

$$AD|_k = \{A_L, D_L, D_{L-1}, \dots, D_1\} \quad (3.41)$$

з коефіцієнтами апроксимації по рівнях розкладання

$$A_L = \{a_{L,p}\} = \sum_{p=1}^{P_L} u[k] \cdot \varphi_{L,p}[k] \quad (3.42)$$

і коефіцієнтами деталізації

$$D_l = \{d_{l,p}\} = \sum_{l=1}^L \sum_{p=1}^{P_l} u[k] \psi_{l,p}[k]; \quad l = \overline{1, L}, \quad (3.43)$$

де φ, ψ – масштабуюча і відповідна їй вейвлет функції; L – кількість рівнів розкладання; P_l – кількість коефіцієнтів на рівні розкладання l .

Коефіцієнти апроксимації визначають згладжений сигнал, а деталізації – його коливання. Для видалення шуму коефіцієнти, менші деякого граничного значення, обнулюються (процедура порогової обробки – трешолдинг T):

$$AD_k^T = T(AD_k).$$

На вході НМ вейвлет коефіцієнти помножуються на відповідні ваги і визначаються рівні активації нейронів, які формують прогноз вейвлет коефіцієнтів для моменту часу $[k + n]$:

$$AD_{k+n}^T = F_{NN}(AD_k^T), \quad (3.44)$$

де F_{NN} – узагальнена функція перетворення НМ.

Для забезпечення асимптотичної стійкості АФА використовуються НМ без зворотних зв'язків і з нелінійними функціями активації (з-за нелінійності ОК і сигналів).

Тоді прогнозування сигналу $\hat{x}[k + n]$ глибиною n виконується за прогнозованими за допомогою НМ коефіцієнтами розкладання $AD_{k+n}^T = \{\hat{a}_{L,p}, \hat{d}_{l,p}\}$ відповідно до виразу ЗДВП:

$$\hat{x}[k + n] = \sum_{p=1}^{P_L} \hat{a}_{L,p} \cdot \varphi_{L,p}[k + n] + \sum_{l=1}^L \sum_{p=1}^{P_l} \hat{d}_{l,p} \psi_{l,p}[k + n]. \quad (3.45)$$

Навчання НВ АФА здійснюється в пакетному режимі, а адаптація – в реальному часі шляхом підстроювання параметрів АФА при обробці поточного блоку за похибкою прогнозу попереднього блоку тривалістю K .

До параметрів АФА, які навчаються (адаптуються), відносяться тип базисного вейвлета ψ , кількість рівнів розкладання L , а також параметри трешолдинга і НМ.

3.2.7. Вейвнет АФА з блоком, який має нескінченну імпульсну характеристику

Для складних динамічних систем (процесів), що характеризуються нестационарністю, стохастичністю і нелінійністю, актуальною є розробка та вибір відповідних засобів для їх ідентифікації і прогнозування.

Наразі перспективним вважається використання при побудові АФА нейронних мереж (НМ), зокрема, НМ із функціями активації у вигляді вейвлет (вейвнетів – **Wavenet Network**). Такі АФА апроксимують нелінійні функції з високою точністю, але мають недостатні прогностичні властивості, що особливо важливо для процесів із значними запізненнями.

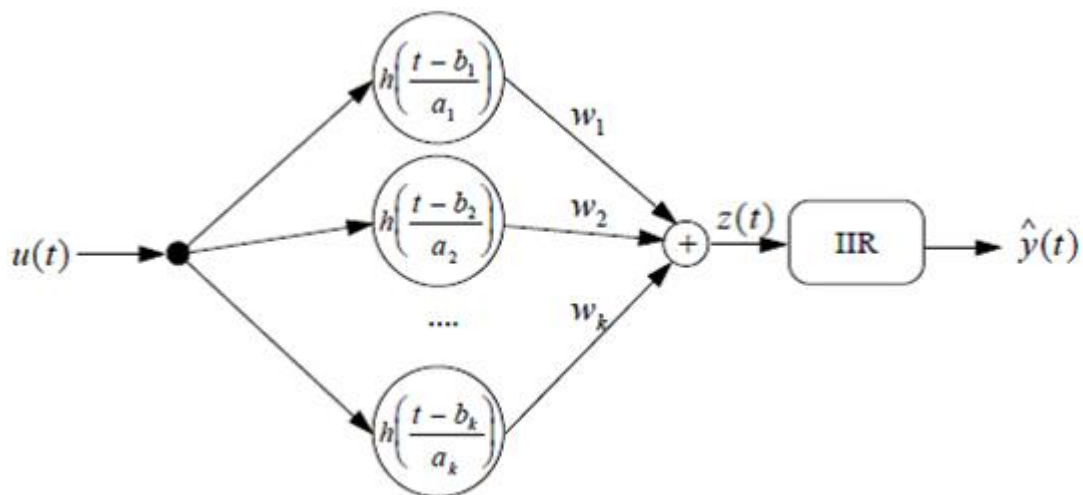
Для подолання цього недоліку було запропоновано будувати АФА у вигляді комбінації вейвнету і фільтра з нескінченною імпульсною характеристикою (блок НІХ), які утворюють структуру Гаммерштейна нелінійної динамічної системи.

Вейвнет є мережею, в якій функція виходу добре локалізована в часовій і в частотній областях. Комбінування каскадної побудови вейвнет з блоком НІХ забезпечує обчислювально ефективне навчання системи і зменшує час досягнення мінімально допустимої похибки системи. В вейвнеті (рис. 3.8) формується оцінка бажаного сигналу $y(t)$ шляхом лінійної комбінації набору вейвлет-функцій $h_{a,b}(t)$ із каскадним розташуванням та подальшого перетворення в динамічному блоці НІХ (ІІР). Вихід такої мережі описується виразом:

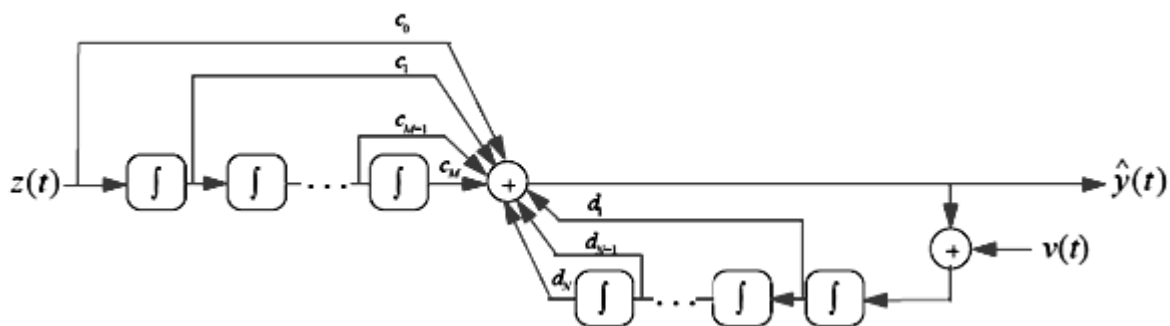
$$\hat{y}(t) = \sum_{i=0}^M c_i z(t-i)u(t) + \sum_{j=1}^N d_j \hat{y}(t-j)v(t), \quad (3.46)$$

де $z(t) = \sum_{k=1}^K w_k h_{a_k, b_k}(t)$. Тут K – число вейвлетів, w_k – k -ий ваговий коефіцієнт,

M і c_i – кількість зворотних зв'язків затримки входу і коефіцієнтами прямої гілки НІХ фільтра, відповідно, N і d_j – це кількість зворотних зв'язків затримки виходу та коефіцієнти зворотної гілки, відповідно.



a



б

Рисунок 3.8. Структура вейвнету із блоком НІХ (а) та структура блоку НІХ (б)

Сигнали $u(t)$ і $v(t)$ є вхідними сигналами системи в момент часу t (вхід $v(t)$, зазвичай, залишається невеликим для стабільності зворотного зв'язку). В процесі адаптації (навчання) кожен вектор коефіцієнтів w , a , b , c і d мережі оновлюється за рівнянням:

$$\bar{c}(n+1) = \bar{c}(n) + \mu_c \Delta \bar{c}; \quad \bar{c} = \{w, a, b, c, d\}, \quad (3.47)$$

де μ – значення коефіцієнта швидкості навчання параметрів; $\Delta \bar{c}$ – градієнт похибки навчання за відповідним коефіцієнтом.

Загальною проблемою НІХ адаптивної мережі є проблема забезпечення стабільності та збіжності. Зокрема, НІХ адаптивні мережі схильні до нестабільності, необмеженого росту коефіцієнтів при адаптації. Крім того, рекурсивна адаптація коефіцієнтів створює рух, в результаті чого мережа стає

нестійкою, навіть якщо адаптація є стабільною. Крім того, використання градієнтного алгоритму адаптації, який застосовується для мінімізації похибки, часто приводить до застрягання в локальному мінімумі.

Незважаючи на ці недоліки потенціал адаптивних мереж з блоком НІХ дуже високий, що і підтримує інтерес до таких структур.

Проблема потенційної адаптивної нестабільності може бути вирішена шляхом контролю та уникнення будь-яких оновлень, які призводять до нестійкості системи або до послідовного зниження швидкості навчання.

3.3. Контрольні питання до глави 3

1. Яка мета дискретизації в задачах спостереження та ідентифікації?
2. Розкрийте суть теорем дискретизації та відновлення Котельникова-Шеннона.
3. Які похибки виникають при дискретизації сигналів з обмеженим спектром?
4. Які особливості дискретизації сигналів з необмеженим спектром?
5. Які особливості відновлення дискретизованих сигналів?
6. Сформулюйте задачу адаптивного прогнозування сигналів і стану об'єктів керування.
7. Розкрийте сутність лінійного адаптивного фільтра-апроксиматора.
8. Опишіть структуру, параметри та процес навчання нейромережевого адаптивного фільтра-апроксиматора.
9. Розкрийте структуру, параметри та процес навчання нейронечіткого адаптивного фільтра-апроксиматора.
10. Опишіть структуру, параметри та процес навчання нечіткого адаптивного фільтра-апроксиматора з субтрактивною кластеризацією.
11. Розкрийте структуру, параметри та процес навчання нечіткого адаптивного фільтра-апроксиматора з кластеризацією С-середніх.
12. Опишіть структуру, параметри та процес навчання нейронного вейвлет адаптивного фільтра-апроксиматора.

13. Розкрийте структуру побудови ґейнґу з блоком із нескінченною імпульсною характеристикою.

Глава 4. Застосування адаптивних фільтрів-апроксиматорів до розв'язування прикладних задач

4.1. Адаптивне оцінювання та ідентифікація процесів на основі вейвнетів

4.1.1. Фільтрація голосу

Наразі вейвнети широко застосовуються для аналізу та машинного синтезу голосу.

Розглянемо та порівняємо процеси навчання між мережами вейвнет із блоком НІХ і без нього при моделюванні жіночої голосової фонемі літери «а».

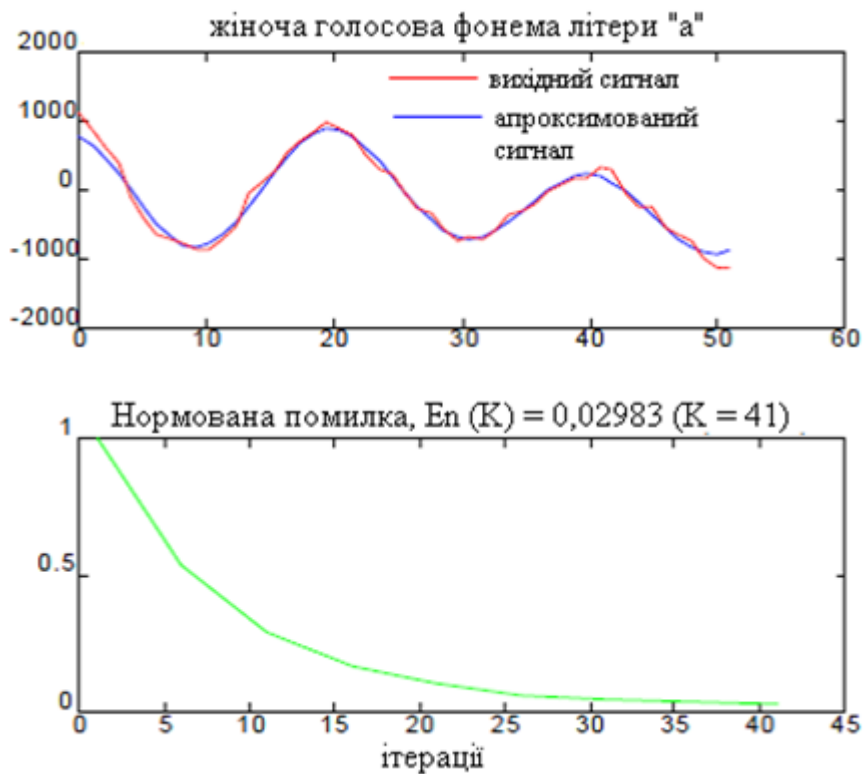


Рисунок 4.1. Процес навчання вейвнету без блоку НІХ

Обидві мережі мають однакові набори вхідних параметрів. Швидкість навчання параметрів вагової функції, масштабування, зсуву, коефіцієнти зворотного зв'язку блоку НІХ і коефіцієнти прямого зв'язку мають значення:

0.01, 0.05, 0.05, 0.02 і 0.02, відповідно. Навчання здійснювалось в пакетному режимі на основі 50 вибірок даних.

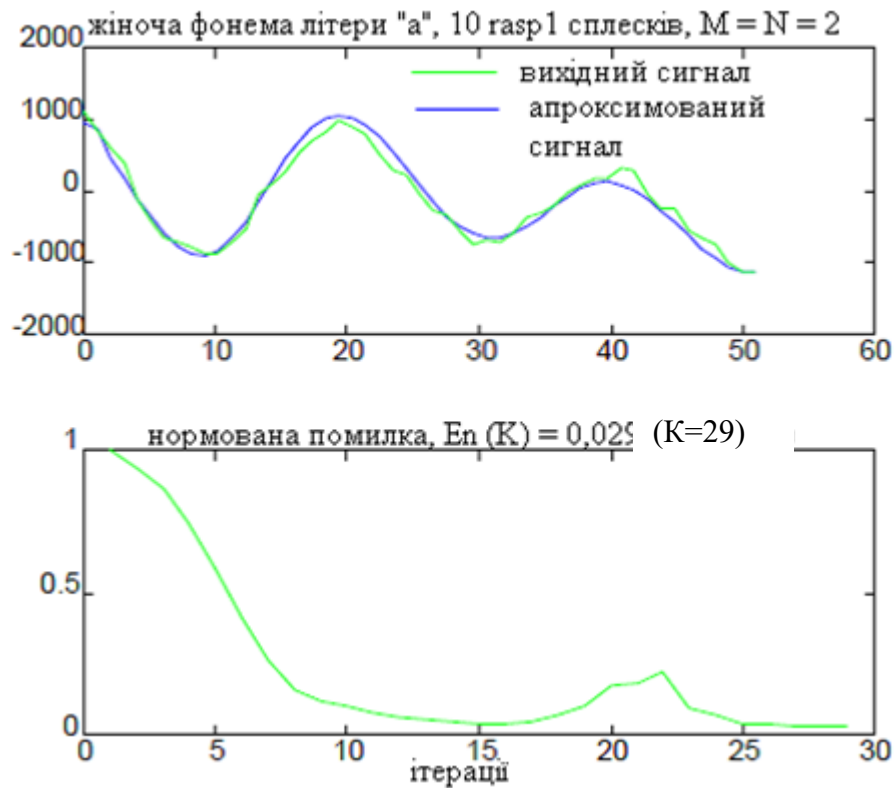


Рисунок 4.2. Процес навчання вейвнету з блоком НІХ

Таблиця 4.1. Кількість ітерацій для досягнення бажаної похибки

Бажана нормалізована похибка	Кількість ітерацій	
	Без блоку НІХ	З блоком НІХ
0.8	4	4
0.5	6	6
0.3	10	7
0.1	20	11
0.05	29	14
0.04	34	27
0.03	41	29

Мережі адаптувались до тих пір, поки очікувана похибка в 0.03 не буде досягнута. Нормованість похибки визначається як $E_n(t) = (e^T \cdot e) / (y^T \cdot y)$, де T – транспонування вектора. В обох мережах використовувались 10 материнських вейвлетів типу раціональних функцій другого порядку (RASP1).

Результати моделювання наведені на рис. 4.1 і 4.2 та в табл. 4.1.

Результати моделювання підтверджують ефективність використання вейвнету з блоком НІХ (бажана нормалізована похибка досягається на 29 ітерації, на відміну від 41 ітерації – для вейнету без блоку НІХ). Тобто, вейвнет з блоком НІХ має в 1,41 раз більшу швидкість збіжності.

4.1.2. Фільтрація голосу за наявності імпульсних перешкод

Однією з переваг використання адаптивних НМ є їх стійкість до завад, наприклад, коли сигнали забруднені негаусовими шумами чи завадами.

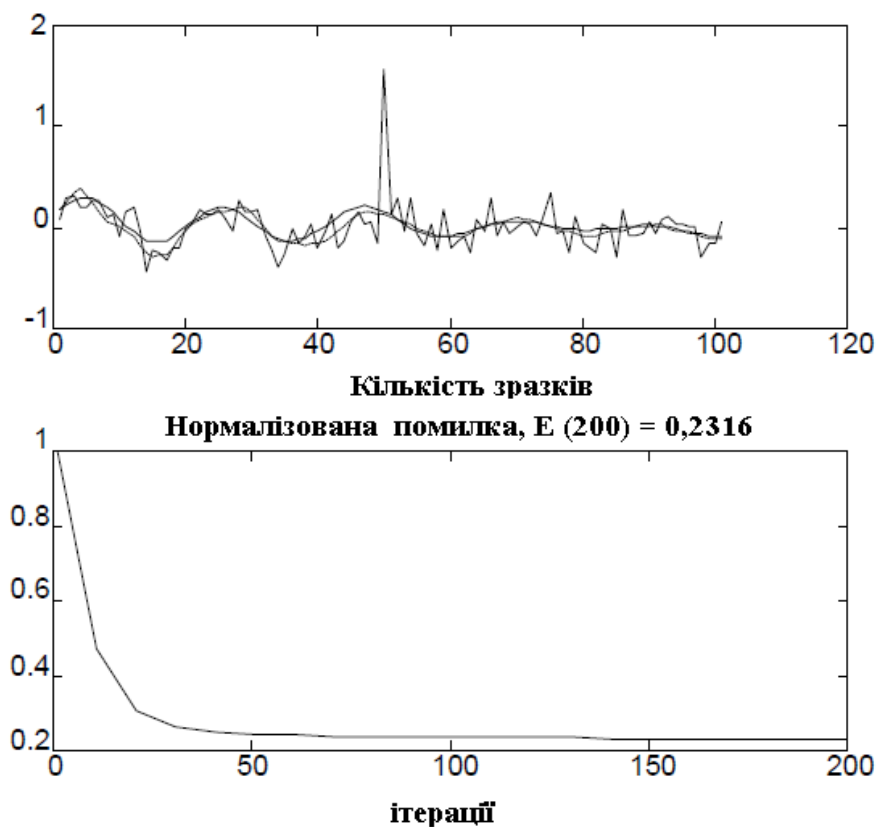


Рисунок 4.3. Моделювання мовних голосових даних, спотворених шумом та імпульсною завадою

Щоб дослідити це розглянемо моделювання голосової жіночої фонемі літери «а» із гаусовим шумом з відношенням сигнал/шум в 10 дБ (рівень шуму $\sigma^2 = 0.1$) та імпульсною завадою при $t = 50$. В вейвнеті були використані 6 вейвлетів Морле без блоку НІХ. Усі параметри векторів w_k та a_k були встановлені на 0 і 8, відповідно. Зсуви b_k кожного вейвлета були встановлені на 5, 25, 46, 70, 90 і 100, відповідно. Навчальний коефіцієнт μ для всіх параметрів був прийнятий 0.005.

Результати моделювання наведені на рис. 4.3. Тут оригінал сигналу представлений переривчастою лінією, забруднений сигнал – тонкою лінією і синтезований мовний сигнал з використанням шести вейвлет Морле – жирною лінією.

Нормована похибка відносно вихідного чистого сигналу склала величину 0.2316 на двохсотій ітерації. Тобто вейвнет без блоку НІХ достатньо добре придушує завади різних типів в мовних даних.

4.1.3. Апроксимація складних функцій

Розглянемо навчання та порівняємо ефективність моделювання складних нелінійних функцій за допомогою НМПП із алгоритмом зворотного поширення похибки та вейнету без блоку НІХ.

Хай складна кусково-лінійна функція, яку треба апроксимувати, має вигляд системи рівнянь:

$$f(x) = \begin{cases} -2.186x & -10 \leq x < -2; \\ 4.246x & -2 \leq x < 0; \\ 10e^{-0.05x-0.5} \cdot \sin((0.03x + 0.7)x) & 0 \leq x \leq 10. \end{cases} \quad (4.1)$$

НМПП мала 2 прихованих шари із функціями активації у вигляді гіперболічного тангенсу та 1 вихідний шар з лінійною функцією активації. Вейвнет мав в прихованому шарі 40 вейвлетів Морле з початковими параметрами навчання: 0.01, 0.001 і 0.001 для адаптації вагових функцій, масштабування і зсуву, відповідно.

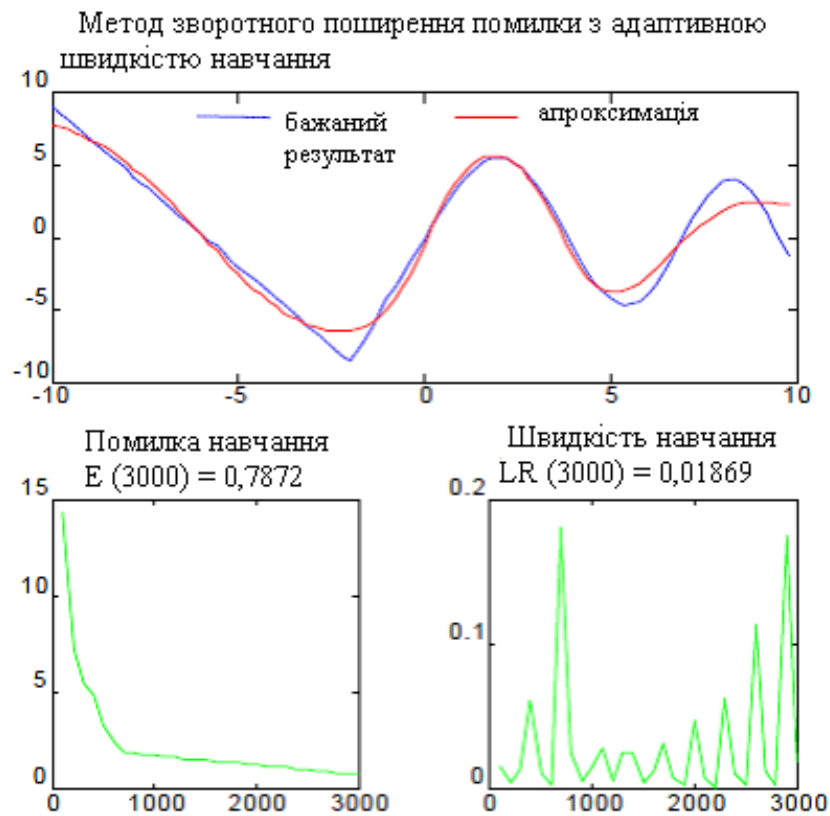


Рисунок 4.4. Апроксимація функції за допомогою НМПП

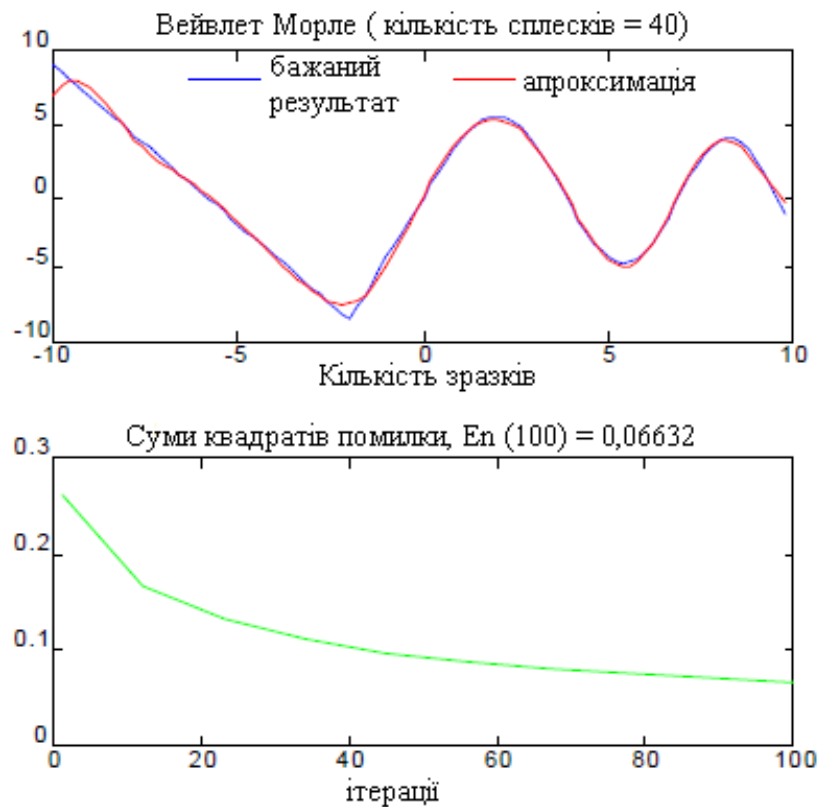


Рисунок 4.5. Апроксимація функції за допомогою вейвнету без блока НІХ

Результати навчання та моделювання НМПП та вейвнету наведені на рис. 4.4 і 4.5, відповідно.

Із графіків видно, що сума квадратів похибки для НМПП склала на трьохтисячній ітерації величину 0.7872, а для вейвнету – 0.0663 всього лиш на сотій ітерації. Тобто вейвнет при апроксимації складних функцій має суттєво кращу швидкість збіжності та саму збіжність.

4.1.4. Ідентифікація нелінійного динамічного процесу за наявності нормально розподіленого шуму

Розглянемо ідентифікацію складного динамічного процесу (системи) із шумом за допомогою вейвнету із блоком НІХ. При цьому процес описується рівнянням:

$$y(n) = \frac{y(n-1)y(n-2)}{1 + y^2(n-1) + y^2(n-2)} + u(n), \quad (4.2)$$

де $u(n) = 0,8 \sin \sin\left(\frac{2\pi n}{50}\right)$, а випадковий шум з нормальним розподілом $\sigma^2 = 0.1$ додається до входу системи.

Були використані 12 вейвлетів RASP1 з коефіцієнтами НІХ $M = N = 1$. Коефіцієнти для кожного параметра w_k , a_k , b_k , c_k , і d_k , мали значення: 0.01, 0.05, 0.05, 0.02 та 0.02, відповідно. Всі початкові ваги були встановлені в 0, масштабування було ініціалізовано до 10 і коефіцієнти НІХ прямого та зворотного зв'язку спочатку налаштовані на 0.1.

Першим, для можливості подальшого порівняння, було змодельовано ідентифікацію процесу без шуму, результати чого показано на рис. 4.6. При цьому, нормована похибка склала 0.02352 на 50-й ітерації.

Для оцінки завадостійкості ідентифікації за допомогою вейвнету з блоком НІХ у виразі (4.2) додано шум з нормальним розподілом $\sigma^2 = 0.1$. Всі мережеві параметри були прийняті такими ж, як і вище.

Результати моделювання представлені на рис. 4.7.

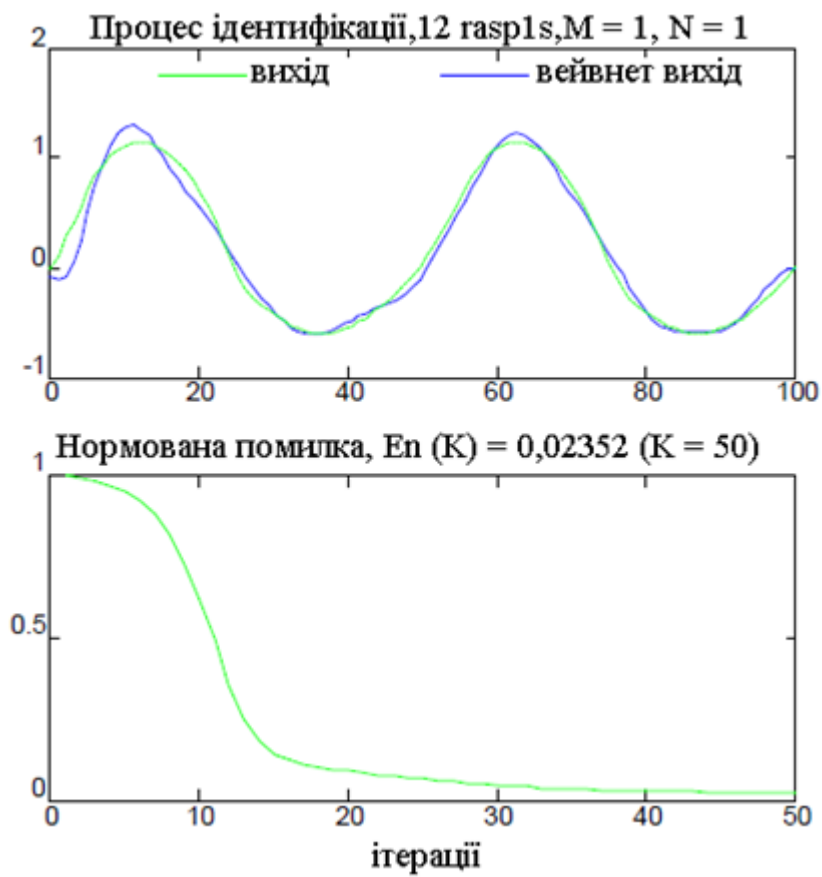


Рисунок 4.6. Ідентифікація нелінійної динамічної системи без шуму

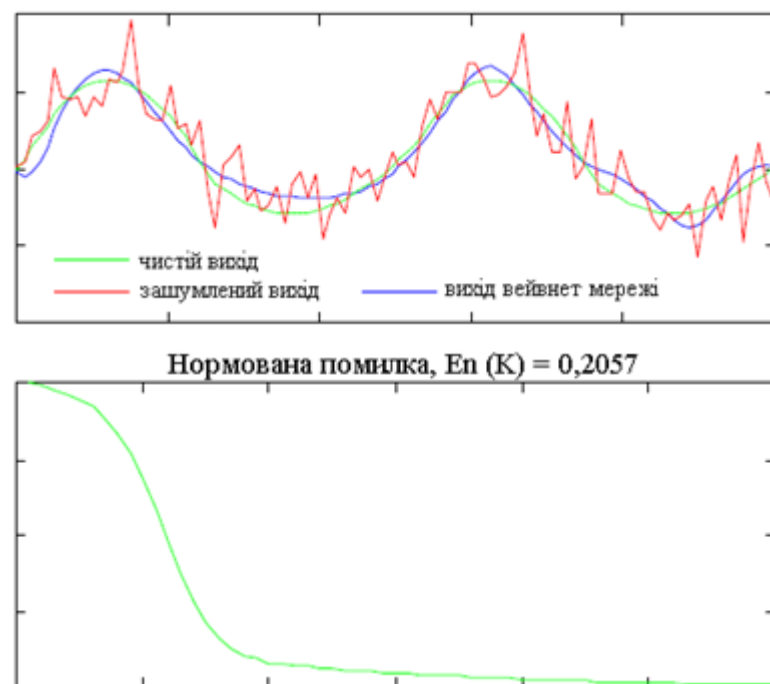


Рисунок 4.7. Ідентифікація нелінійної динамічної системи з шумом

Провівши моделювання, можна зробити висновок про те, що ідентифікація без шуму в системі є більш задовільною, ніж у випадку ідентифікації із шумом, де нормована похибка між чистим виходом системи і відновленим вихідним відгуком мережі склала 0.05744. При цьому, нормована похибка, по відношенню до похибки, між зашумленим сигналом і виходом вейвнету, склала 0.2057 (див. рис. 4.7). Це моделювання підкреслює високу завадостійкість і надійність мережі на основі вейвнету з блоком НІХ.

4.2. Придушення шумів у сигналах мобільного зв'язку з використанням нейронечіткої ідентифікації тракту передачі

Основною проблемою при передачі сигналів мобільному (рухомого) зв'язку є виділення на тлі завад корисного сигналу при його багатопроменевому розповсюдженні. Для цього застосовують різні методи фільтрації сигналів і ідентифікації трактів передачі.

Для вирішення завдань ідентифікації та фільтрації широко використовуються методи систем штучного інтелекту, зокрема, адаптивні нейронні системи нечіткого висновку (Anfis АФА). Вони є нейронною мережею з чіткими сигналами, вагами і функцією активації, але з їх об'єднанням за допомогою операцій нечіткої логіки.

Дослідимо ефективність використання такої системи для оптимального прийому сигналів з нелінійними завадами відповідно до схеми на рис. 4.8.

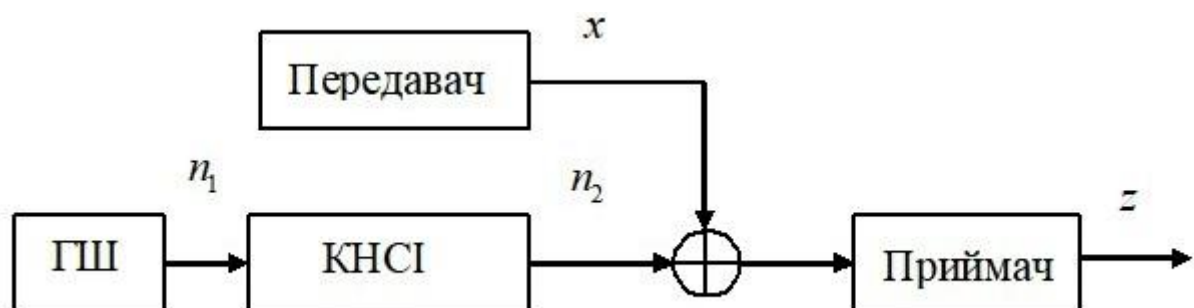


Рисунок 4.8. Схема передачі радіосигналу мобільного зв'язку

Передавач передає інформаційний сигнал x на фоні відомого шуму n_1 , що генерується генератором шуму (ГШ). Відомий шум n_1 зазнає нелінійних спотворень та інтерференції в каналі КНСІ, що має вихідний сигнал n_2 . В результаті адитивний сигнал $z = x + n_2$ вимірюється приймачем.

Тобто, відомі сигнал шуму n_1 і виміряний сигнал z , за якими необхідно відновити інформаційний сигнал x .

Рішення задачі полягає в оцінці шуму $\hat{n}_2$, що можна отримати шляхом ідентифікації зв'язку між n_1 і n_2 з використанням системи нечіткого висновку Anfis. Тоді оцінка сигналу буде дорівнювати $\hat{x} = z - \hat{n}_2$.

Дослідження прийому сигналів з ідентифікацією тракту передачі проводилися при варіації наступних характеристик:

- а) типу інформаційного сигналу x ;
- б) виду моделі тракту розповсюдження;
- в) параметрів системи нечіткого висновку.

Якість прийому оцінювалась за критерієм мінімуму відносної похибки між переданим x і прийнятим $\hat{x}$ сигналами:

$$\varepsilon = \frac{\|x - \hat{x}\|}{\|x\|}. \quad (4.3)$$

Результати моделювання системи прийому з нейронечіткою ідентифікацією для різних типів передаваних сигналів наведені в табл. 4. 2.

Використання системи прийому з нейронечіткою ідентифікацією тракту розповсюдження ефективно для сигналів довільної складності (похибка складає $\varepsilon = 4...5\%$). Очевидно, що додаткова фільтрація сигналів, що приймаються (наприклад, методом ковзного згладжування), зменшить величину похибки.

Результати моделювання системи для різних моделей тракту розповсюдження спотворюючого шуму наведені в табл. 4. 3.

В моделях таблиці змінна k – це такт часу. Аналіз цієї таблиці показує, що використання систем з нейронечіткою ідентифікацією більш ефективно при нелінійних трактах розповсюдження сигналів.

Таблиця 4.2. Похибки прийому для різних типів сигналу

Тип сигналу x	Похибка ε , %	Сигнал/шум, дБ
1. Гаусовий шум	4.05	27.6
2. Лінійна функція $x = 0.5 \cdot (time + 0.4)$	4.47	26.5
3. Показова функція $x = 2^{time+0.9}$	5.18	25.2
4. Сигнал з частотою в функції часу $x = \sin(40 / (time + 0.01))$	3.80	28.1

Таблиця 4.3. Похибки прийому для різних моделей тракту розповсюдження

Рівняння моделі тракту розповсюдження	Похибка ε , %
1. $n_2[k] = n_1[k] + 2 \cdot n_1[k - 1]$	4.70
2. $n_2[k] = 2(n_1[k] + n_1[k - 1])$	3.31
3. $n_2[k] = \sin(n_1[k]) \cdot 2 \cdot n_1[k - 1]$	3.98
4. $n_2[k] = \sin(n_1[k]) \cdot \cos(2 \cdot n_1[k - 1])$	3.42

Дослідження впливу параметрів системи нечіткого висновку Anfis на прийом сигналів з нелінійними завадами та спотвореннями показали, що оптимальними параметрами (у сенсі мінімуму похибки ε) є дзвіноподібна та трикутні функції належності при 2...3 правилах розкладання по кожному входу з розміром кроку навчання 0,1...0,3.

В цілому проведені дослідження дозволили довести ефективність використання систем мобільного (рухомого) зв'язку з нейронечіткою ідентифікацією тракту розповсюдження і визначити її оптимальні параметри.

4.3. Аналіз і моделювання мережевого трафіку

Використовувались експериментальні сигнали трафіку автоматичних телефонних станцій (АТС) на базі систем комутацій: EWSD ДНІПРО (сигнал 1) і F-1500 (сигнал 2) та трафіку, що передавався через Інтернет (сигнал 3).

Система EWSD ДНІПРО надає абонентам розширений набір послуг зв'язку: інтегральне обслуговування, обмін даними і обмін голосовою інформацією, в той час як комутаційна система F-1500 – тільки останню. Також АТС на базі системи комутації EWSD є зоною, а АТС на базі системи комутації F-1500 – міською. Ці особливості побудови та використання систем обумовлюють відмінності їх трафіків. Час агрегації (дискретизації) для сигналів 1 і 2 складає 1 год. Дані реалізації трафіку мережі Інтернет являють собою залежність розміру Ethernet кадрів в байтах від часу. Для того, щоб привести початкові дані до еквідистантної шкали по часовій осі була проведена процедура агрегації з кроком в 5 с.

При моделюванні процесів ідентифікації та прогнозування сигналів 1-3 використовувались як стандартні програми, так і програми, розроблені у середовищі Matlab. Результати наведені на рис. 4.9-4.14 та в таблицях 4.4 і 4.5.

За видом АКФ (рис. 4.9,а) і спектру (рис. 4.11,а) експериментального сигналу 1 можна сказати про нерегулярність породжуючого процесу. Це випливає з того, що його спектр має вид спектру флікер-шуму, тобто спектральна потужність сигналу 1 приблизно зворотнопропорційна частоті. При цьому з вигляду його АКФ не можна з впевненістю стверджувати, що породжуючому процесу притаманна властивість самоподоби (апроксимації АКФ у вигляді повільно (ПУЗ) та швидко (ШУЗ) убуючих залежностей наведені на рис. 4.9,а).

Значення показника Херста (табл. 4.4) сигналу 1 мають в залежності від методів, що використовувались, значення в діапазоні $[0,814; 0,923]$. Тобто це персистентні ряди: якщо вони зростають (убуюють) в попередній період, то, ймовірно, що вони зберігатимуть цю тенденцію певний час в майбутньому. Для таких сигналів події недавнього минулого впливають на процес набагато сильніше, ніж події віддаленого минулого. Крім того, їх вейвлет-перетворення мають самоподібний (фрактальний) характер (гілляста (деревоподібна)

структура зберігається на різних масштабах).

Вигляд графіку дисперсій процесу (рис. 4.13,*а*), а саме їх повільне убуття, також підтверджує наявність даної властивості у досліджуваного процесу, що породжує сигнал 1. Крім того, процес є нелінійним – на це вказують результати проведеного BDS-тесту для залишків лінійної моделі (див. табл. 4.4).

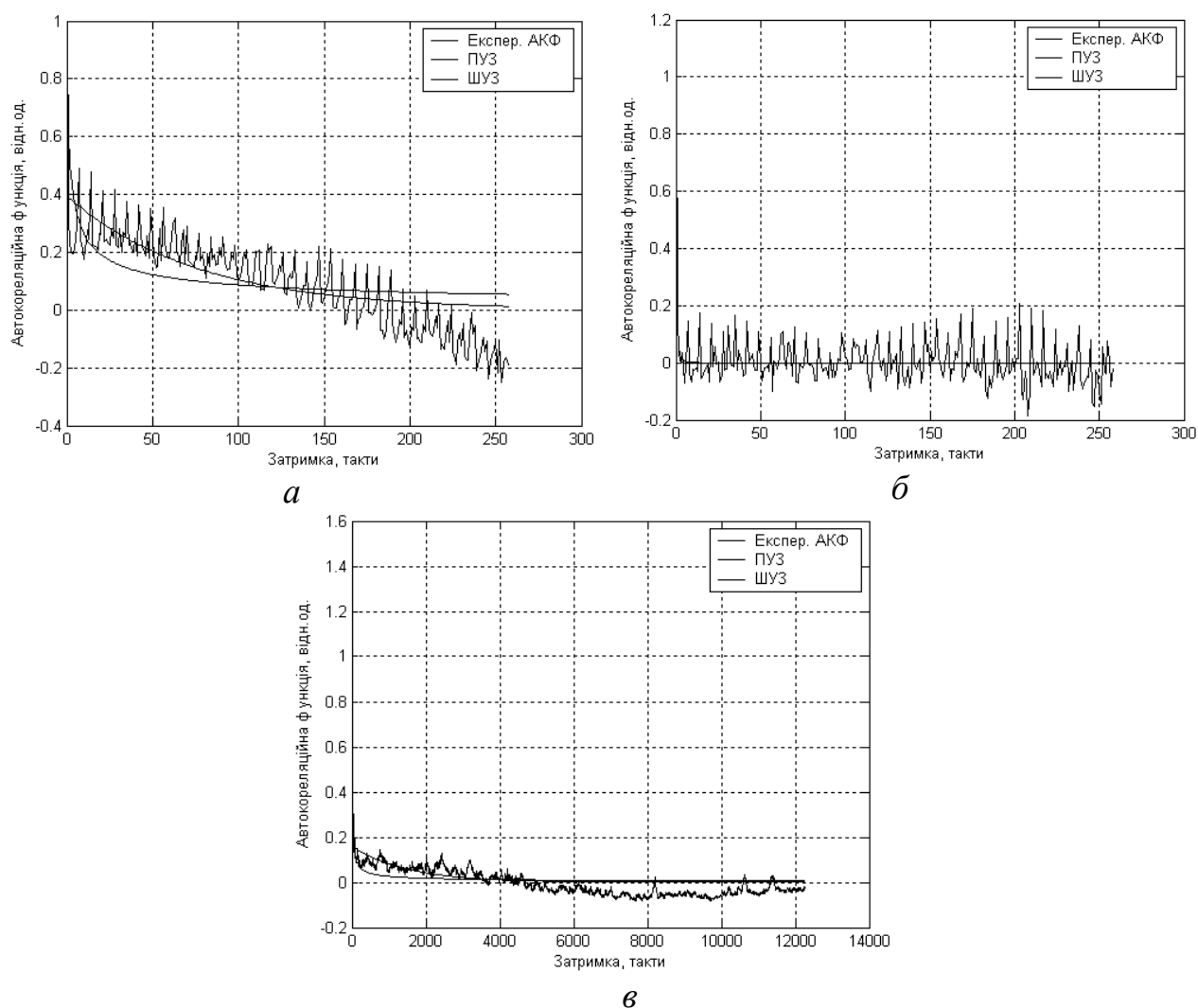


Рисунок 4.9. АКФ експериментальних сигналів 1 (*а*), 2 (*б*) та 3 (*в*)

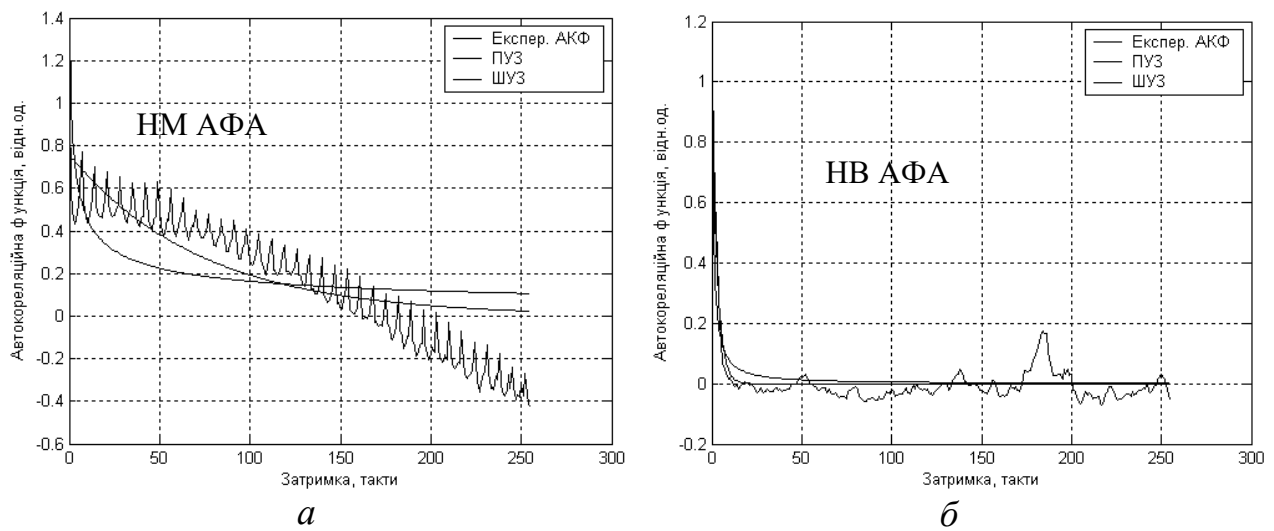


Рисунок 4.10. АКФ модельних сигналів 1 (а) і 3 (б)

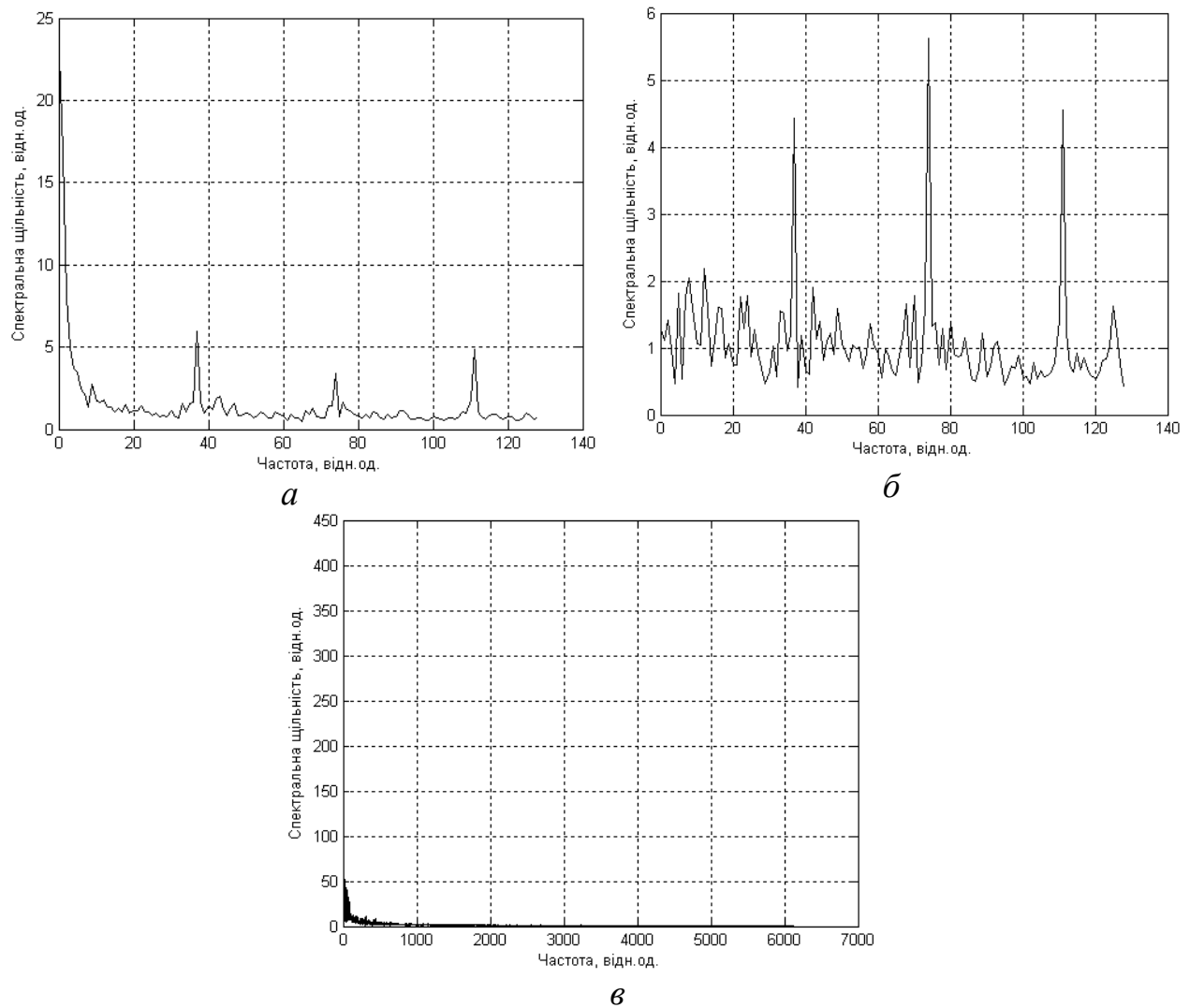
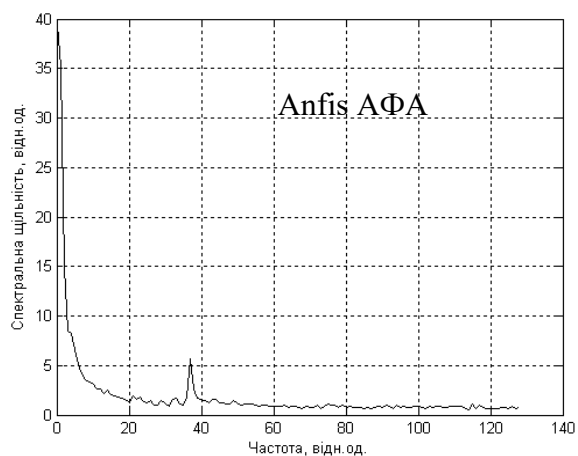
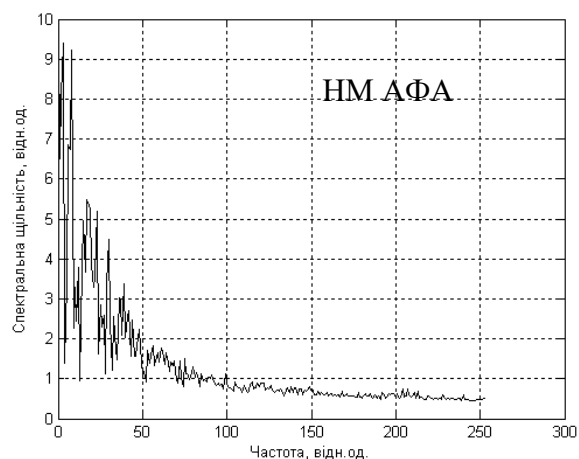


Рисунок 4.11. Спектри експериментальних сигналів 1 (а), 2 (б) та 3 (в)

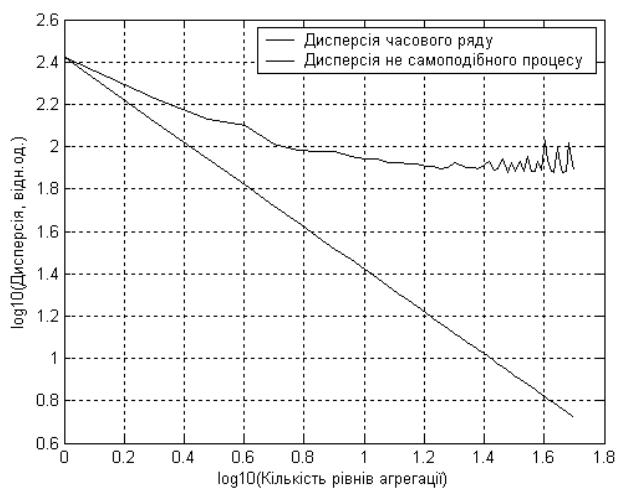


a

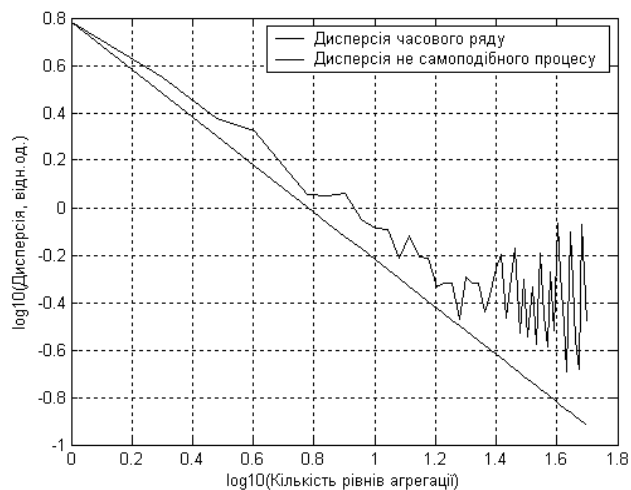


б

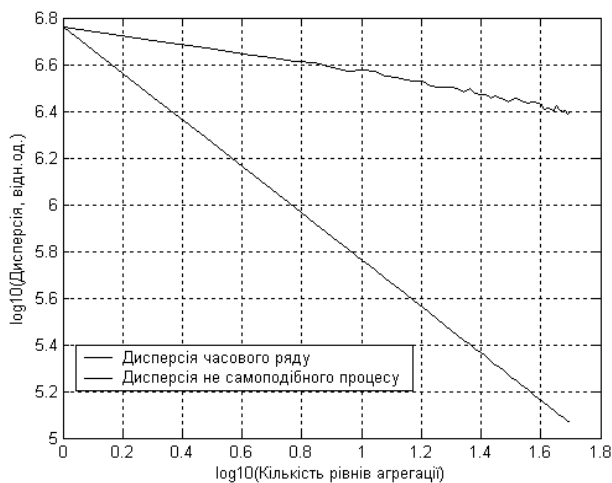
Рисунок 4.12. Спектри модельних сигналів 1 (*a*) і 3 (*б*)



a



б



в

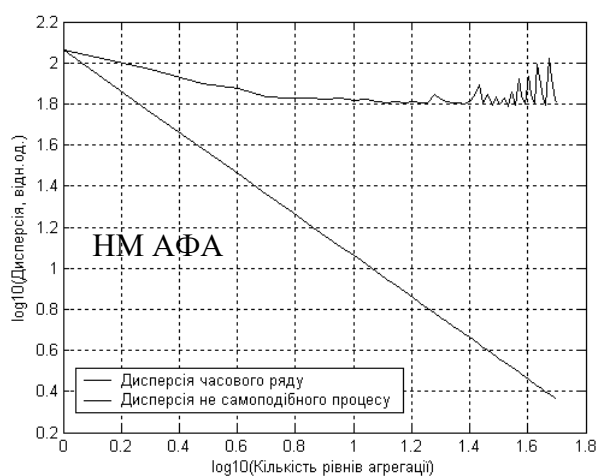
Рисунок 4.13. Дисперсії експериментальних сигналів 1 (*a*), 2 (*б*) та 3 (*в*)

Побудова гістограм експериментального розподілу і визначення їх адекватності теоретичним розподілам за допомогою критеріїв згоди Колмогорова і Пірсона дали змогу встановити, що сигналу 1 відповідає логнормальний закон розподілу.

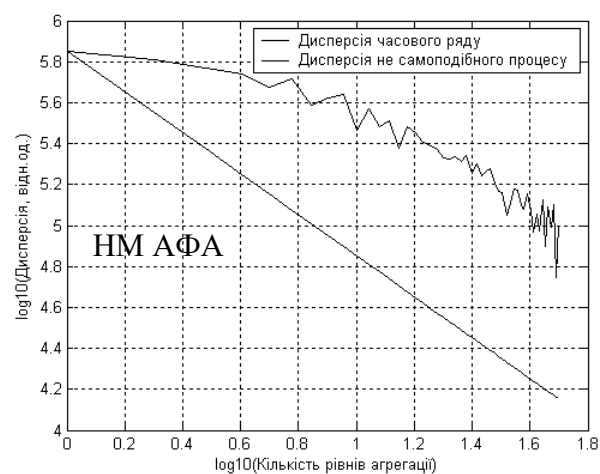
Таблиця 4.4. Характеристики експериментальних сигналів

Показник	Сигнал 1	Сигнал 2	Сигнал 3
АКФ	ШСЗ		ПСЗ
Спектр	Ступеневий закон із сплесками	Енергія представлена в області низьких та високих частот	Ступеневий закон
Вейвлет-перетворення	Деревоподібна структура скейлетона		
Дисперсія	Поволі убуваюча	Зменшується зі зростанням обсягу вибірки	Поволі убуваюча
Закон розподілу за критерієм: - Колмогорова - Пірсона	Логнормальний Логнормальний	Логнормальний Логнормальний	Парето Парето
Показник «важкості хвоста»	~5,4	~4,5	~1,4
BDS-тест	Хаотичний нелінійний процес	Стохастичний нелінійний процес	Хаотичний нелінійний процес

Показник Херста			
- періодограмний метод	0,923	0,909	0,980
- метод агрегованої дисперсії	0,904	0,668	0,888
- R/S аналіз	0,814	0,671	0,913
Кореляційна розмірність	3,584	4,326	2,956
Кореляційна ентропія	0,954	0,599	0,640
Інтервал прогнозованості	1,694	2,697	2,525
Розмірність атрактора	$4 \leq d \leq 6$	$5 \leq d \leq 8$	$2 \leq d \leq 4$



a



б

Рисунок 4.14. Дисперсії модельних сигналів 1 (*a*) і 3 (*б*)

Швидкий спад АКФ експериментального сигналу 2 (рис. 4.9,*б*) свідчить про стохастичність процесу, що його породжує. Енергія спектру сигналу (рис. 4.11,*б*) представлена як в області низьких, так і високих частот. Значення показника

Херста (табл. 4.4) змінюється в діапазоні $[0,814; 0,923]$. Графік дисперсії сигналу 2 (рис. 4.13,б) свідчить про відсутність властивості самоподоби. Підтвердженням стохастичності та нелінійності процесу є результат BDS-тесту.

Породжуючий процес експериментального сигналу 3 має всі властивості, притаманні самоподібним процесам: гіперболічне убування АКФ із зростанням часової затримки (рис. 4.9,в), ступеневий закон графіка його спектральної щільності (рис. 4.11,в) та поволі убуваючу дисперсію (рис. 4.13,в). Сигнал 3 має (табл. 4.4) розподіл з «важким хвостом» (показник «важкості хвоста», визначений за методом Хілла та шляхом побудови графіка додаткового розподілу в подвійному логарифмічному масштабі, становить $\alpha = 1,4$), гістограма експериментального розподілу згідно з критеріями згоди Колмогорова і Пірсона відповідає розподілу Парето, скейлетон вейвлет-перетворення сигналу має деревоподібну структуру, а показник Херста має діапазон значень $[0,888; 0,980]$. Процес також є нелінійним відповідно до результатів BDS-тесту.

Характеристики модельних сигналів

Таблиця 4.5

Показник	Сигнал 1			Сигнал 3	
	Anfis АФА	НМ АФА	НВ АФА	НМ АФА	НВ АФА
АКФ	ШСЗ		ПСЗ	ШСЗ	
Спектр	Ступеневий закон із сплесками			Ступеневий закон	
Вейвлет- перетворення	Деревоподібна структура скейлетона				
Дисперсія	Поволі убуваюча				
Закон розподілу за критерієм: - Колмогорова	Логнормальний	Логнормальний	Логнормальний	Парето	

- Пірсона	Логнормальний		Гамма-розподіл	Логнормальний	Експоненціальний
Показник «важкості хвоста»	~6,5	~6,1	~9,4	~1,2	~1,5
BDS-тест	Хаотичний нелінійний процес				
Показник Херста:					
- періодограмний метод	0,999	0,997	1,013	0,936	0,974
- метод агрегованої дисперсії	0,967	0,979	1,039	0,687	1,454
- R/S аналіз	0,837	0,992	0,849	0,788	0,764
Кореляційна розмірність	4,134	3,717	2,738	3,992	3,483
Кореляційна ентропія	1,029	0,807	0,788	0,826	0,643
Інтервал прогнозованості	1,551	2,003	2,046	3,063	2,987
Розмірність атрактора	$5 \leq d \leq 7$	$4 \leq d \leq 6$	$4 \leq d \leq 6$	$2 \leq d \leq 5$	$2 \leq d \leq 4$

Гіпотези про тотожність розподілу значень сигналів приймалися з вірогідністю 95 %.

Значення кореляційних ентропій і розмірностей атракторів експериментальних сигналів наведені в табл. 4.4, при цьому інтервали точної передбачуваності (глибина точного прогнозу) склали $T_{C_1} = 1,694$, $T_{C_2} = 2,697$ і $T_{C_3} = 2,525$ тактів дискретизації.

При моделюванні у якості моделей трафіку (сигналів 1 і 3)

використовувались лінійний, Anfis, НМ і НВ АФА. В результаті їх адаптації та навчання АКФ (рис. 4.10), спектри (рис. 4.12) та дисперсії (рис. 4.14) модельних сигналів якісно відповідають графікам експериментальних сигналів (рис. 4.9, 4.11, 4.13).

Крім того, якісні та числові значення характеристик модельних сигналів в цілому відповідають характеристикам експериментальних сигналів (див. табл. 4.4 і 4.5).

В результаті моделювання встановлено, що відносна середньоквадратична похибка апроксимації експериментальних сигналів за допомогою АФА не перевищує 0,046.

Статистична перевірка за непараметричним критерієм знаків дозволила встановити адекватність експериментальних і модельних сигналів з рівнем значущості 0,01.

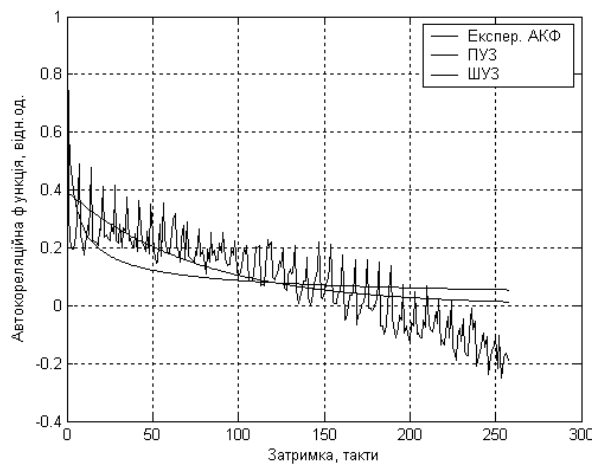
4.4. Мультифрактальний аналіз та моделювання мережевого трафіку

Для аналізу та моделювання використовувались експериментальні сигнали трафіку автоматичної телефонної станції на базі системи комутацій: EWSD (сигнал 1) та трафіку, що передавався через мережу Інтернет (сигнал 3).

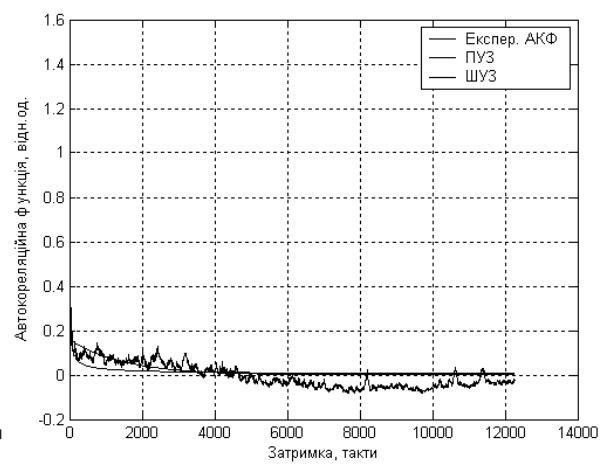
Для моделювання мережевого трафіку використовувались мультифрактальні вейвлет моделі (МВМ) двох типів: МВМ з бета-розподілом та гібридна МВМ, в якій бета розподіл використовується на грубих масштабах, а розподіл точечних мас – на точних масштабах.

Результати оцінки характеристик та моделювання сигналів трафіку наведені на рис. 4.15-4.17 та в табл. 4.6.

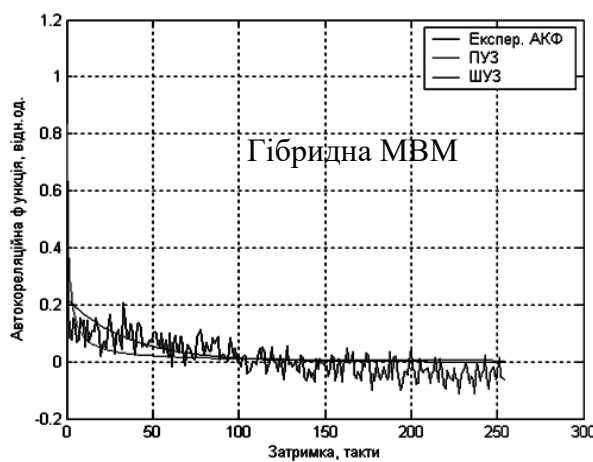
За видом АКФ (рис. 4.15, *a* і *б*), спектрів (рис. 4.16, *a* і *б*) та дисперсій (рис. 4.17, *a* і *б*) експериментальних сигналів можна сказати про самоподобу їх породжуючих процесів, оскільки АКФ мають довготривалі кореляції, енергія спектрів спадає по ступеневому закону, а дисперсії поволі убувають. (На графіках АКФ також наведені їх апроксимації у вигляді повільно (ПУЗ) та швидко (ШУЗ) убуваючих залежностей).



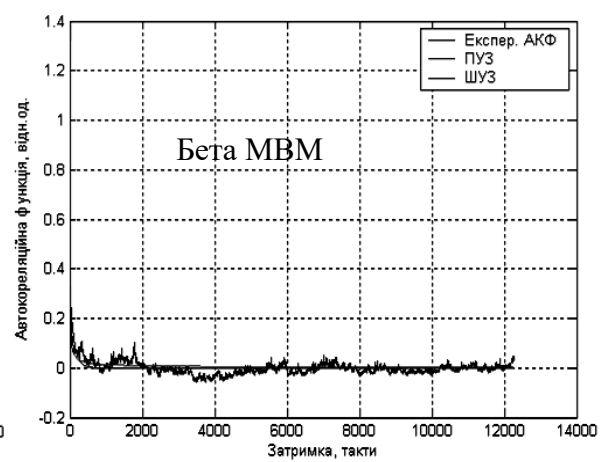
a



б



в



г

Рисунок 4.15. АКФ експериментальних сигналів 1 (*a*) і 3 (*б*)
та модельних сигналів 1 (*в*) і 3 (*г*)

Моделювання сигналів трафіку мультифрактальними вейвлет моделями підтверджує їх якісні оцінки (рис. 4.15, *в* і *г*; рис. 4.16, *в* і *г*; рис. 4.17, *в* і *г*). Крім того, статистична перевірка за непараметричним критерієм знаків дозволила встановити адекватність експериментальних і модельних сигналів з рівнем значущості 0,01.

Розрахунки дозволили визначити (табл. 4.6), що експериментальні і модельні сигнали 1 і 3 підпорядковуються субекспоненційним законам розподілу з показниками «важкості хвоста» [5,3; 5,7] і [1,2; 1,5] відповідно (гіпотези про тотожність розподілу значень сигналів приймалися з вірогідністю 95 %).

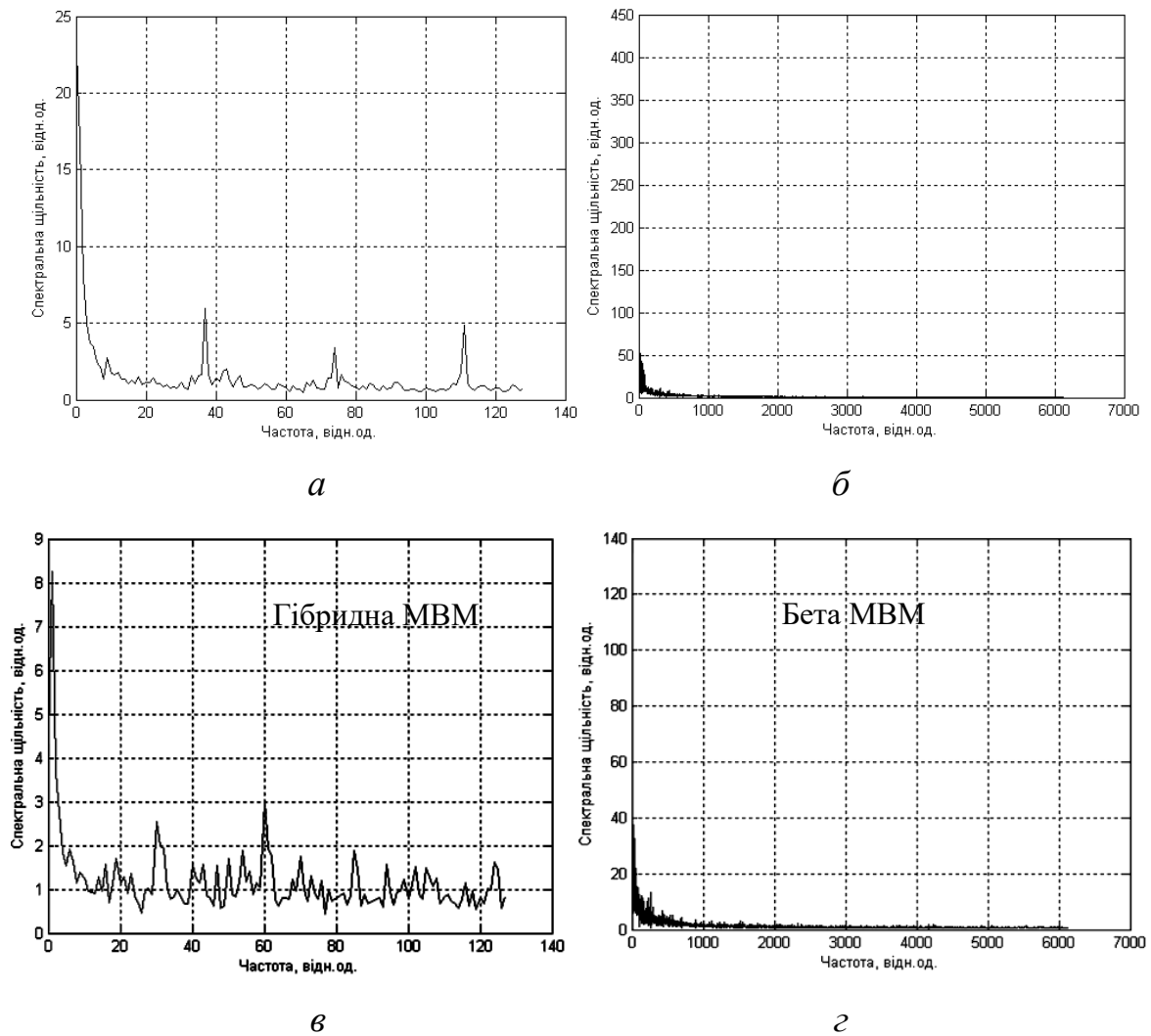
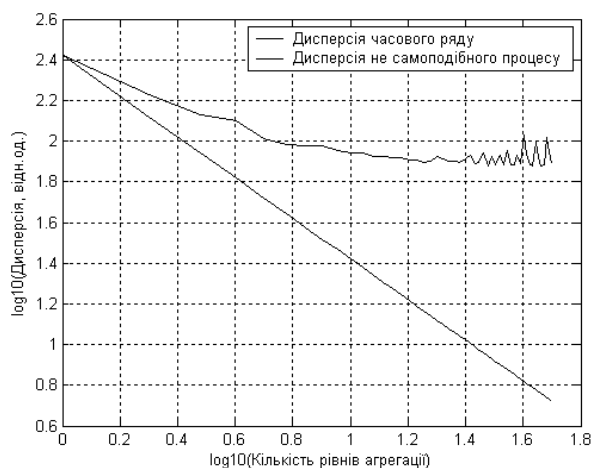


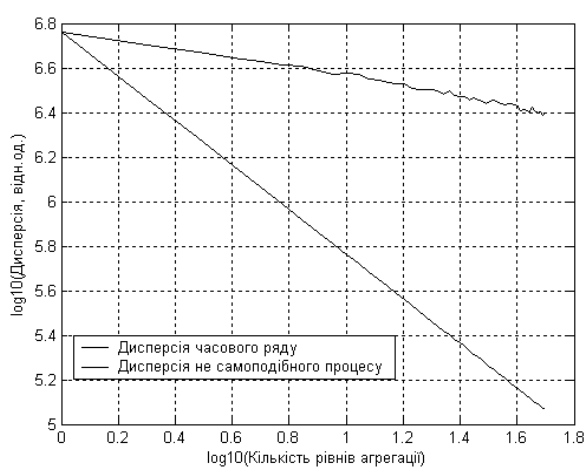
Рисунок 4.16. Спектри експериментальних сигналів 1 (*a*) і 3 (*б*)
та модельних сигналів 1 (*в*) і 3 (*г*)

Числові значення характеристик модельних сигналів в цілому відповідають характеристикам експериментальних сигналів, при цьому бета MBM мають меншу похибку визначення характеристик, ніж гібридні MBM.

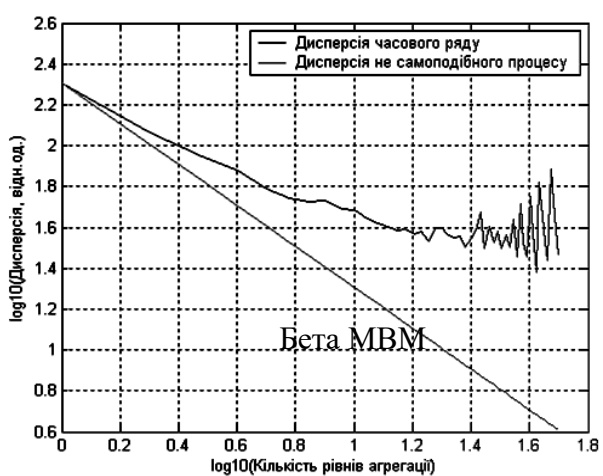
Проведений аналіз детрендових флуктуацій (АДФ) дозволив отримати наступні результати. Поведінка сигналу 1 (рис. 4.18, *a*) на інтервалі від 10 до 86 тактів некорельована, оскільки показник скейлінга $\alpha_1 = 0,5$.



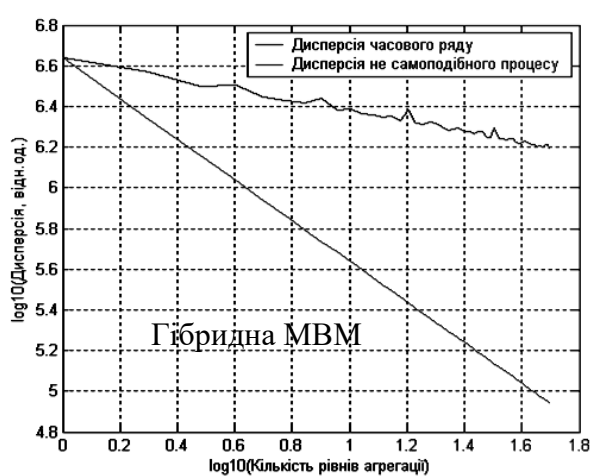
а



б



в



г

Рис. 4.17. Дисперсії експериментальних сигналів 1 (а) і 3 (б)
та модельних сигналів 1 (в) і 2 (г)

Таблиця 4.6. Характеристики сигналів

Показник	Сигнал 1			Сигнал 3		
	Експериментальний	Бета MBM	Гібридна MBM	Експериментальний	Бета MBM	Гібридна MBM
АКФ	Поволі убуваюча залежність					
Спектр	Ступеневий закон із сплесками			Ступеневий закон		

Вейвлет-перетворення	Деревоподібна структура скейлетона					
Дисперсія	Поволі убуваюча					
Закон розподілу за критерієм: - Колмогорова - Пірсона	Логнормальний Логнормальний		Гамма-розподіл Нормальний	Парето Парето		Парето Логнормальний
Показник «важкості хвоста»	~5,4	~5,7	~5,3	~1,4	~1,5	~1,2
BDS-тест	Хаотичний нелінійний процес		Стохастичний нелінійний процес	Хаотичний нелінійний процес		
Показник Херста	0,923	0,891	0,936	0,980	0,940	0,884
- періодограмний метод						
- метод агрегованої дисперсії						
- R/S аналіз	0,814	0,910	0,776	0,913	0858	0,877
Кореляційна розмірність	3,584	3,710	3,898	2,956	2,978	2,864
Кореляційна ентропія	0,954	0,898	0,827	0,640	0,742	0,857
Інтервал прогнозування	1,694	1,800	1,954	2,525	2,178	1,886

Розмірність атрактора	$4 \leq d \leq 6$	$4 \leq d \leq 6$	$4 \leq d \leq 6$	$2 \leq d \leq 4$	$2 \leq d \leq 5$	$2 \leq d \leq 5$
--------------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

Проміжку часу від 87 до 129 тактів відповідає $\alpha_2 = 1,79 > 1$, що свідчить про кореляції, проте ступеневу залежність вони не мають. На інтервалі більше 129 тактів показник приймає значення $0,5 < \alpha = 0,92 < 1$, що властиве для персистентних процесів, які мають тренд.

Для сигналу 3 значення $\alpha = 0.98$ близьке до 1 (рис. 4.18, б), що властиво для 1/f-шума (флікер-шума), для якого енергія спектра зворотно пропорційна частоті.

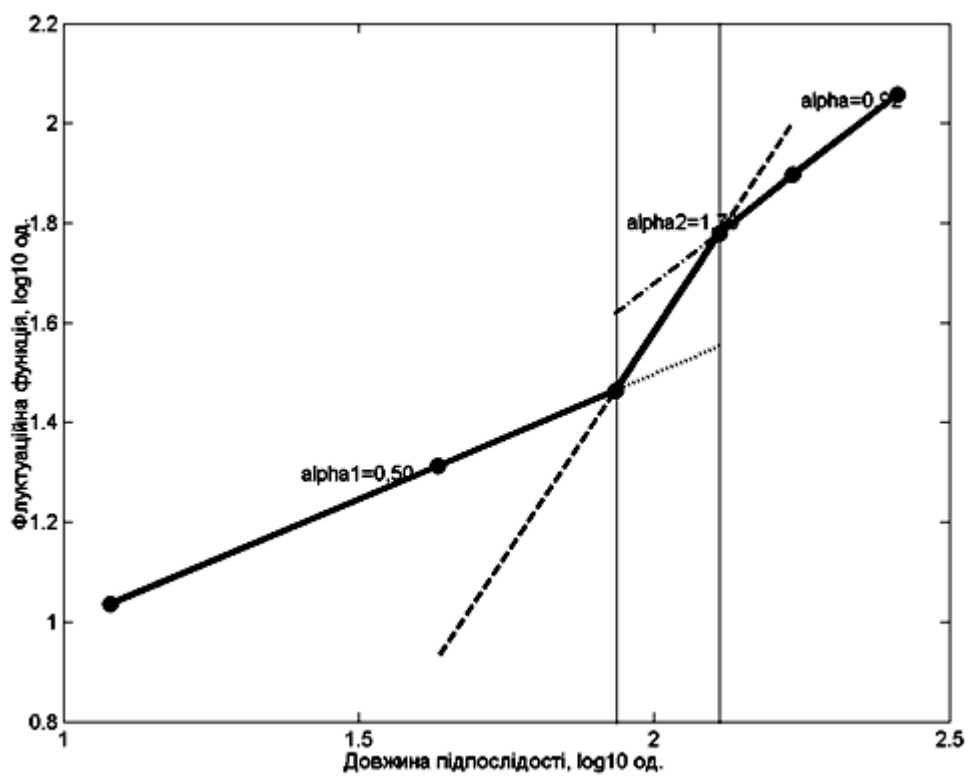
Слід відмітити, що для експериментальних сигналів отримані значення показника скейлінга α співпадають зі значеннями показника Херста (табл. 4.6), отриманими періодограмним методом.

Результати мультифрактального АДФ наведені на рис. 4.19. При обмеженні значень порядку $-5 \leq q \leq 5$ на масштабах $s < 2^5 = 32$ трафіки (сигнали 1 та 3) можна розглядати як монофрактальні процеси.

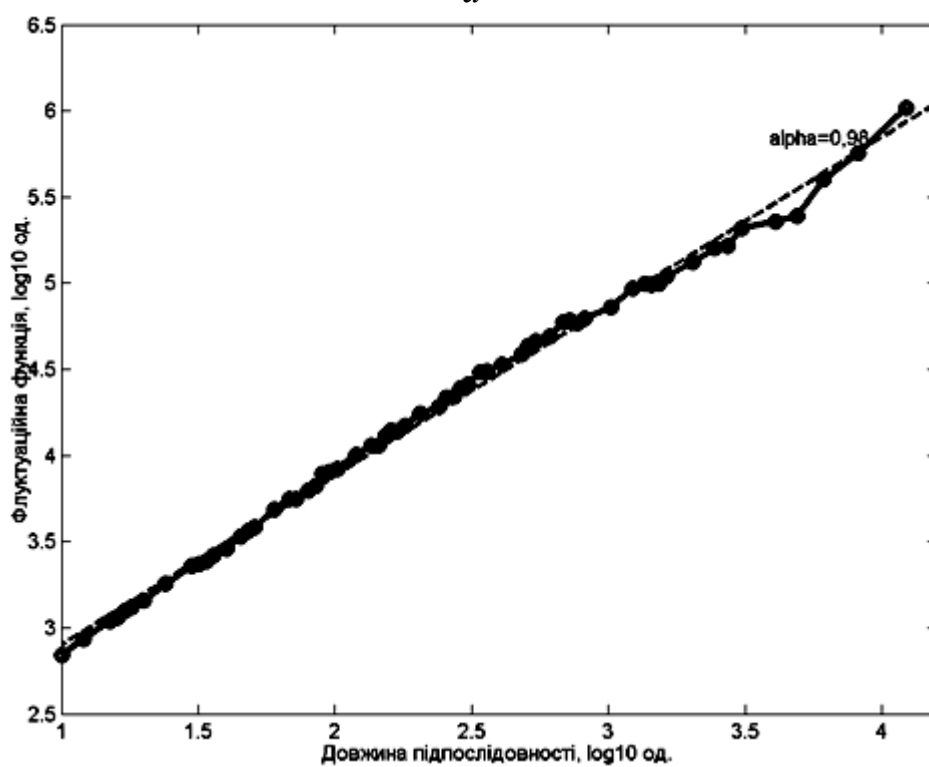
В результаті проведеного дослідження сигналів з використанням методу МВП були отримані наступні результати. Графіки часткових функцій $Z(q,s)$ (рис. 4.20) показують, що при варіації значення їх порядку $-5 < q < 5$ на масштабі $s \leq 2^3 = 8$ експериментальні сигнали виявляють монофрактальні властивості.

Графіки скейлінгових експонент $\tau(q)$, показників Гельдера (узагальнених показників Херста) $h(q)$ та спектрів сингулярності $D(h)$ експериментальних сигналів наведені на рис. 4.21-4.23, відповідно.

З графіків випливає, що експериментальний сигнал 1 при варіації значення порядку $-5 \leq q \leq 5$ породжується монофрактальним процесом: скейлінгова експонента є прямою лінією, показник Гельдера практично не змінюється, а спектр сингулярностей є компактною множиною точок.

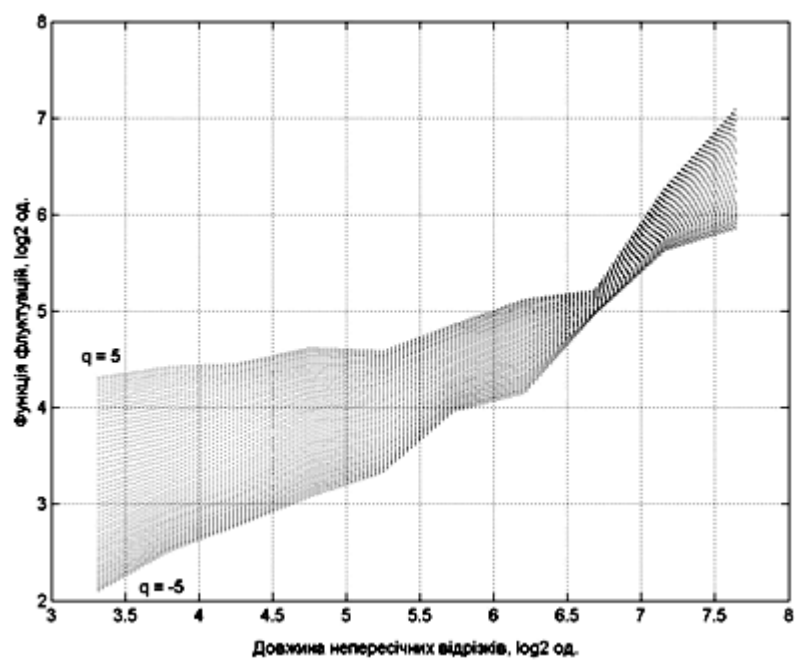


a

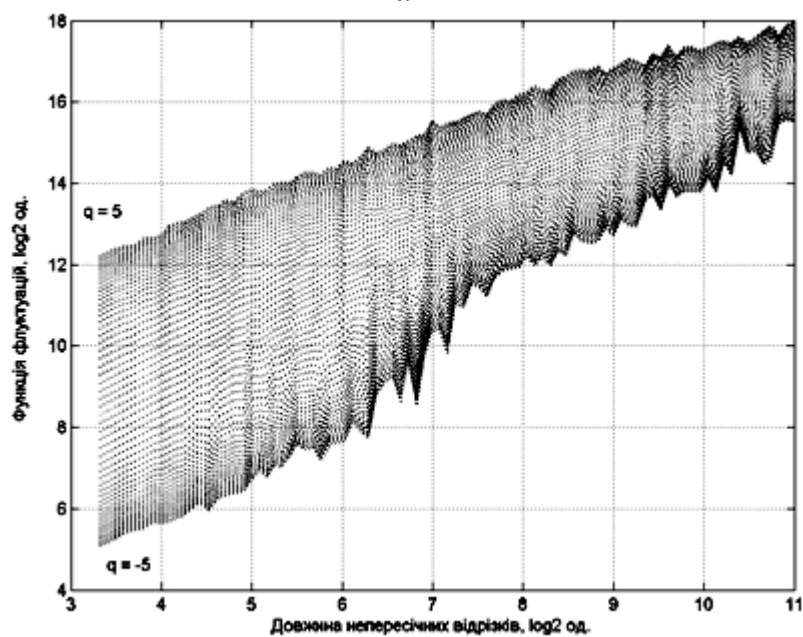


b

Рисунок 4.18. Флуктуаційні функції експериментальних сигналів 1 (*a*) і 3 (*b*) за АДФ

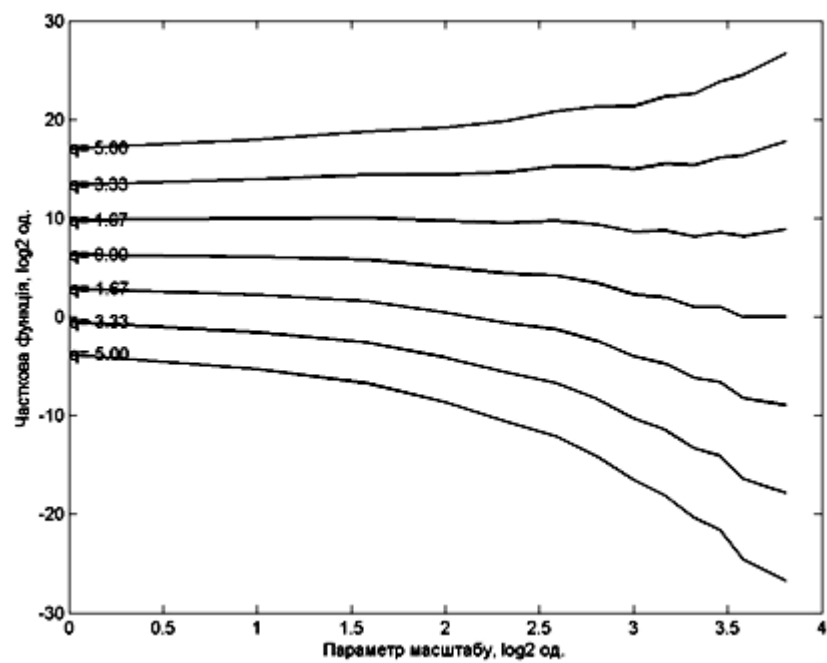


a

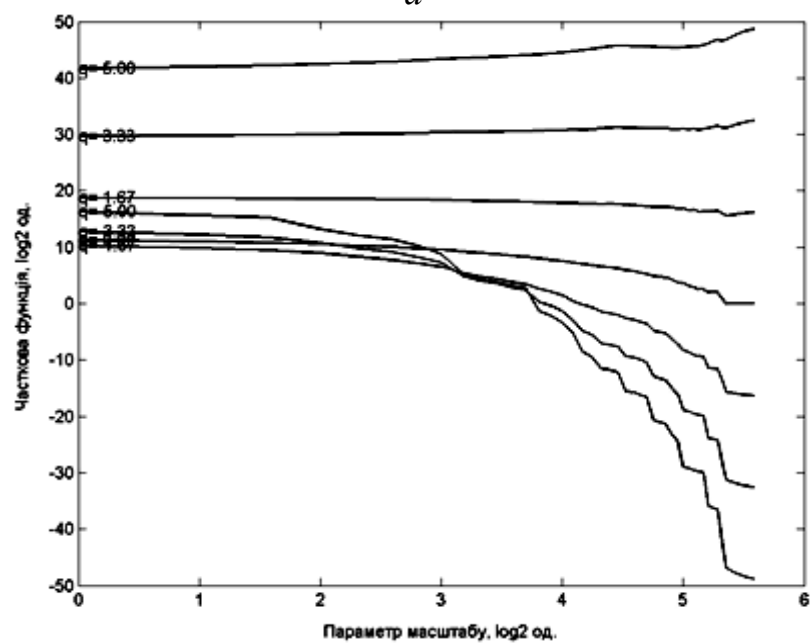


б

Рисунок 4.19. Флуктуаційні функції експериментальних сигналів 1 (*a*) і 3 (*б*) за мультифрактальним АДФ



a



б

Рисунок 4.20. Часткові функції експериментальних сигналів 1 (*a*) і 3 (*б*)

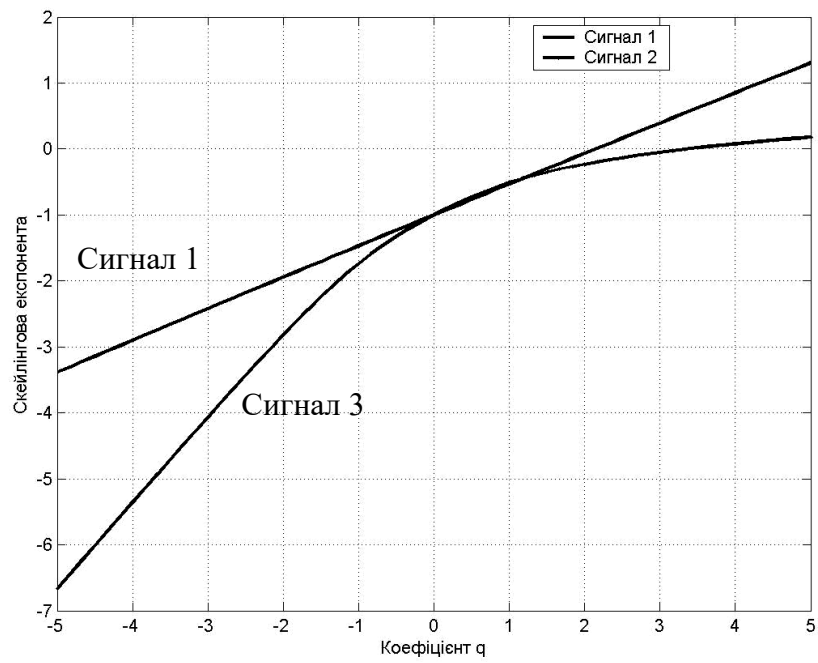


Рисунок 4.21. Скейлінгові експоненти експериментальних сигналів

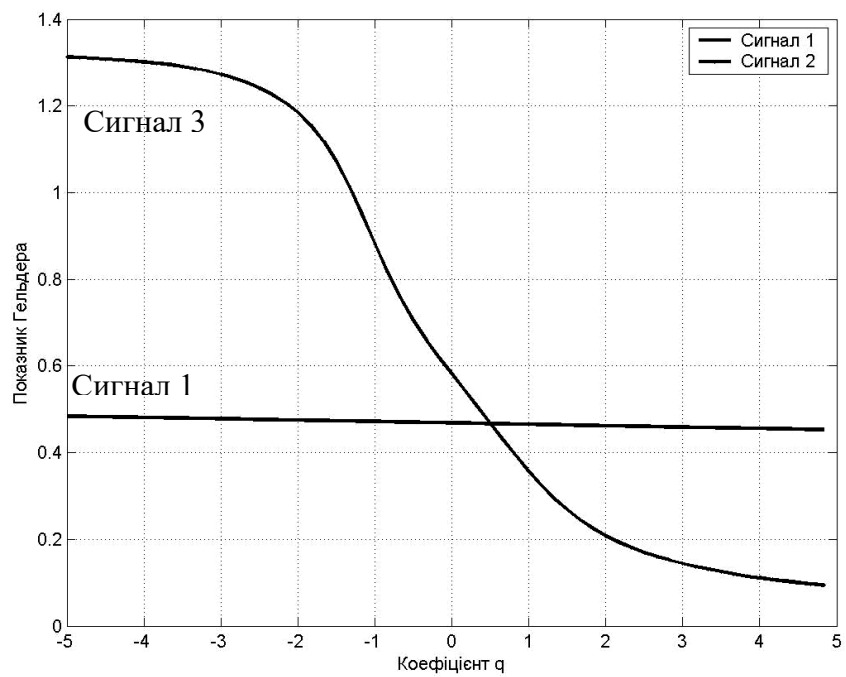


Рисунок 4.22. Показники Гельдера експериментальних сигналів

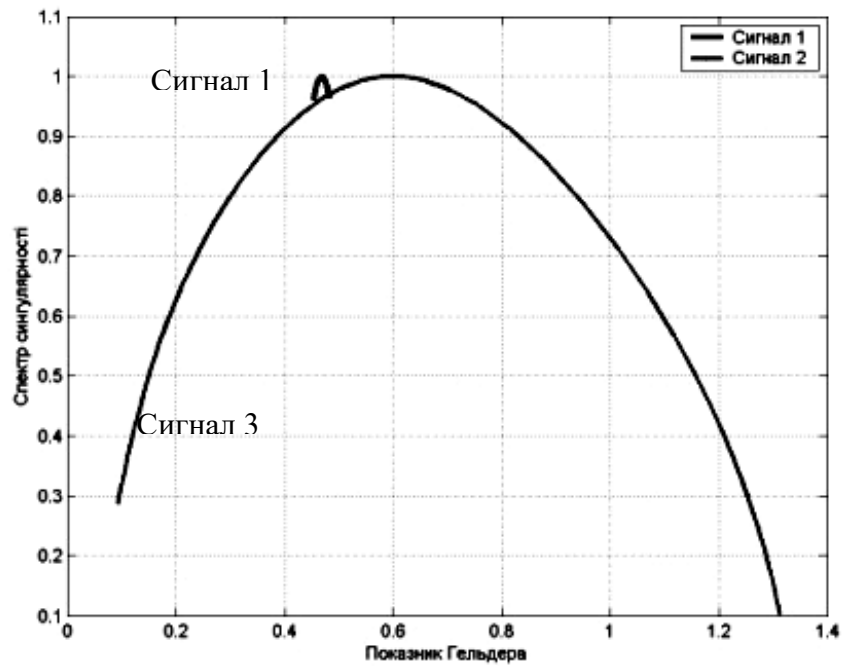


Рисунок 4.23. Спектри сингулярностей експериментальних сигналів

Сигнал 3 в цілому виявив мультифрактальні властивості: його скейлінгова експонента є кривою лінією, показник Гельдера залежить від значення порядку, а спектр сингулярностей не є компактною множиною. Разом з тим, при обмеженні варіацій значень порядку експериментальний сигнал 3 може бути апроксимований монофрактальними моделями.

4.5. Прогнозування вмісту кремнію в чавуні під час його виплавлення

Прогнозування вмісту хімічного складу чавуну на виході (що є одним з основних показників плавки) може значно підвищити якість контролю процесу доменної печі.

У якості АФА використовувались Anfis, Genfis2 та Genfis3, а як метод параметричної оптимізації був використаний метод золотого перетину, за яким змінювались наступні параметри АФА:

- для Anfis – належна функція прихованого шару;
- для Genfis2 – діапазон впливу центру кластера R_c ;
- для Genfis3 – кількість кластерів k_c .

Початкова послідовність експериментальних даних вмісту кремнію в чавуні була розділена на навчальну (А) та тестову (В) зразки нарівно, а в якості критерію оптимізації була використана відносна середньоквадратична похибка між реальними (Y^*) та передбачуваними ($\hat{Y}$) значеннями, обчислена на тестовому зразку (В):

$$C = \frac{\|Y_B^*[k+n] - \hat{Y}_B[k+n]\|}{\|Y_B^*[k+n]\|}. \quad (4.4)$$

Решта параметрів АФА (тренувальна функція для Anfis, нечіткий логічний алгоритм для Genfis3) змінювались шляхом перебору.

Глибина прогнозу приймалась на 5 циклів випуску чавуну, а глибина пам'яті – 4 цикли.

Похибка прогнозу обчислювалась як відносна похибка на кожному такті прогнозу:

$$C_{pr} = \|Y^*[k] - \hat{Y}[k]\| / \|Y^*[k]\|. \quad (4.5)$$

Результати прогнозування вмісту кремнію в чавуні на виході наведені на рис. 4.24, де позначено: TREND – часовий сигнал хімічного складу чавуну на виході, PREDICT – його прогноз, URR – показник режиму прогнозу.

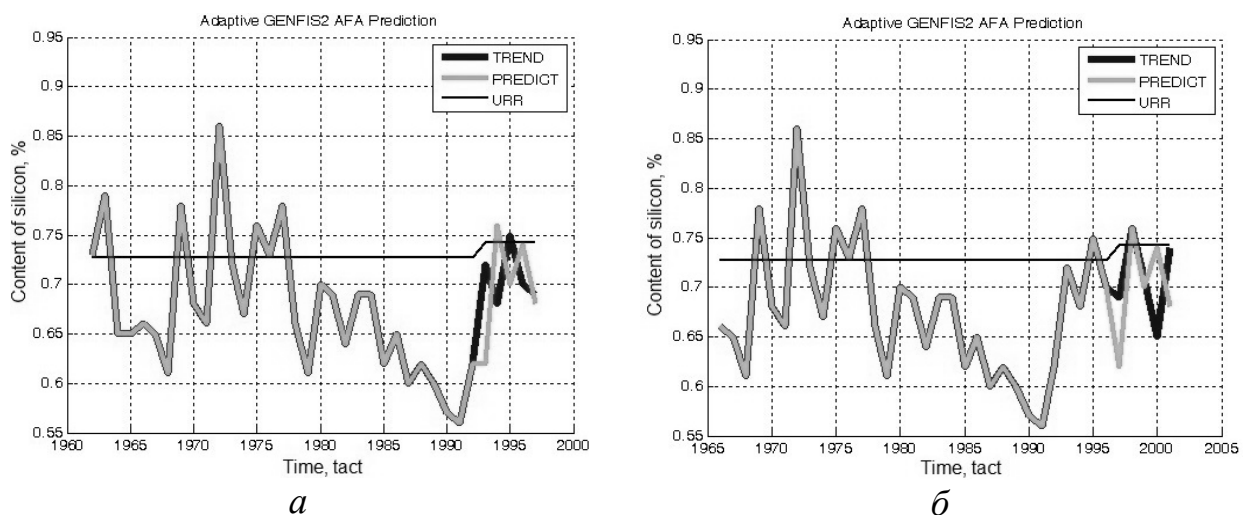


Рисунок 4.24. Прогноз вмісту кремнію протягом 1 (а) та 5 (б) циклів наперед

Відносні похибки прогнозу склали: Anfis – 0,0653; Genfis2 – 0,0563; Genfis3 – 0,0625. Встановлено, що мінімуму критерію (4.4) відповідає АФА Genfis2 з впливом центру кластера $R_c = 0,4$. При цьому забезпечується і найменше значення відносної похибки прогнозу (4.5), що для Genfis2 дорівнює 0,0563.

Статистична перевірка за критерієм знаків показала значимість отриманих результатів.

4.6. Прогнозування теплового стану доменної печі та характеристик якості чавуну

Ідентифікація має на меті отримання математичних моделей ОК, придатних для використання в контурах керування. Вона виконується на основі експериментальних даних про ОК, які вимагають, як правило, попередньої обробки. Це дозволяє отримати характеристики ОК, використовувані далі для формування її математичної моделі.

Відповідно до отриманих експериментальних даних, середній час між випусками чавуну (тривалість плавок) складає $\bar{t}_{pl} = 116$ хв (близько 2 годин), а його середньоквадратичне відхилення – $\sigma_{pl} = 20,2$ хв. Тоді коефіцієнт варіації тривалості плавок – $K_{pl} = \sigma_{pl} / \bar{t}_{pl} = 0,174$ а значення інтервалу керування прийняте рівним $\Delta t = 20$ хв.

Керування здійснюється за контрольованим рівнем розплаву (РР), що характеризує тепловий стан доменної печі. Для визначення динамічних властивостей сигналу РР по його миттєвих значеннях був отриманий тренд РР, згладжений за допомогою вейвлет фільтру по середньому значенню часу плавки $\bar{t}_{pl}$, а також обчислена АКФ тренда РР і його спектральна щільність (рис. 4.25).

Визначимо характеристики стану процесу, що породжує тренд РР доменної печі. По вигляду часо-частотних характеристик тренда РР можна сказати про нерегулярність процесу, який його породжує. Це витікає з нерегулярного виду часової реалізації сигналу тренда (див. рис. 4.25,а), авторегресійного виду його

АКФ (рис. 4.25,б) і наявності значної частини енергії спектру сигналу в низькочастотній області (див. рис. 4.25,в).

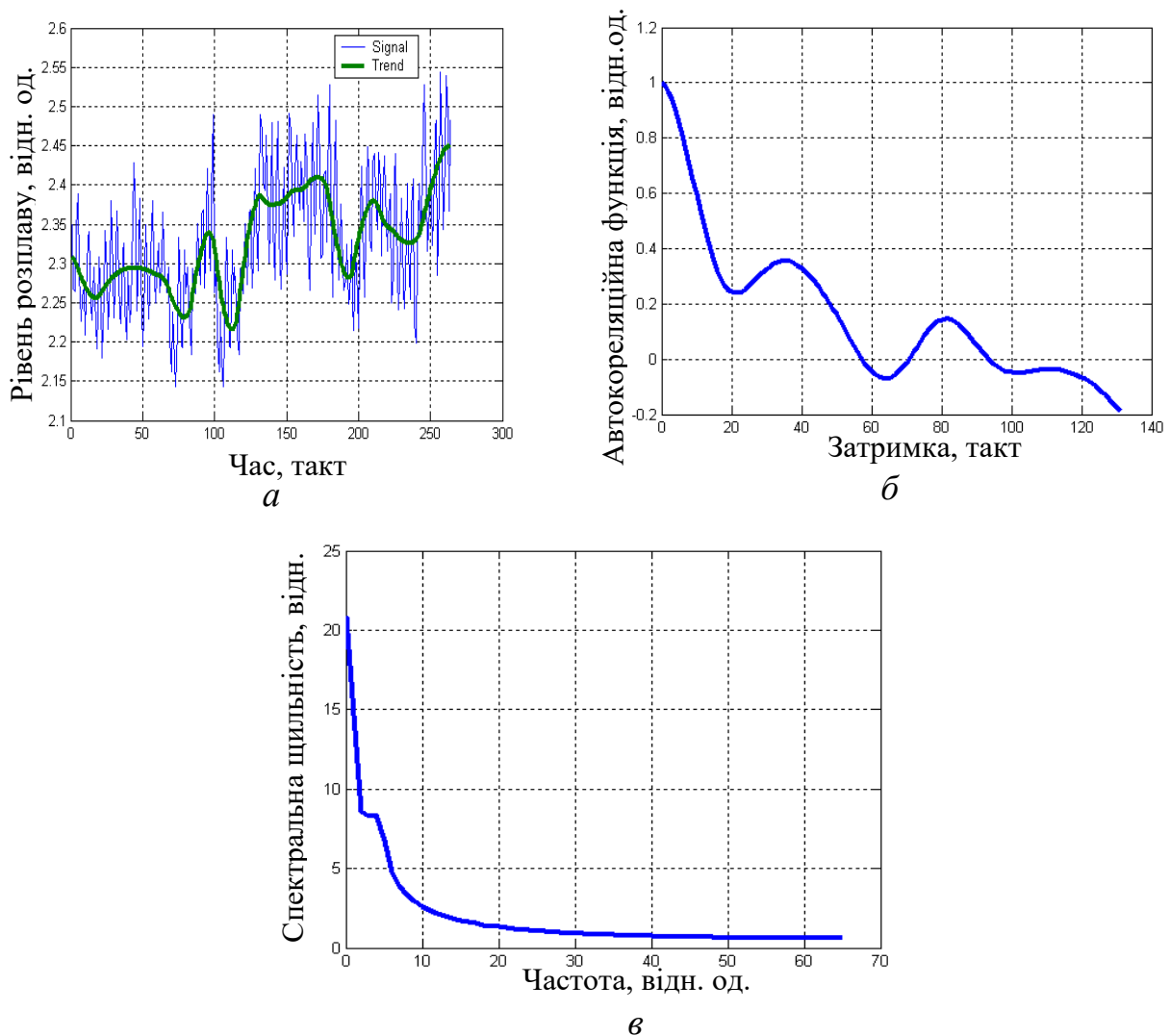


Рисунок 4.25. Реалізація РР (Signal) і його тренда (Trend) (а), автокореляційна функція (б) і спектральна щільність (в) тренда

Розрахунки виконані за допомогою стандартних програм дозволили визначити для тренда РР его кореляційну ентропію $K_c = 0,49$, кореляційну розмірність $D_c = 2,21$ та показник Херста $H = 0,21$. При цьому, інтервал точної передбачуваності (глибина точного прогноза) склав $T_c = 6,17$ такту (довжиною $\Delta t = 20$ хв).

Для визначення розмірності d фазового простору (глибини пам'яті) тренда РР обчислена її оцінка зверху $d \leq 5$. Кореляційна розмірність D_c практично перестає збільшуватися (входить в насичення) при розмірності фазового простору $d \geq 3$. Таким чином отримуємо, що $3 \leq d \leq 5$.

Показник Херста $H = 0,21 < 0,5$ характеризує тренд РР, як ергодичний мінливий процес, що складається з частих спадів-підйомів. Тобто збільшення (зменшення) значень процесу у минулому означає їх вірогідне зменшення (збільшення) в майбутньому.

Для прогнозування тренда РР використовувалися лінійний (Lin), нейронечіткий (Anfis), нейровейвлетний (WVNN) і нейромережевий (NN) АФА.

Як Lin АФА використовувався адаптивний КІХ-фільтр 8-го порядку з нормалізованим градієнтним алгоритмом і кроком адаптації, рівним 0,75.

В WVNN АФА використовувалися вейвлети типу симлет п'ятого порядку з двома рівнями розкладання і каскадної НМ з 32 нейронами в прихованому шарі з сигмоїдальними функціями активації і 1 лінійним нейроном у вихідному шарі.

Для побудови Anfis і NN АФА використовувалася структура Вінера-Гаммерштейна з кількістю ліній затримок на вході (глибиною пам'яті), рівній розмірності фазового простору тренда РР $d = 4$, визначеною вище.

Як Anfis АФА використовувалася система нечіткого висновку структури Сугено з дзвіноподібною функцією належності, двома рівнями розкладання по кожному входу і 15 епохами навчання.

В NN АФА використовувалася каскадна НН прямого розповсюдження з 64 нейронами в прихованому шарі з сигмоїдальними функціями активації і 1 лінійним нейроном у вихідному шарі, а кількість циклів навчання обмежувалася 300.

Початкова послідовність тренда РР розбивалася на навчальну і перевіірочну послідовності нарівно, а як критерій оптимізації використовувалася відносна середньоквадратична похибка між реальними і прогнозованими значеннями тренда РР в перевіірочній послідовності (8.44).

Моделювання АФА виконувалося з глибиною прогнозу 40 тактів ($40 \cdot \Delta t = 40 \cdot 20 \text{ хв} = 800 \text{ хв} \approx 13 \text{ год}$). Результати представлені на рис. 4.26.

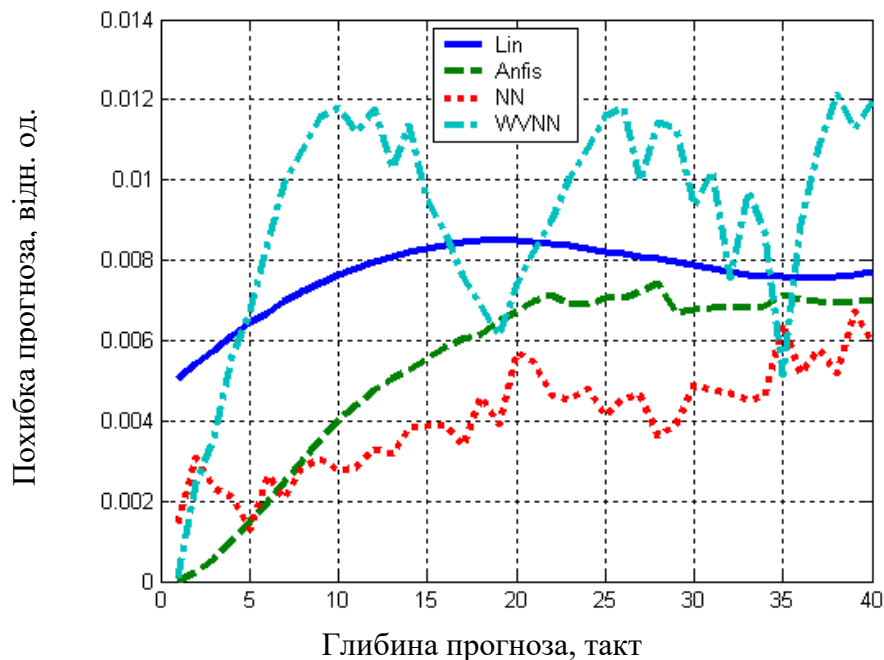


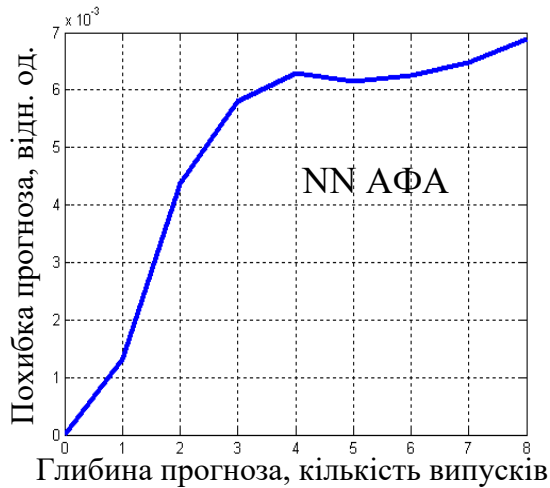
Рисунок 4.26. Похибки прогнозу тренда РР для різних АФА: лінійного (Lin), нейронечіткого (Anfis), нейромережевого (NN) і нейровейвлетного (WVNN)

Аналіз показує, що похибка прогнозу не перевищує 0,012. При цьому, найменші похибки має NN АФА.

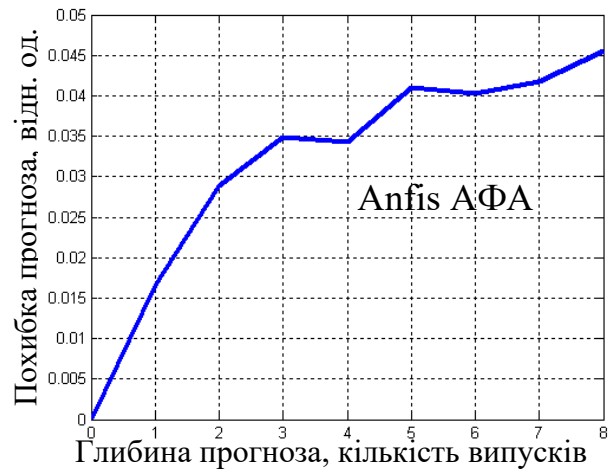
Аналогічно виконувались оцінки похибок прогнозування показників якості чавуну на випуску. Глибина прогнозу складала 8 випусків ($8 \cdot \bar{t}_{pl} = 8 \cdot 116 \text{ хв} = 928 \text{ хв} \approx 15 \text{ год}$). На рис. 4.27 наведені графіки похибок тих АФА, які дали кращі результати.

При моделюванні всі показники були занормовані в межі ± 1 , а їх середні значення склали для: температури чавуну – 1481 °С, вмісту кремнію – 0,755 % і вмісту сірки – 0,022 %.

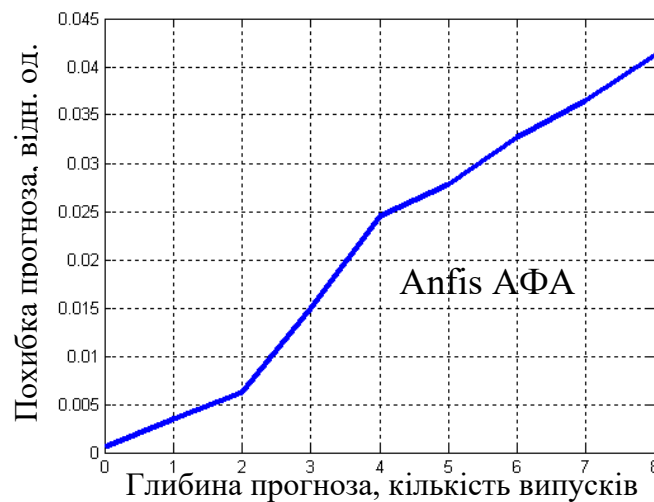
Аналіз результатів моделювання показує, що відносна похибка прогнозу для температури чавуну не перевищує 1 %, а для вмісту кремнію і сірки – 5 % при глибині прогнозу до 8 випусків чавуну.



a



б



в

Рисунок 4.27. Похибки прогнозу температури чавуна (*a*) та вмісту кремнія (*б*) і сірки (*в*) в ньому

Статистична перевірка за непараметричним критерієм знаків підтвердила, що результати прогнозу з рівнем значущості 0,01 адекватні експериментальним реалізаціям.

4.7. Контрольні запитання до глави 4

1. Які особливості фільтрації голосу за допомогою вейвнет?
2. Які переваги та недоліки апроксимації складних функцій за допомогою вейвнет?
3. Які особливості ідентифікація нелінійного динамічного процесу із нормально розподіленим шумом?
4. Як здійснюється придушення шумів сигналів мобільного зв'язку з нейронечіткою ідентифікацією тракту передачі?
5. В чому полягає комплексна оцінка характеристик при моделюванні трафіка.
6. Розкрийте зміст часо-частотного аналізу трафіка.
7. Розкрийте зміст статистичного аналізу трафіка.
8. Розкрийте зміст фрактального аналізу трафіка.
9. Розкрийте зміст методики оцінки характеристик мережевого трафіка.
10. Розкрийте зміст моделей мережевого самоподібного трафіка.
11. В чому полягає мультифрактальний аналіз трафіка.
12. В чому полягає мультифрактальний аналіз детрендових флуктуацій.
13. Розкрийте зміст методу максимумів модулів вейвлет-перетворення для аналізу мережевого трафіку.
14. Які особливості моделювання трафіка із використанням блочно-орієнтованих структур?
15. Як вирішується задача прогнозування вмісту кремнію в чавуні при його плавці за допомогою адаптивних фільтрів-апроксиматорів?
16. Які особливості прогнозування теплового стану доменної печі та показників якості чавуну за допомогою адаптивних фільтрів-апроксиматорів?

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Gorev V., Gusev A., Korniienko V. The use of the Kolmogorov–Wiener filter for prediction of heavy-tail stationary processes // CEUR Workshop Proceedings, 2022. Vol. 3156. P. 150–159.
2. Gorev V., Gusev A., Korniienko V., Shedlovska Y. On the use of the Kolmogorov – Wiener filter for heavy-tail process prediction // Journal of Cyber Security and Mobility. 2023. Vol. 12, Issue 3. P. 315–338.
3. Gusev O. Yu., Gorev V. M., Korniienko V. I. Theory of Adaptive Filtration : tutorial. Dnipro, NTU «DP», 2019. 156 p.
4. Горєв В. М. Теорія адаптивної фільтрації. Методичні рекомендації до практичних занять дисципліни для здобувачів-магістрів спеціальності 172 Телекомунікації та радіотехніка. Ч. 1. Дніпро : НТУ «ДП», 2022. 29 с.
5. Горєв В. М. Теорія адаптивної фільтрації. Методичні рекомендації до практичних занять з дисципліни для здобувачів-магістрів спеціальності 172 Телекомунікації та радіотехніка. Ч 2. Дніпро : НТУ «ДП», 2022. 29 с.
6. Miller S., Childers D. Probability and Random Processes With Applications to Signal Processing and Communications. Second edition. Amsterdam : Elsevier/Academic Press, 2012. 598 p.
7. Новицький І. В., Ус С. А. Випадкові процеси : навч. посіб. Дніпро : НГУ, 2011. 125 с.
8. Diniz P. S. R. Adaptive Filtering. Algorithms and Practical Implementation. Cham : Springer Nature Switzerland, 2020. 495 p.

Навчальне видання

Горєв В'ячеслав Миколайович
Корнієнко Валерій Іванович
Гусєв Олександр Юрійович

**АДАПТИВНА ФІЛЬТРАЦІЯ:
ТЕОРІЯ ТА ЗАСТОСУВАННЯ**

Навчальний посібник

Видано авторській редакції.

Підписано до друку 10.09.2024. Формат 30×84¹/₁₆ .

Папір офсетний. Друк цифровий.

Обл.-вид. арк. 9,25. Ум. друк. арк. 8,6.

Наклад 300 пр. Зам. № 003

Підготовлено до видання

в Національному технічному університеті «Дніпровська політехніка».

Свідоцтво про внесення до Державного реєстру ДК № 1842 від 11.06.2004.

49005, м. Дніпро, просп. Дмитра Яворницького, 19.

Видавництво «Журфонд»

49001, Дніпро, вул. Старокозацька, 8.

Свідоцтво про внесення до Державного реєстру

ДК № 684 від 21.11.2001 р.

Віддруковано:

ПП Вахмістров О.Є.,

м. Дніпро, вул. Писаржевського, буд. 18.

Горєв В. М.

Г67 Адаптивна фільтрація: теорія та застосування : навч. посіб. / В. М. Горєв, В. І. Корнієнко, О. Ю. Гусєв ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : Журфонд, 2025. – 146 с.

ISBN 978-966-934-634-6

Викладено основи теорії адаптивної фільтрації даних із застосуванням відомих засобів – фільтра Калмана та фільтрів, побудованих на RLS- і LMS-алгоритмах, а також тих, що створені штучним інтелектом (нейронних мережах, системах нечіткого висновку тощо).

Розглянуто застосування адаптивних фільтрів до розв'язування задач спостереження, ідентифікації, фільтрації голосу, аналізу й моделювання трафіку телекомунікаційних систем та прогнозування технологічних змінних. Набуті знання дають можливість засвоїти принципи побудови та сферу застосування адаптивних фільтрів.

Посібник призначено для здобувачів ступеня магістра спеціальності 172 Електронні комунікації та радіотехніка. Також він може бути корисним студентам інших технічних або фізико-математичних спеціальностей.

УДК 004.021