

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



В.І. Корнієнко, О.Ю. Гусєв, О.В. Герасіна

**ІНТЕЛЕКТУАЛЬНЕ МОДЕЛЮВАННЯ
НЕЛІНІЙНИХ ДИНАМІЧНИХ ПРОЦЕСІВ
У СИСТЕМАХ КЕРУВАННЯ, КІБЕРБЕЗПЕКИ,
ТЕЛЕКОМУНІКАЦІЙ**

Підручник

За загальною редакцією професора В.І. Корнієнка

Дніпро
НТУ «ДП»
2020

УДК 004.942 +004.89+681.5.015

К 67

Рекомендовано вченою радою як підручник для студентів спеціальностей 125 Кібербезпека та 172 Телекомунікації та радіотехніка (протокол № 5 від 28.05.2020).

Рецензенти:

Л.Ф. Політанський – д-р техн. наук, проф., завідувач кафедри радіотехніки та інформаційної безпеки (Чернівецький національний університет ім. Ю. Федьковича);

С.В. Толюпа – д-р техн. наук, проф., професор кафедри кібербезпеки та захисту інформації (Київський національний університет ім. Т. Шевченка);

М.О. Алексеєв – д-р техн. наук, проф., декан факультету інформаційних технологій (Національний технічний університет «Дніпровська політехніка»).

Корнієнко В.І.

Інтелектуальне моделювання нелінійних динамічних процесів у
К 67 системах керування, кібербезпеки, телекомунікацій: підручник /
В.І. Корнієнко, О.Ю. Гусєв, О.В. Герасіна ; за заг. ред. В.І. Корнієнка ;
М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». –
Дніпро : НТУ «ДП», 2020. – 536 с.

ISBN 978–966–350–735–4

Розглянуто інтелектуальні методи обробки інформації: технології штучних нейронних мереж, методи та алгоритми їх навчання, архітектури побудови нейронних мереж та їх ефективність. Викладено основи нечіткої логіки та алгоритми побудови систем нечіткого висновку. Наведено методи еволюційного моделювання та агентські алгоритми оптимізації рою частинок; алгоритми мурашиної колонії та бджолиного рою. Подано теоретичні основи моделювання нелінійних динамічних процесів. Викладено основи нелінійної динаміки та теорії хаосу, фрактальний аналіз динамічних процесів та їх частотно-часовий аналіз. Розглянуто практичні застосування інтелектуальних методів моделювання нелінійних динамічних процесів у системах керування, кібербезпеки та телекомунікацій.

Для магістрів спеціальностей 125 Кібербезпека та 172 Телекомунікації та радіотехніка денної та заочної форм навчання.

УДК 004.942

ISBN 978–966–350–735–4

© В.І. Корнієнко, О.Ю. Гусєв, О.В. Герасіна, 2020
© НТУ «Дніпровська політехніка», 2020

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	10
ВСТУП	13
ЧАСТИНА І. ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ОБРОБКИ ІНФОРМАЦІЇ	20
Глава 1. НЕЙРОННІ МЕРЕЖІ	20
1.1. Основи технології штучних нейронних мереж	24
1.1.1. Визначення.....	24
1.1.2. Області використання нейронних мереж	26
1.1.3. Біологічний нейрон	26
1.1.4. Історія розвитку штучних нейронних мереж.....	27
1.1.5. Структура та властивості штучного нейрона. Функції активації	28
1.1.6. Класифікація нейронних мереж	32
1.1.7. Топологія нейронних мереж	33
1.2. Навчання нейронних мереж	38
1.2.1. Навчання із вчителем.....	38
1.2.2. Алгоритм зворотного поширення похибки	40
1.2.3. Навчання нейронних мереж без вчителя	43
1.3. Архітектури нейронних мереж	48
1.3.1. Перцептрони	48
1.3.2. Нейронні мережі зустрічного розповсюдження	51
1.3.3. Нейронні мережі з радіальними базисними функціями.....	55
1.3.4. Нейронні мережі Хопфілда і Хемінга	58
1.3.5. Ймовірнісні нейронні мережі.....	64
1.3.6. Узагальнено-регресійні нейронні мережі	66
1.3.7. Лінійні нейронні мережі.....	67
1.4. Ефективність нейронних мереж	68
Глава 2. НЕЧІТКА ЛОГІКА І СИСТЕМИ НЕЧІТКОГО ВИСНОВКУ	70
2.1. Вступ.....	70
2.2. Нечіткі множини.....	72
2.2.1. Визначення нечітких множин	72
2.2.2. Основні характеристики нечітких множин	73
2.2.3. Функції належності нечітких множин	74
2.2.4. Методи побудови функцій належності	78
2.2.5. Операції над нечіткими множинами	80
2.3. Нечіткі змінні та відношення	84
2.3.1. Нечіткі та лінгвістичні змінні	84
2.3.2. Нечіткі числа.....	85
2.3.3. Нечіткі відношення та операції над ними	86
2.4. Системи нечіткого висновку	88
2.4.1. Нечіткі висновки	88
2.4.2. Алгоритм Мамдані (Mamdani).....	90
2.4.3. Алгоритм Сугено (Sugeno).....	91

2.4.4. Алгоритм Цукамото (Tsukamoto)	92
2.4.5. Алгоритм Ларсена (Larsen)	93
2.4.6. Методи приведення до чіткості	95
2.5. Нечітка кластеризація	96
2.5.1. Кластерний аналіз	96
2.5.2. Субтрактивна кластеризація	99
2.5.3. Нечітка кластеризація <i>C</i> -середніх	100
2.6. Ефективність систем з нечіткою логікою	102
2.7. Нейронечіткі мережі	103
Глава 3. ЕВОЛЮЦІЙНЕ МОДЕЛЮВАННЯ	108
3.1. Вступ	108
3.2. Біологічні передумови й загальна схема еволюційних алгоритмів	110
3.3. Генетичні алгоритми	114
3.3.1. Формування початкової популяції	116
3.3.2. Оцінка осіб популяції	117
3.3.3. Відбір (селекція)	117
3.3.4. Схрещування	118
3.3.5. Мутація	119
3.3.6. Формування нового покоління	121
3.3.7. Зупинка алгоритму	122
3.4. Приклади реалізацій генетичних алгоритмів	122
3.4.1. Канонічний генетичний алгоритм	122
3.4.2. Генетичний алгоритм Genitor	123
3.4.3. Гібридні генетичні алгоритми	124
3.4.4. Острівний генетичний алгоритм	125
3.4.5. Генетичний алгоритм перехресного відбору	125
3.4.6. Теорема схем	126
3.5. Еволюційна стратегія	129
3.6. Еволюційне програмування	133
3.7. Диференціальна еволюція	136
3.8. Метод групового урахування аргументів	139
3.8.1. Метод найменших квадратів	139
3.8.2. Загальна схема побудови алгоритмів методу	141
3.8.3. Алгоритм з коваріаціями і з квадратичними описами	143
3.8.4. Критерії селекції методу	144
Глава 4. АГЕНТСЬКІ АЛГОРИТМИ ОПТИМІЗАЦІЇ	147
4.1. Алгоритм оптимізації роєм часток	147
4.1.1. Канонічний алгоритм рою часток	147
4.1.2. Модифікації канонічного алгоритму рою часток	152
4.1.3. Топології сусідства часток	158
4.1.4. Алгоритми з динамічної топологією сусідства часток	161
4.2. Мурашина оптимізація	163

4.2.1. Біонічні передумови.....	165
4.2.2. Алгоритм безперервної оптимізації колонією мурах.....	166
4.2.3. Алгоритм безперервно взаємодіючої колонії мурах.....	175
4.2.4. Безперервний ортогональний алгоритм мурашиної колонії.....	175
4.3. Оптимізація бджолиним роєм.....	180
4.3.1. Біонічні передумови.....	181
4.3.2. Бджолиний алгоритм.....	183
4.3.3. Алгоритм колонії штучних бджіл.....	188

Частина II. МОДЕЛЮВАННЯ НЕЛІНІЙНИХ ДИНАМІЧНИХ ПРОЦЕСІВ

Глава 5. НЕЛІНІЙНА ДИНАМІКА ТА ХАОС.....	196
5.1. Динамічні системи.....	196
5.1.1. Визначення динамічної системи, класифікація.....	196
5.1.2. Регулярні і хаотичні атрактори.....	200
5.1.3. Системи з дискретним часом. Відображення послідування.....	203
5.1.4. Стійкість динамічних систем.....	208
5.2. Біфуркації динамічних систем.....	217
5.2.1. “М’які” і “жорсткі” біфуркації. Катастрофи.....	218
5.2.2. Біфуркація “триразова рівновага”.....	220
5.2.3. Біфуркація Андронова–Хопфа.....	222
5.2.4. Біфуркації граничних циклів.....	223
5.3. Детермінований хаос в динамічних системах.....	228
5.3.1. Детермінованість і хаос.....	230
5.3.2. Стійкість і нелінійність.....	231
5.3.3. Нестійкість і нелінійне обмеження.....	232
5.3.4. Детермінований хаос.....	235
5.3.5. Перемішування.....	236
5.3.6. Ймовірнісні властивості детермінованих систем.....	238
5.3.7. Детермінований хаос – властивість матеріального світу.....	239
5.3.8. Дивні хаотичні атрактори.....	240
5.3.9. Дивні нехаотичні і хаотичні недивні атрактори.....	241
5.4. Сценарії розвитку хаосу. Типи атракторів.....	242
5.4.1. Перехід до хаосу через послідовність біфуркацій подвоєння періоду.....	244
5.4.2. Жорсткі переходи до хаосу. Криза і перемежана.....	249
5.4.3. Перехід до хаосу через руйнування квазіперіодичних коливань.....	251
5.4.4. Грубі і негрубі динамічні системи.....	252
5.4.5. Квазігіперболічні атрактори. Атрактори типу Лоренца.....	254
5.4.6. Квазіатрактори і їх властивості.....	256
5.5. Реконструкція динамічних систем.....	257
5.5.1. Визначення розмірності вкладення і реконструкція атрактора.....	261
5.5.2. Розрахунок старшого показника Ляпунова по часовому ряду.....	264

5.5.3. Реконструкція динамічної системи	266
5.5.4. Приклад реконструкції динамічної системи	267
Глава 6. ФРАКТАЛЬНИЙ АНАЛІЗ	272
6.1. Основні визначення.....	272
6.1.1. Поняття фракталів.....	272
6.1.2. Класифікація фракталів	274
6.2. Стохастичні шуми	277
6.2.1. Рівняння Ланжевена.....	278
6.2.2. Гаусові процеси.....	280
6.2.3. Вінерівський процес	280
6.2.4. Білий гаусовий шум	281
6.2.5. Пуассонівський процес.....	283
6.2.6. Білий дробовий шум	283
6.2.7. Кольоровий шум.....	284
6.3. Фрактальні шуми.....	285
6.3.1. Самоподібні степеневі закони	285
6.3.2. Фрактальний метод нормованого розмаху Херста (R/S-аналіз) ...	285
6.3.3. Класифікація фрактальних шумів	287
6.3.4. Рожевий шум	289
6.3.5. Коричневий шум і самоподібні процеси на фондовій біржі	292
6.3.6. Чорні шуми і розливи Нілу	293
6.4. Фрактальна розмірність.....	296
6.4.1. Загальне поняття фрактальної розмірності	296
6.4.2. Узагальнені фрактальні розмірності	299
6.4.3. Основні властивості розмірності Хаусдорфа і інформаційна розмірність	302
6.4.4. Кореляційна розмірність	304
6.4.5. Розмірність подібності і її експериментальне визначення	305
6.4.6. Визначення фрактальної розмірності фазової траєкторії і поверхні.....	307
6.4.7. Визначення фрактальної розмірності для часових рядів.....	309
6.4.8. Дисперсійна розмірність	311
6.4.9. Визначення розмірності за максимумами	314
6.4.10. Визначення хаотичності режиму і наявності дивного атратора	318
6.5. Мультифрактальний скейлінг-спектр	319
6.5.1. Послідовність показників маси.....	320
6.5.2. Функція мультифрактального спектру	323
6.5.3. Геометричні властивості мультифрактального скейлінг-спектра	326
6.6. Фрактальний аналіз часових рядів	330
Глава 7. ЧАСТОТНО-ЧАСОВИЙ АНАЛІЗ	337
7.1. Частотно-часові властивості базисних функцій.....	337
7.1.1. Площина частота-час	337

7.1.2. Обмежене в часі Фур'є-перетворення	340
7.1.3. Принцип невизначеності	341
7.2. Базисні функції частотно-часового аналізу	342
7.3. Безперервне вейвлет-перетворення	344
7.4. Властивості безперервного вейвлет-перетворення	347
7.5. Дискретне вейвлет-перетворення	353
7.5.1. Дискретизація масштабу	354
7.5.2. Дискретизація масштабу і зсуву. Фрейми	356
7.5.3. Приклади вейвлетів для дискретного перетворення	358
7.6. Швидкий алгоритм дискретного вейвлет-перетворення	364
7.7. Узагальнений алгоритм частотно-часового вейвлет-аналіза	368
7.8. Перетворення Гільберта-Гуанга	370
7.8.1. Емпіричний метод декомпозиції	371
7.8.2. Спектральний аналіз Гільберта	377
7.8.3. Емпірична декомпозиція шумових сигналів	378

ЧАСТИНА ІІІ. МОДЕЛЮВАННЯ СКЛАДНИХ ПРОЦЕСІВ В КЕРУВАННІ, КІБЕРБЕЗПЕЦІ ТА ТЕЛЕКОМУНІКАЦІЯХ

Глава 8. ІДЕНТИФІКАЦІЯ ТА ПРОГНОЗУВАННЯ СКЛАДНИХ НЕЛІНІЙНИХ ПРОЦЕСІВ ЯК ОБ'ЄКТІВ КЕРУВАННЯ	385
8.1. Адаптивне прогнозування сигналів і стану об'єктів керування	385
8.1.1. Лінійний адаптивний фільтр-апроксиматор	387
8.1.2. Нейромережевий адаптивний фільтр-апроксиматор	388
8.1.3. Нейронечіткий адаптивний фільтр-апроксиматор	389
8.1.4. Нечіткий адаптивний фільтр-апроксиматор з субтрактивною кластеризацією	390
8.1.5. Нечіткий адаптивний фільтр-апроксиматор з кластеризацією С- середніх	391
8.1.6. Нейронний вейвлет адаптивний фільтр-апроксиматор	393
8.2. Ідентифікація нелінійних процесів за часовими реалізаціями	395
8.2.1. Характеристики нелінійних хаотичних процесів	396
8.2.2. Визначення характеристик стану нелінійних процесів	397
8.3. Приклади ідентифікації породжуючих процесів та прогнозування їх стану	401
8.3.1. Ідентифікація процесу дроблення руди та прогнозування її крупності	401
8.3.2. Прогнозування вмісту кремнію в чавуні при його плавці	405
8.3.3. Прогнозування теплового стану доменної печі та характеристик якості чавуну	413
8.4. Структурно-параметрична ідентифікація об'єктів керування	418
8.4.1. Постановка задачі	418
8.4.2. Структурна оптимізація	419
8.4.3. Критерії оптимізації та обмеження	422
8.4.4. Структура і базисні функції моделі	424

8.4.5. Приклад структурно-параметричної ідентифікації нелінійного об'єкту	428
---	-----

Глава 9. КОМПЛЕКСНА ОЦІНКА, ІДЕНТИФІКАЦІЯ ТА ПРОГНОЗУВАННЯ ТРАФІКУ В ІНФОРМАЦІЙНИХ ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖАХ	432
9.1. Комплексна оцінка характеристик трафіку	433
9.1.1. Часо-частотний аналіз.	434
9.1.2. Статистичний аналіз.	435
9.1.3. Фрактальний аналіз.	439
9.1.4. Методика оцінки характеристик мережевого трафіку.	441
9.2. Моделювання самоподібного трафіку	443
9.2.1. Моделі мережевого трафіку.	443
9.2.2. Приклад аналізу та моделювання мережевого трафіку	445
9.3. Мультифрактальний аналіз трафіку	453
9.3.1. Мультифрактальні моделі.	455
9.3.2. Мультифрактальний аналіз детрендових флуктуацій.	457
9.3.3. Метод максимумів модулів вейвлет-перетворення.	459
9.3.4. Приклад аналізу експериментальних даних та моделювання.	460
9.4. Структурно-параметрична ідентифікація та прогнозування трафіку	470
9.4.1. Приклад структурно-параметричної ідентифікації трафіку.	440
9.4.2. Приклад моделювання трафіку із використанням блочно-орієнтованих структур	474
9.4.3. Приклад адаптивного нечіткого прогнозування трафіку.....	476

Глава 10. ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНИХ МЕТОДІВ ПРИ РІШЕННІ ПРИКЛАДНИХ ЗАВДАНЬ	480
10.1. Адаптивне оцінювання та ідентифікація процесів за допомогою вейвнет	480
10.1.1. Структура вейнету з блоком із нескінченною імпульсною характеристикою.	480
10.1.2. Фільтрація голосу.	482
10.1.3. Фільтрація голосу при імпульсних завадах.	485
10.1.4. Апроксимація складних функцій.	486
10.1.5. Ідентифікація нелінійного динамічного процесу із нормально розподіленим шумом.	488
10.2. Оцінка систем захисту інформації від витоку	490
10.2.1. Оцінка DLP-систем з використанням засобів нечіткої логіки	490
10.2.2. Оцінка DLP-систем з використанням штучних нейронних мереж	493
10.3. Методика нечіткої оцінки технічного стану комп'ютерної системи в задачах інформаційної безпеки	494
10.4. Методика виявлення DoS-атак в інформаційно-комунікаційних мережах з використанням штучних нейронних мереж	498

10.4.1. Інформативні ознаки DoS-атак.	499
10.4.2. Критерії ефективності нейронних мереж.	502
10.4.3. Дослідження ефективності нейронних мереж для виявлення DoS-атак.	503
10.5. Нечітка маршрутизація пакетного трафіку з використанням мурашиного алгоритму навчання	503
10.5.1. Протоколи маршрутизації і алгоритми керування трафіком в IP-мережах.	508
10.5.2. Алгоритм маршрутизації на основі систем нечіткої логіки	509
10.5.3. Навчання системи нечіткого висновку на основі мурашиного алгоритму.	511
10.5.4. Оцінка ефективності алгоритму маршрутизації.	511
10.6. Нечіткі алгоритми IP-маршрутизації в телекомунікаційних мережах	513
10.6.1. Маршрутизація як процес керування.....	514
10.6.2. Нечітка маршрутизація в рамках протокола RIP.....	515
10.6.3. Нечітка маршрутизація в рамках протокола SNMP	519
10.7. Система нечіткого вибору міжсегментного інтервалу в транспортному протоколі телекомунікаційної мережі	521
10.8. Придушення шумів сигналів мобільного зв'язку з нейронечіткою ідентифікацією тракту передачі	525
10.9. Модифіковані мурашині алгоритми оптимізації у бездротових сенсорних мережах	528
10.9.1. Мурашиний алгоритм із використанням семафорів.	529
10.9.2. Мурашиний алгоритм із розділенням часу збирання інформації.	531
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ	534

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

- АДФ – аналіз детрендових флуктуацій
АКФ – автокореляційна функція
АТС – автоматична телефонна станція
АФА – адаптивний фільтр-апроксиматор
БО – багатокритеріальна оптимізація
БП – багат шаровий персептрон
БСМ – бездротова сенсорна мережа
ВА – вейвлет-аналіз
ВП – вейвлет перетворення
ГА – генетичний алгоритм
ГП – генетичне програмування
ДВП (DWT) – дискретне вейвлет перетворення
ДЕ – диференційна еволюція
ДС – динамічна система
ДФЕ – дробний факторний експеримент
ЕА – еволюційний алгоритм
ЕП – еволюційне програмування
ЕС – еволюційна стратегія
ЗДВП – зворотне дискретне вейвлет перетворення
ЗДР – звичайне диференціальне рівняння
ІТМ – інформаційна телекомунікаційна мережа
ІХ – імпульсна характеристика
КІХ – кінцева імпульсна характеристика
ЛДБ – лінійний динамічний блок
ЛХП – ляпуновські характеристичні показники
МА – мурашиний алгоритм
МВМ – мультифрактальна вейвлет модель
МГУА – метод групового урахування аргументів

МІВ – метод імітації відпалу
ММВП – (метод) максимумів модулів вейвлет перетворення
МПП – метод порогового прийняття
НВ – нейро-вейвлетний (фільтр)
НІХ – нескінченна імпульсна характеристика
НМ – нейронна мережа
НМПР – нейронна мережа прямого розповсюдження
НСБ – нелінійний статичний блок
ОК – об’єкт керування
ПВП – прямий випадковий пошук
ПУЗ – повільно убуваюча залежність
ПФЕ – повний факторний експеримент
РБФ (RBF)– радіальна базисна функція
ФБР – фрактальний броунівський рух
ФГШ – фрактальний гаусівський шум
ФН – функція належності
ЦП – центральний процесор
ШВП – швидке вейвлет перетворення
ШІ – штучний інтелект
ШУЗ – швидко убуваюча залежність
АВС – колонія штучних бджіл
АСО – оптимізація колонією мурах
Anfis – адаптивна нейронна система нечіткого висновку
В – бджолиний (алгоритм)
САСО – безперервна оптимізація колонією мурах
СІАС – безперервно взаємодіюча мурашина колонія
СНС – перехресний вибір популяції
СОАС – безперервний ортогональний (алгоритм) мурашиної колонії
DLP – програма (система) захисту інформації від її витоку
DoS – (інформаційна атака) типу «відмова в обслуговуванні»

DSPSO – динамічна соціометрична оптимізація роєм часток
EMD – емпіричний метод декомпозиції
FARIMA – фрактальна авторегресійна інтегральна модель з ковзним середнім
Genfis2 – алгоритм субтрактивної кластеризації
Genfis3 – алгоритм кластеризації C -середніх
GRNN – узагальнено-регресійна нейронна мережа
ННТ – перетворення Гільберта-Гуанга
HSA – спектральний аналіз Гільберта
IDPS – програма (система) виявлення та попередження (інформаційних) атак
IMF – функція внутрішньої моди
PNN – ймовірнісна нейронна мережа
PSO – оптимізація роєм часток
RASP1 – раціональна функція другого порядку (вейвлет)

ВСТУП

Наразі характерним є широке застосування методів систем штучного інтелекту (нейронних мереж, систем нечіткого висновку, еволюційного моделювання тощо) та методів нелінійної динаміки (теорії хаосу та фрактального аналізу) для моделювання складних процесів в системах керування, кібербезпеки та телекомунікацій.

Їх актуальність обумовлена, по-перше, спроможністю методів нелінійної динаміки більш узагальнено з єдиних позицій описувати складні процеси в системах різної природи, а, по-друге, здатністю інтелектуальних методів розв'язувати оптимізаційні задачі, що погано формалізуються, та використанням для моделювання ефективних і універсальних апроксиматорів (нейронних мереж та систем нечіткого висновку).

Основна мета написання цієї книги полягає в наданні розуміння читачам о єдності законів та методів розв'язання задач в системах різної природи.

Для побудови інтелектуальних систем використовують різні підходи. При *структурному підході* здійснюють спроби побудови інтелектуальних систем шляхом моделювання структури людського мозку. Однією з перших таких спроб був перцептрон Ф. Розенблатта. Пізніше виникли й інші моделі, які наразі відомі під терміном «*нейронні мережі*». Ці моделі розрізняються за будовою окремих нейронів, за топологією зв'язків між ними і за алгоритмами навчання. Для нейронних моделей характерна більша виразність, легке розпаралелювання алгоритмів, а також пов'язана з цим висока продуктивність паралельно реалізованих нейронних мереж.

Основою для *логічного підходу* служить булева алгебра, яка має свій подальший розвиток у вигляді числення предикатів, в якому вона розширена за рахунок введення предметних символів, відносин між ними, кванторів існування та загальності. Домогтися більшої виразності логічного підходу дозволяє такий напрям, як *нечітка логіка*.

Теорію нечітких множин запропонував в 1965 р. професор Лотфі Заде, який розширив класичне поняття множини, допустивши, що характеристична функція (*функція належності* елемента множині) може приймати будь-які значення в інтервалі $[0; 1]$, а не тільки значення 0 або 1. Подальші роботи Л. Заде і його послідовників заклали міцний фундамент нової теорії і створили передумови для впровадження методів нечіткого висновку в інженерну практику.

Для більшості логічних методів характерна велика трудомісткість, оскільки під час пошуку доказу можливий повний перебір варіантів. Тому даний підхід вимагає ефективної реалізації обчислювального процесу, і його працездатність, зазвичай, гарантується при порівняно невеликому розмірі бази даних.

При побудові інтелектуальних систем за *еволюційним підходом* основна увага приділяється побудові початкової моделі і правилам, за якими вона може змінюватися (еволюціонувати). Причому модель може бути складена за допомогою найрізноманітніших методів (нейронних мереж, алгоритмів нечіткої логіки тощо). Далі система, побудована за еволюційним підходом, на підставі перевірки моделей відбирає найкращі з них (за певним критерієм), на підставі яких за самими різними правилами генеруються нові моделі, з яких знову обираються найкращі й так далі.

Прикладом систем штучного інтелекту, побудованих за еволюційним підходом є генетичні алгоритми, еволюційні алгоритми та еволюційні стратегії, які використовуються для розв'язання задач оптимізації складних систем різного походження. Суть їх парадигми полягає у використанні базових принципів теорії біологічної еволюції – відбору, мутації і відтворення осіб.

Ще один широко використовуваний підхід до побудови інтелектуальних систем – *імітаційний*, за яким моделюється властивість людини копіювати те, що роблять інші, не вдаючись у подробиці, навіщо це потрібно. Найчастіше ця здатність економить йому багато часу, особливо на початку його життя.

До імітаційних алгоритмів оптимізації відносять інспіровані живою природою *агентські алгоритми* (ройовий інтелект), серед яких отримали поширення алгоритми рою часток, мурашиної колонії та бджолиного рою.

Динамічний підхід до опису систем самого різного походження відомий з часів Ньютона. Він є основою аналізу більшості класичних явищ: спочатку будується відповідна математична модель у вигляді динамічних рівнянь, а потім тим або іншим способом вивчаються їх рішення, які, в принципі, можна зіставити з експериментальними даними.

Нелінійна динаміка – наукова дисципліна, що вивчає загальні закономірності нелінійних динамічних систем і процесів. Її важливість полягає в тому, що Мир в основному нелінійний і процеси, що відбуваються в ньому, – нелінійні.

Багато властивостей нелінійних систем істотно відрізняються від властивостей лінійної: нелінійна система не має властивості суперпозиції, яка в лінійній ситуації дозволяє конструювати будь-яке рішення з певного набору часткових рішень; зазвичай реалізується не один, а множина різних режимів функціонування (атрактори), які залежать від початкового стану, параметрів системи і зовнішніх дій; можливі детерміновані процеси, які зовні виглядають, як випадковий процес (детермінований хаос); при зміні параметрів, що управляють, відбувається зміна режимів функціонування системи, причому системи демонструють біфуркації і катастрофи (один режим втрачає стійкість, гине, йому на зміну приходить інший і так далі).

Нелінійна динаміка акумулює досягнення трьох наук: нелінійної термодинаміки необоротних процесів, що відповідає на питання чому у відкритих системах виникають дисипативні структури; власне нелінійної динаміки, що пояснює як саме виникають ці дисипативні структури і за якими законами змінюються в часі і просторі; та синергетики – науки про процеси самоорганізації, але рухається далі оскільки займається будь-якими системами і не тільки процесами самоорганізації, а процесами організації за цілеспрямованою дією зовнішніх сил, як сумісних, так й індивідуальних.

Загально визнано, що фрактальна геометрія є корисним інструментом для кількісного опису самих різних структур. Це можуть бути як абстрактно-математичні побудови, часові ряди, функціонування якої-небудь складної системи (наприклад, в теорія ігор, рух ціни товару на біржі тощо), так і природні об'єкти астрономії, геофізики, фізики, хімії, біології і медицини. Останніми роками технологія фрактального аналізу привертає все більшу увагу в області обробки сигналів і зображень.

У техніці аналізу природних фрактальних структур важливе місце займають методи, засновані на розрахунку ентропії К. Шеннона, А. Колмогорова, А. Ренї. Ентропійний підхід дає універсальний ключ як для розрахунків мультифрактальних спектрів розмірності систем, так і для осмислення і інтерпретації результатів їх фрактального аналізу. Це стало можливим, коли була виявлена глибока у фундаментальному плані і достатньо проста форма зв'язку між цими поняттями.

Ентропія асоціюється з мірою хаотичності системи, тоді як фрактал – з гранично впорядкованою, самоподібною і масштабно інваріантною структурою. Подальший розвиток теорії інформації розкрив глибокий зв'язок між інформацією і структурою. Зрозуміло, що ці поняття практично нерозривні – кожна структура несе в собі інформацію, і структура без інформації немислима, так само як і навпаки – неможлива інформація сама по собі – без носія, структури. Таким чином, інформація є універсальною, первинною категорією і скрізь присутня у вигляді структур. Стало також зрозуміло, що ентропія як міра хаотичності і мультифрактальність як міра структурної впорядкованості є взаємно доповнюючими поняттями.

Традиційні методи аналізу даних, як правило, засновані на лінійних і стаціонарних припущеннях. При цьому теорія вейвлетів дає зручний і ефективний інструмент для вирішення багатьох практичних завдань. Основна область застосування вейвлетних перетворень – аналіз і обробка сигналів і функцій, нестационарних в часі або неоднорідних в просторі, коли результати аналізу повинні містити не тільки загальну частотну характеристику сигналу

(розподіл енергії сигналу по частотних складових), але і відомості про певні локальні координати, на яких проявляють себе ті або інші групи частотних складових, або на яких відбуваються швидкі зміни частотних складових сигналу.

У більшості реальних систем, отримані дані найімовірніше будуть не тільки нестационарними, але і нелінійними. Необхідна умова представлення таких даних полягає в тому, щоб мати адаптивний базис. У недавно розробленому методі Гільберта-Гуанга намічено рішення цієї задачі. Метод включає емпіричний метод розкладання і спектральний аналіз Гільберта. У всіх вивчених випадках цей метод дав результати, набагато чіткіші, ніж традиційні методи аналізу в частотно-енергетичних часових представленнях.

Підручник складається із трьох частин. В *першій частині* (глави 1-4) розглянуті інтелектуальні методи обробки інформації. В *главі 1* викладені основи технології штучних нейронних мереж, методи та алгоритми їх навчання, архітектури побудови нейронних мереж та їх ефективність.

В *главі 2* викладені основи нечіткої логіки та алгоритми побудови систем нечіткого висновку: визначення нечітких множин, їх основні характеристики та функції над ними; нечіткі змінні та їх відношення; алгоритми Мамдані, Сугено, Цукамото та Ларсена; алгоритми нечіткої кластеризації; ефективність систем з нечіткою логікою та гібридні нейронечіткі мережі.

В *главі 3* викладені методи еволюційного моделювання: біологічні передумови й загальна схема еволюційних алгоритмів; генетичні алгоритми та приклади їх реалізацій; алгоритми еволюційних стратегій та програмування, диференціальної еволюції та методу групового урахування аргументів.

В *главі 4* викладені агентські алгоритми оптимізації: канонічний та модифіковані алгоритми рою часток; біонічні передумови та алгоритми мурашиної колонії; біонічні передумови та алгоритми бджолиного рою.

В *другій частині* (глави 5-7) розглянуто моделювання нелінійних динамічних процесів. В *главі 5* викладені основи нелінійної динаміки та теорії хаосу: визначення динамічної системи і її стійкість, регулярні і хаотичні

атрактори; біфуркації динамічних систем; детермінований хаос; сценарії розвитку хаосу і типи атракторів; реконструкція моделей динамічних систем.

В *главі 6* викладений фрактальний аналіз динамічних процесів: поняття і класифікація фракталів; стохастичні шуми; фрактальні самоподібні шуми; розмірність Хаусдорфа та інформаційна, кореляційна і дисперсійна фрактальні розмірності; мультифрактальний скейлінг-спектр та фрактальний аналіз часових рядів.

В *главі 7* викладений частотно-часовий аналіз сигналів, породжуваних нелінійними динамічними процесами: базові функції частотно-часового аналізу та їх властивості; безперервне та дискретне вейвлет-перетворення; узагальнений алгоритм частотно-часового вейвлет-аналізу; перетворення Гільберта-Гуанга та емпірична декомпозиція шумових сигналів.

В *третьій частині* (глави 8-10) розглянуто моделювання складних нелінійних процесів в системах керування, кібербезпеки та телекомунікацій.

В *главі 8* викладена ідентифікація та прогнозування складних нелінійних процесів як об'єктів керування: адаптивне прогнозування за допомогою лінійного, нейромережевого, нейронечіткого, нечітких з кластерізацією та нейронного вейвлет фільтрів-апроксиматорів; ідентифікація нелінійних динамічних процесів за їх часовими реалізаціями; приклади ідентифікації породжуючих процесів та прогнозування їх стану; структурно-параметрична ідентифікація об'єктів керування.

В *главі 9* викладені комплексна оцінка, ідентифікація та прогнозування трафіку в інформаційних телекомунікаційних мережах: комплексна оцінка характеристик трафіку за допомогою час-частотного, статистичного та фрактального аналізів; моделі та приклад моделювання мережевого трафіку; мультифрактальний аналіз та приклад аналізу експериментальних даних і моделювання трафіку; приклади структурно-параметричної ідентифікації та прогнозування трафіку.

В *главі 10* викладені практичні застосування інтелектуальних методів при рішенні завдань в керуванні, кібербезпеці та телекомунікаціях: адаптивне

оцінювання та ідентифікація процесів за допомогою вейвнет; оцінка систем захисту від витоку інформації з використанням засобів нечіткої логіки та нейронних мереж; методика нечіткої оцінки стану комп'ютерної системи в задачі інформаційної безпеки; методика виявлення DoS-атак в інформаційно-комунікаційних мережах з використанням нейронних мереж; нечітка маршрутизація пакетного трафіку з використанням мурашиного алгоритму навчання; нечіткі алгоритми IP-маршрутизації в телекомунікаційних мережах; придушення шумів сигналів мобільного зв'язку з нейронечіткою ідентифікацією тракту передачі; модифіковані мурашині алгоритми оптимізації у бездротових сенсорних мережах.

Контрольні запитання, що наведені в кінці кожної глави, спрямовані на закріплення матеріалу та поглиблене його осмислення.

В основу підручника покладені лекції, що читаються авторами в Національному технічному університеті «Дніпровська політехніка». Глави 1-4 написані к.т.н., доцентом Герасіною О.В., глави 5-7 – к.ф.-м.н., професором Гусевим О.Ю., глави 8-10, а також загальна редакція – д.т.н., професором Корнієнком В.І.

Метою підручника є формування компетентностей та вмінь застосування інтелектуальних методів обробки інформації та моделювання складних нелінійних динамічних процесів в системах різної природи та призначення.

Підручник призначений для підготовки магістрів денної і заочної форм навчання спеціальностей 125 «Кібербезпека» і 172 «Телекомунікації та радіотехніка» та відповідає програмам дисциплін «Інтелектуальні системи кіберзахисту», «Моделювання систем» і «Сучасні методи обробки інформації».

Підручник може бути корисним для підготовки бакалаврів та магістрів спеціальностей галузей знань 12 «Інформаційні технології», 15 «Автоматизація та приладобудування» та 17 «Електроніка та телекомунікації».

ЧАСТИНА І.

ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ОБРОБКИ ІНФОРМАЦІЇ

Перша частина цієї книги передує викладу основного матеріалу книги, присвяченій власне інтелектуальному моделюванню нелінійних динамічних процесів в керуванні, кібербезпеці та телекомунікаціях.

Термін *інтелект* (intelligence) походить від латинського intellectus – що означає розум, розсудок, розумові здібності людини. Відповідно *штучний інтелект* (artificial intelligence) (ШІ) – властивість автоматичних систем брати на себе окремі функції інтелекту людини, наприклад, обирати і приймати оптимальні рішення на основі раніше отриманого досвіду і раціонального аналізу зовнішніх впливів.

Наразі існують різні підходи до побудови інтелектуальних систем (або систем ШІ). Цей розподіл не є історичним, коли одна думка поступово змінює іншу, і різні підходи існують і зараз. Крім того, оскільки по-справжньому повних інтелектуальних систем в даний час немає (теорія створення систем ШІ інтенсивно створюється), то наразі не можна сказати остаточно, що якийсь підхід є більш правильним, а якийсь більш помилковим.

Під *структурним підходом* мається на увазі спроби побудови інтелектуальних систем шляхом моделювання структури людського мозку. Однією з перших таких спроб був перцептрон Ф. Розенблатта. Основною структурною одиницею, що моделюється, в перцептронах (як і в більшості інших варіантів моделювання мозку) є нейрон.

Пізніше виникли й інші моделі, які наразі відомі під терміном «*нейронні мережі*» (НМ) (вони розглянуті у главі 1). Ці моделі розрізняються за будовою окремих нейронів, за топологією зв'язків між ними і за алгоритмами навчання. Серед найбільш відомих варіантів НМ можна назвати мережі прямого розповсюдження, НМ Хопфілда, НМ зі зворотним поширення похибки тощо.

Для моделей, побудованих за мотивами людського мозку характерна більша виразність, легке розпаралелювання алгоритмів, а також пов'язана з цим

висока продуктивність паралельно реалізованих НМ. НМ найбільш успішно застосовуються в задачах розпізнавання образів, в тому числі сильно зашумлених, а також інших систем, про що більш докладно розглянуто в розділі 1.1.2.

Логічний підхід виник тому, що саме здатність до логічного мислення дуже сильно відрізняє людину від тварин. Основою для логічного підходу служить булева алгебра. Кожен програміст знайомий з нею і з логічними операторами (наприклад, оператор *if ... else*). Свій подальший розвиток булева алгебра отримала у вигляді числення предикатів, в якому вона розширена за рахунок введення предметних символів, відносин між ними, кванторів існування та загальності. Практично кожна система ШІ, що побудована на логічному принципі, являє собою машину доведення теорем. При цьому вихідні дані зберігаються в базі даних у вигляді аксіом, а правила логічного висновку – як відношення між ними. Крім того, кожна така машина має блок генерації цілей, і система виводу намагається довести дану мету як теорему. Якщо мета доведена, то трасування застосованих правил дозволяє отримати ланцюжок дій, необхідних для реалізації поставленої мети. Потужність такої системи визначається можливостями генератора цілей і машиною доведення теорем.

Домогтися більшої виразності логічного підходу дозволяє такий напрям, як *нечітка логіка*, яка розглянута у главі 2. Основною її відмінністю є те, що правдивість висловлювання може приймати в ній крім «так»/«ні» (1/0) ще й проміжні значення – не знаю (0,5), пацієнт швидше живий, ніж мертвий (0,75), пацієнт швидше мертвий, ніж живий (0,25). Даний підхід більше схожий на мислення людини, оскільки він на питання рідко відповідає тільки так чи ні. Хоча в окремих випадках, наприклад при тестуванні знань, приймаються якраз відповіді з розряду класичної булевої алгебри.

Для більшості логічних методів характерна велика трудомісткість, оскільки під час пошуку доказу можливий повний перебір варіантів. Тому даний підхід вимагає ефективної реалізації обчислювального процесу, і хороша робота зазвичай гарантується при порівняно невеликому розмірі бази даних.

При побудові інтелектуальних систем за *еволюційним підходом* основна увага приділяється побудові початкової моделі і правилам, за якими вона може змінюватися (еволюціонувати). Причому модель може бути складена за допомогою найрізноманітніших методів (НМ, алгоритми нечіткої логіки тощо). Далі система, побудована за еволюційним підходом, на підставі перевірки моделей відбирає найкращі з них (за певним критерієм), на підставі яких за самими різними правилами генеруються нові моделі, з яких знову обираються найкращі й так далі.

Взагалі, еволюційних моделей як таких не існує, існують тільки еволюційні алгоритми навчання, але моделі, які отримані при еволюційному підході мають деякі характерні особливості, що дозволяє виділити їх в окремий клас. Такими особливостями є перенесення основної роботи розробника з побудови моделі на алгоритм її модифікації, а також те, що отримані моделі практично не надають нових знань про середовище, що оточує систему ШІ, тобто вона стає як би річчю в собі.

Прикладом системи ШІ, побудованої за еволюційним підходом є *генетичні алгоритми, еволюційні алгоритми, еволюційні стратегії*, які розглянуті в главі 3.

Ще один широко використовуваний підхід до побудови інтелектуальних систем – *імітаційний*. Даний підхід є класичним для кібернетики і пов'язаний з одним з її базових понять – «чорним ящиком». *Чорний ящик* – це пристрій, програмний модуль або набір даних, інформація про внутрішню структуру та зміст якого повністю відсутня, але відомі специфікації вхідних і вихідних даних. Об'єкт, поведінка якого імітується, якраз і являє собою такий «чорний ящик». Нам не важливо, що у нього і у моделі всередині і як він функціонує, головне, щоб наша модель в аналогічних ситуаціях вела себе точно так само.

Таким чином, при імітаційному підході моделюється інша властивість людини – здатність копіювати те, що роблять інші, не вдаючись у подробиці, навіщо це потрібно. Найчастіше ця здатність економить йому масу часу, особливо на початку його життя.

Основним недоліком імітаційного підходу також є низька інформаційна здатність більшості моделей, побудованих з його допомогою.

Варто зазначити, що на практиці дуже чіткої межі між різними підходами до побудови систем ІІІ немає. Дуже часто зустрічаються змішані системи, де частина роботи виконується за одним підходом, а частина – за іншим.

Глава 1. НЕЙРОННІ МЕРЕЖІ

1.1. Основи штучних нейронних мереж

1.1.1. Визначення.

Під *нейронними мережами* (НМ) маються на увазі обчислювальні структури, які моделюють прості біологічні процеси, що зазвичай асоціюються з процесами людського мозку. НМ, що адаптуються і навчаються, представляють собою розпаралелені системи, здатні до навчання шляхом аналізу позитивних й негативних впливів. Елементним перетворювачем в даних мережах є штучний нейрон або просто нейрон, названий так за аналогією із біологічним прототипом.

Наразі запропоновано і вивчено велика кількість моделей нейроподібних елементів і НМ, ряд з яких розглянуто в цій главі.

Термін «нейронні мережі» сформувався в 40 роках ХХ століття в середовищі дослідників, які вивчали принципи організації і функціонування біологічних НМ. Основні результати, отримані в цій області, пов'язані з іменами американських дослідників У. Маккалока, У. Піттса, Ф. Розенблатта, Д. Хебба, Дж. Хопфілда та інших.

1.1.2. Області використання нейронних мереж.

Класифікація образів. Завдання полягає у вказівці належності вхідного образу, представленого вектором ознак, одному або декільком попередньо визначеним класам. До відомих додатків відносяться розпізнавання букв, розпізнавання мови, класифікація сигналу електрокардіограми, класифікація клітин крові тощо.

Кластеризація / категоризація. При вирішенні задачі кластеризації, яка відома також як класифікація образів «без вчителя», відсутня навчальна вибірка з мітками класів. Алгоритм кластеризації заснований на подібності образів, й розміщує близькі образи в один кластер. Відомі випадки застосування кластериза-

ції для отримання знань, стиснення даних, дослідження властивостей даних тощо.

Апроксимація функцій. Припустимо, що існує навчальна вибірка $((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N))$ (пари даних вхід-вихід), яка генерується невідомою функцією $F(x)$, що спотворена шумом. Завдання апроксимації полягає в знаходженні оцінки невідомої функції $F(x)$. Апроксимація функцій необхідна при вирішенні численних інженерних й наукових задач моделювання.

Передбачення / прогноз. Нехай задані n дискретних відліків $\{y(t_1), y(t_2), \dots, y(t_k)\}$ в послідовні моменти часу $t_1, t_2, \dots, t_k; k = \overline{1, n}$. Завдання полягає в передбаченні значення $y(t_{k+1})$ в деякий майбутній момент часу t_{k+1} . Передбачення / прогноз має значний вплив на прийняття рішень в науці, техніці та бізнесі. Типовими додатками техніки передбачення / прогнозу є передбачення цін на фондовій біржі, прогноз погоди тощо.

Оптимізація. Численні проблеми в математиці, статистиці, техніці, науці, медицині та економіці можуть розглядатися як проблеми оптимізації. Завданням алгоритму оптимізації є знаходження такого рішення, яке задовольняє системі обмежень й максимізує або мінімізує цільову функцію. Класичним прикладом завдання оптимізації є відома задача комівояжера.

Пам'ять, що адресується за змістом / асоціативна пам'ять. У моделі обчислень фон Неймана звернення до пам'яті доступно тільки за допомогою адреси, яка не залежить від змісту пам'яті. Більш того, якщо допущена помилка в обчисленні адреси, то може бути знайдена зовсім інша інформація. Асоціативна пам'ять доступна за вказівкою заданого змісту. Вміст пам'яті може бути викликано навіть по частковому входу або спотвореного змісту. Асоціативна пам'ять надзвичайно бажана при створенні мультимедійних інформаційних баз даних.

Керування. Розглянемо динамічну систему, задану сукупністю $\{u(t), y(t)\}$, де $u(t)$ є вхідним керуючим впливом, а $y(t)$ – виходом системи в момент часу t . В системах керування з еталонною моделлю метою керування є розрахунок та-

кого вхідного впливу $u(t)$, при якому система прямує по бажаній траєкторії, що диктується еталонною моделлю.

1.1.3. Біологічний нейрон.

Нервова система людини, побудована з елементів, званих нейронами, і має приголомшуючу складність. Біля 10^{11} нейронів беруть участь в приблизно 10^{15} передавальних зв'язках, що мають довжину метр і більше. Кожен нейрон володіє якостями, спільними з іншими елементами тіла, але його унікальною здатністю є приймання, обробка і передача електрохімічних сигналів по нервових шляхах, які утворюють комунікаційну систему мозку.

Нейрон (від латинського «neuron», дав.-грецького «νεῦρον» – «волокно», «нерв») є особливою біологічною клітиною, яка обробляє інформацію (рис. 1.1). Термін «нейрон» для позначення нервових клітин введений Г. В. Вальдейером у 1891 році.

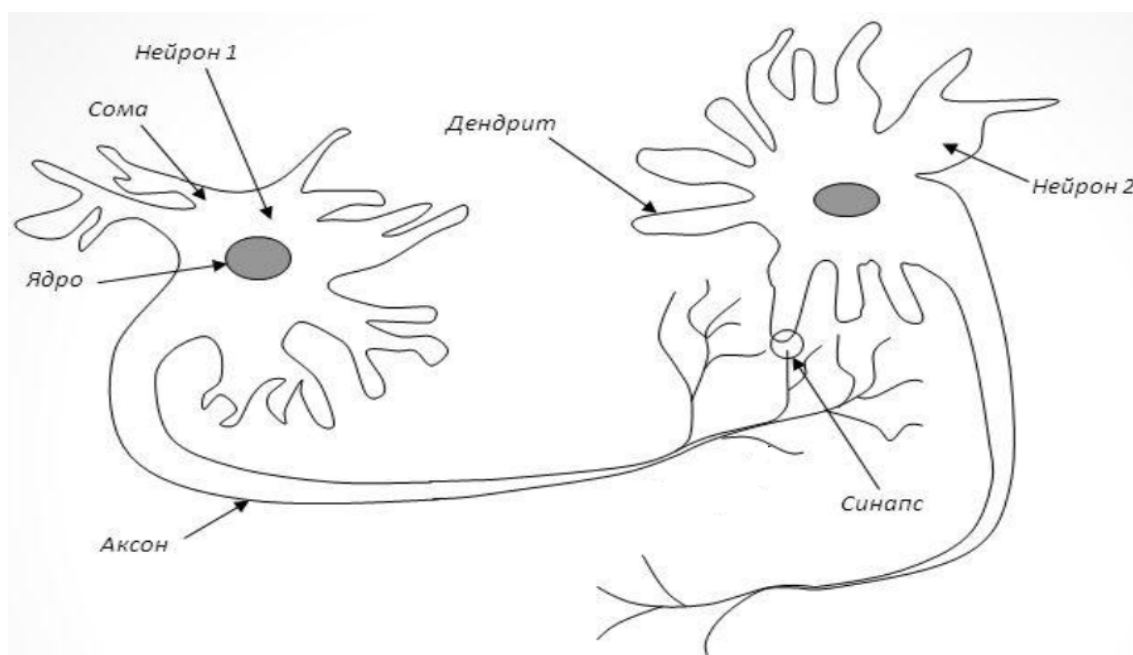


Рис. 1.1. Структура пари типових біологічних нейронів

Типовий нейрон складається з *тіла* (соми), а також відростків нервових волокон двох типів – *дендритів* (від грецького «déndron» – дерево), по яким

приймаються імпульси, і єдиного *аксона* (від грецького «ἄξων» – вісь), по якому нейрон може передавати імпульс. Тіло нейрона включає *ядро*, яке містить інформацію про спадкові властивості, та *плазму*, що володіє молекулярними засобами для виробництва необхідних нейрону матеріалів. Нейрон отримує сигнали (імпульси) від аксонів інших нейронів через дендрити (приймачі) і передає сигнали, що генеровані тілом клітки, вздовж свого аксона (передавача), який в кінці розгалужується на волокна. На кінчиках цих волокон знаходяться спеціальні утворення – *синапси* (від грецького «synapsis» – з'єднання), які впливають на величину імпульсів.

Синапс є елементарною структурою і функціональним вузлом між двома нейронами (волокно аксона одного нейрона і дендрит іншого). Коли імпульс досягає синапсичного закінчення, вивільняються хімічні речовини, звані нейротрансмітерами. Нейротрансмітери дифундують через синапсичну щілину, збуджуючи або гальмуючи, в залежності від типу синапсу, здатність нейрона-приймача генерувати електричні імпульси. Результативність синапсу може налаштовуватися сигналами, що проходять через нього, отже синапси можуть навчатися в залежності від активності процесів, в яких вони беруть участь. Ця залежність від передісторії діє як пам'ять, яка, можливо, відповідальна за пам'ять людини. Важливо відзначити, що ваги синапсів можуть змінюватися з часом, а отже, змінюється й поведінка відповідних нейронів.

Нейрони з'єднуються один з одним, формуючи біологічні НМ.

1.1.4. Історія розвитку штучних нейронних мереж.

1943 р. – У. Маккалок та У. Піттс формалізують поняття НМ у фундаментальній статті про логічне обчислення ідей і нервової активності.

1948 р. – Норберт Вінер публікує роботу про кібернетику. Основною ідеєю є представлення складних біологічних процесів математичними моделями.

1949 рік – Д. Хебб пропонує перший алгоритм навчання.

1958 р. – Ф. Розенблатт винаходить перцептрон. Перцептрон набуває популярності – його використовують для розпізнавання образів, прогнозування погоди тощо.

1960 р. – Б. Уїдроу спільно зі своїм студентом Хоффом на основі дельта-правила розробили адаптивний суматор Adaline, який відразу почав використовуватися для завдань прогнозування і адаптивного управління. Зараз Adaline є стандартним елементом багатьох систем обробки сигналів.

У 1961 р. під керівництвом М.М. Бонгарда розроблено програму «Кора» (комбінаторного розпізнавання), що реалізує алгоритм класифікації за методом зваженого голосування правил.

У 1969 р. М. Мінський публікує формальний доказ обмеженості перцептрону і показує, що він нездатний вирішувати деякі завдання, пов'язані з інваріантністю уявлень. Інтерес до НМ різко спадає.

1974 р. – Пол Дж. Вербос, і А.І. Галушкін одночасно винаходять алгоритм зворотного поширення похибки для навчання багаточарових перцептронів.

У 1975 р. Фукусіма представляє когнітрон – мережу, яка само організовується та призначена для інваріантного розпізнавання образів, але це досягається тільки за допомогою запам'ятовування практично усіх станів образу.

1982 р. – Дж. Хопфілд показав, що НМ із зворотними зв'язками може бути системою, що мінімізує енергію (так звана мережа Хопфілда). Т. Кохоненом представлені моделі мережі, що навчається без учителя (НМ Кохонена).

1986 р. – Д. Румельхартом, Дж. Хінтоном і Р. Вільямсом, а також незалежно і одночасно С.І. Барцевим та В.О. Охоніним пере відкритий та істотно розвинений метод зворотного поширення похибки. Почався вибух інтересу до НМ.

1.1.5. Структура та властивості штучного нейрона. Функції активації.

Штучний нейрон складається з елементів трьох типів: помножувачів (синапсів), суматора і нелінійного перетворювача (рис. 1.2). Синапси здійснюють зв'язок між нейронами, помножують вхідний сигнал на число, що характеризує силу зв'язку – вагу синапсу. Суматор виконує додавання сигналів, що надходять

по синапсичним зв'язкам від інших нейронів, та зовнішніх вхідних сигналів. Нелінійний перетворювач реалізує нелінійну функцію одного аргументу – виходу суматора, яка називається *функцією активації* або передавальною функцією нейрона.

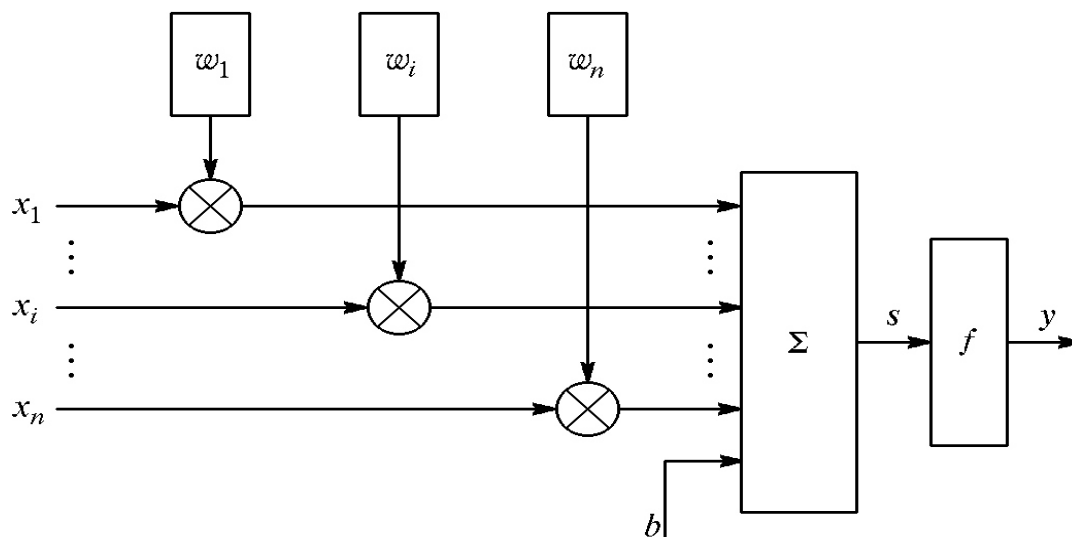


Рис. 1.2. Структура штучного нейрона

Нейрон в цілому реалізує скалярну функцію векторного аргументу. Математична модель нейрона описується співвідношенням:

$$s = \sum_{i=1}^n w_i x_i + b ; \quad (1.1)$$

$$y = f(s) ;$$

де w_i – вага синапсу ($i=1\dots n$); b – значення зміщення; s – результат підсумовування; x_i – компонент вхідного вектора (вхідний сигнал) ($i=1\dots n$); y – вихідний сигнал нейрона; n – число входів нейрона; f – нелінійне перетворення (функція активації або передаточна функція).

У загальному випадку вхідний сигнал, вагові коефіцієнти і зміщення можуть приймати дійсні значення, а в багатьох практичних завданнях – лише деякі фіксовані значення. Вихід y визначається видом функції активації та може бути як дійсним, так і цілим.

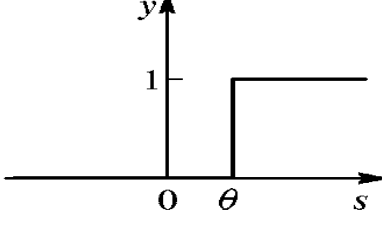
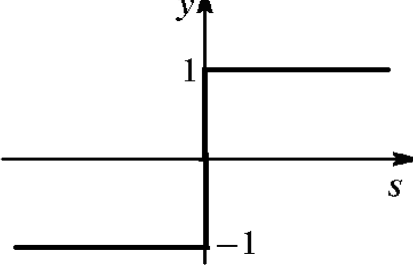
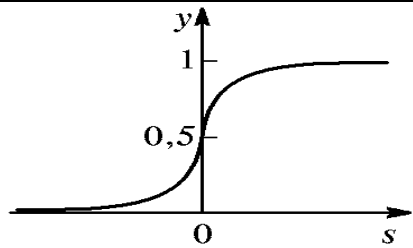
Синапсичні зв'язки з позитивними вагами називають збуджуючими, а з негативними – гальмуючими.

Описаний обчислювальний елемент можна вважати спрощеною математичною моделлю біологічних нейронів. Щоб підкреслити відмінність нейронів біологічних і штучних, другі іноді називають нейроноподібними елементами або формальними нейронами.

На вхідний сигнал s нелінійний перетворювач відповідає вихідним сигналом $f(s)$, який представляє собою вихід у нейрона. Приклади найбільш поширених активаційних функцій представлені в табл. 1.1.

Функції активації нейронів

Таблиця 1.1

Назва	Формула	Графік	Область значень
Порогова	$f(s) = \begin{cases} 0, & s < \theta, \\ 1, & s \geq \theta \end{cases}$		(0,1)
Знакова (сигнатурна)	$f(s) = \begin{cases} 1, & s > 0, \\ -1, & s \leq 0 \end{cases}$		(-1,1)
Логістична сигмоїдальна	$f(s) = \frac{1}{1 + e^{-s}}$		(0,1)

Назва	Формула	Графік	Область значень
Гіперболічний тангенс	$f(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}}$		(-1,1)
Радіальна базисна (гаусова)	$f(s) = e^{-s^2}$		(0,1)
Лінійна	$f(s) = s$		$(-\infty, \infty)$
Напівлінійна	$f(s) = \begin{cases} s, & s > 0, \\ 0, & s \leq 0 \end{cases}$		$(0, \infty)$
Трикутна	$f(s) = \begin{cases} 1 - s , & s \leq 1, \\ 0, & s > 1 \end{cases}$		(0,1)
Напівлінійна з насиченням	$f(s) = \begin{cases} 0, & s \leq 0, \\ s, & 0 < s < 1, \\ 1, & s \geq 1 \end{cases}$		(0,1)

Назва	Формула	Графік	Область значень
Лінійна з насиченням	$f(s) = \begin{cases} -1, & s \leq -1, \\ s, & -1 < s < 1, \\ 1, & s \geq 1 \end{cases}$		(-1,1)

Вибір функції активації визначається: специфікою завдання, зручністю реалізації на комп'ютері, а також алгоритмом навчання, оскільки деякі алгоритми накладають обмеження на вид функції активації.

НМ притаманний принцип паралельної обробки сигналів, який досягається шляхом об'єднання великого числа нейронів в так звані шари і з'єднання певним чином нейронів різних шарів, а також (в деяких конфігураціях) – і нейронів одного шару між собою, причому обробка взаємодії всіх нейронів ведеться пошарово.

1.1.6. Класифікація нейронних мереж.

НМ являє собою сукупність нейроподібних елементів, певним чином з'єднаних один з одним і з зовнішнім середовищем за допомогою зв'язків, що визначаються ваговими коефіцієнтами.

Побудова НМ вирішується в два етапи:

1. *Вибір типу (архітектури) мережі.* Не обов'язково будувати НМ «з нуля», існує декілька десятків різних нейромережових архітектур, ефективність яких доказана математично.

2. *Навчання мережі,* тобто підбір таких значень ваг, щоб мережа працювала належним чином. В використовуваних на практиці НМ кількість ваг може становити десятки тисяч, тому навчання – дійсно складний процес. Для багатьох архітектур розроблені спеціальні алгоритми навчання, які дозволяють налаштувати ваги мережі певним чином.

В залежності від функцій, які виконуються нейронами в мережі, можна виділити їх три типи:

1. *Вхідні нейрони*, на які подається вектор, що кодує вхідний вплив або образ зовнішнього середовища. В цих нейронах зазвичай не здійснюються обчислювальні процедури, а інформація передається з входу на вихід нейрона шляхом зміни його активації.

2. *Вихідні нейрони*, вихідні значення яких представляють виходи НМ.

3. *Проміжні нейрони*, що становлять основу НМ.

У більшості нейронних моделей тип нейрону пов'язаний із його розташуванням в мережі. Якщо нейрон має тільки вхідні зв'язки, то це вхідний нейрон, якщо навпаки – вихідний нейрон. Однак можливий випадок, коли вихід топологічно внутрішнього нейрону розглядається як частина виходу мережі. В процесі функціонування мережі здійснюється перетворення вхідного вектора у вихідний, деяка обробка інформації. Конкретний вид виконуваного мережею перетворення даних обумовлюється не тільки характеристиками нейроподібних елементів, а й особливостями її архітектури, а саме топологією зв'язків між нейронами, вибором певних підмножин нейроподібних елементів для введення і виведення інформації, способами навчання мережі, наявністю або відсутністю конкуренції між нейронами, напрямком і способами керування і синхронізації передачі інформації між нейронами.

1.1.7. Топологія нейронних мереж.

З точки зору топології можна виділити три основні типи НМ:

- повно зв'язані (рис. 1.3,*а*);
- багат шарові (рис. 1.3,*б*);
- слабо зв'язані (з локальними зв'язками) (рис. 1.3,*в*).

У повно зв'язаних НМ (див. рис. 1.3,*а*) кожен нейрон передає свій вихідний сигнал іншим нейронам, у тому числі й самому собі. Усі вхідні сигнали подаються усім нейронам. Вихідними сигналами мережі можуть бути усі або деякі вихідні сигнали нейронів після кількох тактів функціонування мережі.

У багатошарових НМ (див. рис. 1.3,б) нейрони об'єднуються у шари. Шар містить сукупність нейронів з єдиними вхідними сигналами. Число нейронів у шарі може бути будь-яким і не залежить від кількості нейронів в інших шарах. У загальному випадку мережа складається з множини шарів, пронумерованих зліва направо. Зовнішні вхідні сигнали подаються на входи нейронів вхідного шару (його часто нумерують як нульовий), а виходами мережі є вихідні сигнали останнього шару. Крім вхідного і вихідного шарів в багатошаровій НМ є один або кілька прихованих шарів. Зв'язки від виходів нейронів деякого шару q до входів нейронів наступного шару ($q+1$) називаються послідовними.

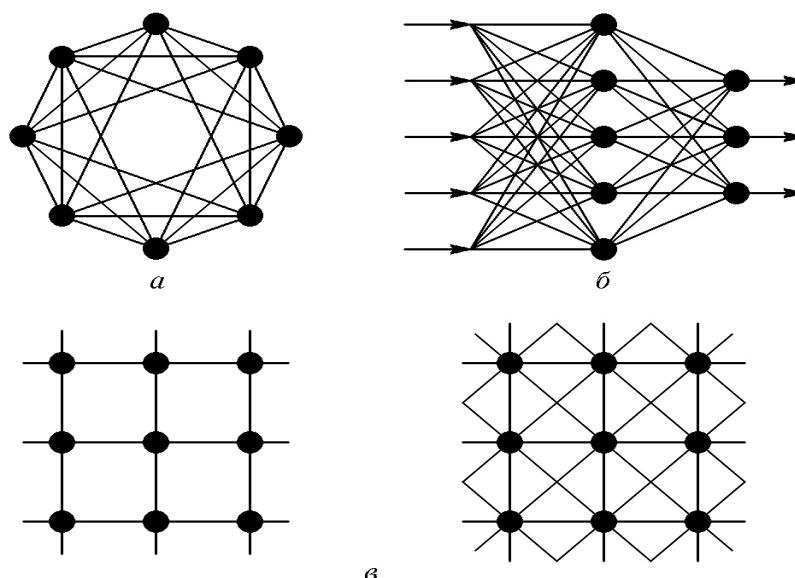


Рис. 1.3. Топологія НМ: *a* – повно зв'язана мережа, *б* – багатошарова НМ з послідовними зв'язками, *в* – слабо зв'язані мережі

Серед багатошарових НМ (див. рис. 1.3,б) виділяють наступні типи.

1. *Монотонні НМ*. Це окремий випадок багатошарових мереж з додатковими умовами на зв'язки і нейрони. Кожен шар крім останнього (вихідного) розбитий на два блоки: збуджуючий і гальмуючий. Зв'язки між блоками теж поділяються на гальмуючі і збуджуючі. Якщо від нейронів блоку *A* до нейронів блоку *B* ведуть тільки збуджуючі зв'язки, то це означає, що будь-який вихідний сигнал блоку є монотонною не спадаючою функцією будь-якого вихідного сиг-

налу блоку A . Якщо ж ці зв'язки тільки гальмуючі, то будь-який вихідний сигнал блоку B є не зростаючою функцією будь-якого вихідного сигналу блоку A . Для нейронів монотонних мереж необхідна монотонна залежність вихідного сигналу нейрона від параметрів вхідних сигналів.

2. *НМ без зворотних зв'язків.* У таких мережах нейрони вхідного шару отримують вхідні сигнали, перетворюють їх і передають нейронам першого прихованого шару, і так далі аж до вихідного, який видає сигнали для інтерпретатора і користувача. Якщо не визначено інше, то кожен вихідний сигнал шару q подається на вхід всіх нейронів шару $(q+1)$; проте можливий варіант сполюки шару q з довільним шаром $(q+p)$, $p \neq 1$.

Класичним варіантом НМ без зворотних зв'язків є мережі прямого розповсюдження (рис. 1.4).

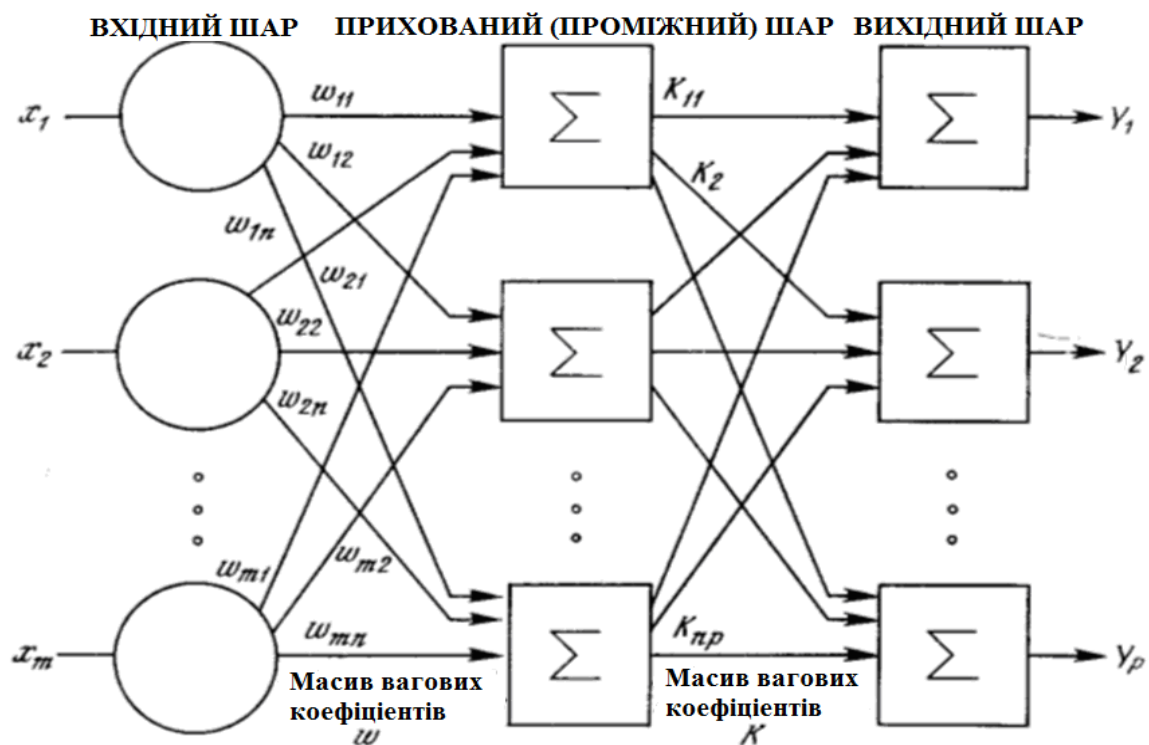


Рис. 1.4. Двошарова НМ прямого розповсюдження

3. *Мережі зі зворотними зв'язками.* У мережах із зворотними зв'язками інформація з наступних шарів передається на попередні.

Прикладом НМ зі зворотними зв'язками є частково-рекурентні мережі Елмана і Жордана. Такі мережі реалізують нелінійні моделі, які можуть бути застосовані для оптимального керування процесами, що змінюються за часом, тобто зворотній зв'язок дозволяє забезпечити адаптивне запам'ятовування минулих часових подій.

Мережа Елмана (рис. 1.5) відрізняється від НМ Жордана тим, що в ній сигнали на вхідний шар йдуть не від виходів мережі, а від виходів нейронів прихованого шару.

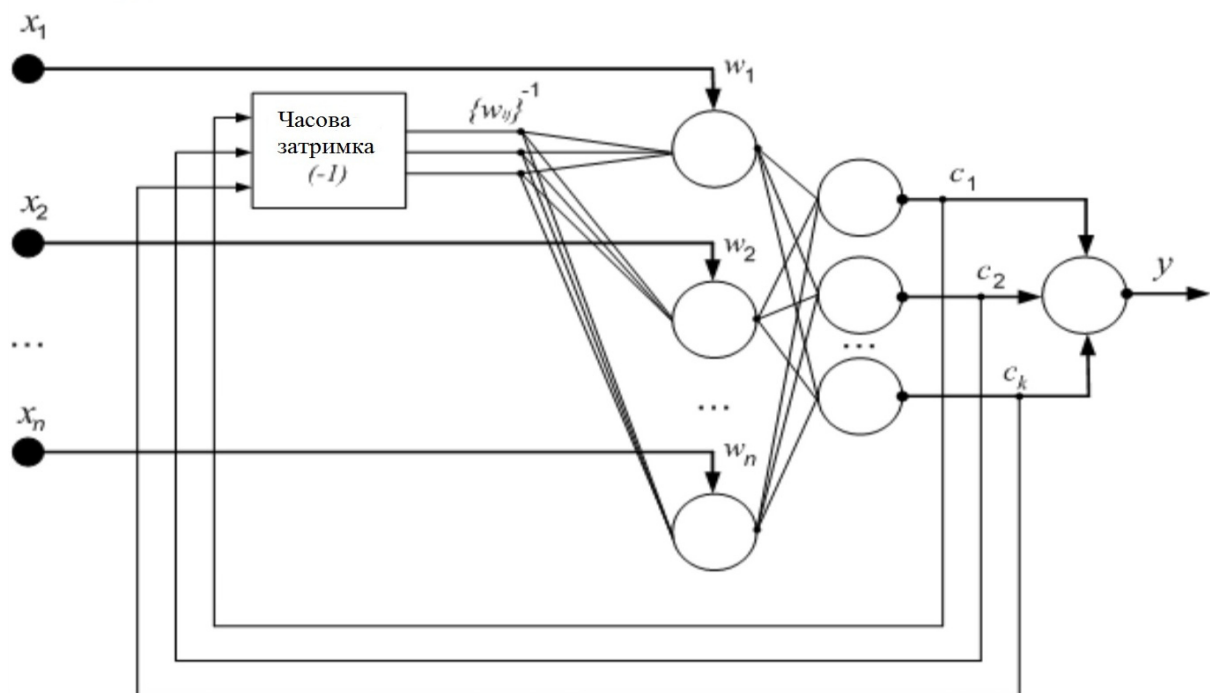


Рис. 1.5. НМ Елмана

У НМ Елмана (див. рис. 1.5) виходи прихованого шару $\{c_1, c_2, \dots, c_k\}$ подаються з часовою затримкою на вхідні нейрони з ваговими коефіцієнтами $\{w_{ij}\}^{-1}$, де i – індекс нейрона, на який подається сигнал ($i=1, 2, \dots, n$), j – індекс вихідного сигналу нейроне прихованого шару ($j=1, 2, \dots, k$).

В основі мережі Жордана лежить багатошаровий перцептрон. Зворотний зв'язок реалізується через подачу на вхідний шар не тільки вхідних даних, але й сигналів виходу мережі з затримкою на один або декілька тактів, що дозволяє

врахувати передісторію спостережених процесів і накопичити інформацію для вироблення раціональної стратегії керування. Також до НМ зі зворотними зв'язками належать НМ Хопфілда та Хемінга, які детально розглянуті в п. 1.3.4.

Взагалі немає чітко визначеної процедури для вибору кількості нейронів і кількості шарів в штучній НМ. Чим більше кількість нейронів і шарів, тим ширше можливості мережі, тим повільніше вона навчається і працює і тим більш нелінійною може бути залежність вхід - вихід.

Кількість нейронів і шарів пов'язано зі складністю завдання, із кількістю даних для навчання, із необхідною кількістю входів і виходів мережі, а також із наявними ресурсами: пам'яттю і швидкістю комп'ютера, на якому моделюється мережа.

Якщо в мережі занадто мало нейронів або шарів, то:

1) мережа недостатньо навчиться і похибка при роботі мережі залишиться великою;

2) на виході мережі не будуть передаватися високочастотні (різкі) коливання функції $y(x)$, що апроксимується.

Перевищення необхідної кількості нейронів теж заважає роботі мережі. Оскільки, якщо нейронів або шарів занадто багато, то:

1) швидкодія буде низькою, а пам'яті буде потрібно багато (на фон-нейманівських комп'ютерах);

2) НМ перенавчиться: вихідний вектор буде передавати незначні і несуттєві деталі в досліджуваній залежності $y(x)$, наприклад, шум або помилкові дані;

3) залежність виходу від входу виявиться різко нелінійною: вихідний вектор буде суттєво і непередбачувано змінюватися при малій зміні вхідного вектора x ;

4) мережа буде нездатна до узагальнення: в області, де немає або мало відомих точок функції $y(x)$ вихідний вектор буде випадковим і непередбачуваним, хоча і буде адекватним задачі, яка вирішується.

1.2. Навчання нейронних мереж.

1.2.1. Навчання із вчителем.

Коли йде мова про використання НМ, майже завжди мають на увазі певні процедури їх навчання. НМ є адаптивною системою, життєвий цикл якої складається з двох незалежних фаз – навчання і роботи мережі. Процес навчання НМ показано на рис. 1.6.



Рис. 1.6. Процес навчання НМ

Коли після багаторазового пред'явлення прикладів змінювані параметри (ваги) НМ стабілізуються, і вона дає правильні відповіді на всі (або майже всі) приклади з бази даних, то в такому випадку говорять, що «НМ навчена». У програмних реалізаціях можна бачити, що в процесі навчання величина похибки поступово зменшується. Коли величина похибки досягає нуля або прийняттого малого рівня, навчання зупиняють, а отриману НМ вважають навченою і готовою до застосування на нових даних.

Таким чином, навчання НМ вважається закінченим, коли вона правильно виконує перетворення на тестових прикладах і подальше навчання не викликає

значної зміни вагових коефіцієнтів. Далі мережа виконує перетворення раніше невідомих їй даних на основі сформованої нею в процесі навчання моделі процесу. Мережа успішно працює до тих пір, поки істотно не зміниться реальна модель відображуваного явища (наприклад, у разі виникнення ситуації, інформація про яку ніколи (або у достатньому обсязі) не пред'являлась мережі при навчанні). Після цього НМ може бути донавчена (адаптована) з урахуванням нової інформації, причому при донавчанні попередня інформація не губиться, а узагальнюється із новою.

Слід зазначити, що вся інформація, яку НМ має про завдання, міститься в наборі прикладів. Тому якість навчання мережі безпосередньо залежить від кількості прикладів в навчальній вибірці, а також від того, наскільки повно ці приклади описують це завдання. Вважається, що для повноцінного тренування НМ потрібно хоча б кілька десятків (а краще сотень) прикладів. Отже навчання НМ є складним і наукомістким процесом, і від того, наскільки якісно він буде виконаний, залежить здатність НМ вирішувати поставлені перед нею завдання під час експлуатації.

Математично процес навчання НМ можна описати наступним чином.

В процесі функціонування НМ формує вихідний сигнал Y , реалізуючи деяку функцію $Y=G(X)$. Якщо архітектура мережі задана, то вид функції G визначається значеннями синапсичних ваг і зміщень мережі.

Нехай рішенням деякої задачі є функція $Y=F(X)$, задана параметрами вхідних-вихідних даних $(X^1, Y^1), (X^2, Y^2), \dots, (X^N, Y^N)$, для яких $Y^k=F(X^k)$ ($k=1, 2, \dots, N$).

Навчання полягає в пошуку (синтезі) функції G , близькою до F в сенсі деякої функції похибки E (див. рис. 1.6).

Якщо обрано множину навчальних прикладів – пар (X^k, Y^k) (де $k=1, 2, \dots, N$) і спосіб обчислення функції похибки E , то навчання НМ перетворюється в задачу багатовимірної оптимізації, що має дуже велику розмірність. При цьому, оскільки функція E може мати довільний вигляд, навчання у загальному випадку – багато екстремальна (полімодальна) не опукла задача оптимізації.

Для вирішення цієї задачі можуть використовуватись такі ітераційні алгоритми.

1. Алгоритми локальної оптимізації з обчисленням частинних похідних першого порядку: градієнтний алгоритм або метод найшвидшого спуску, методи з одновимірною і двовимірною оптимізацією цільової функції в напрямку антиградієнту, метод сполучених градієнтів, тощо.

2. Алгоритми локальної оптимізації з обчисленням частинних похідних першого і другого порядку: метод Ньютона, методи оптимізації з розрідженими матрицями Гессе, квазіньютонівські методи, метод Гауса-Ньютона, метод Левенберга-Марквардта, тощо.

3. Стохастичні алгоритми оптимізації: пошук у випадковому напрямку, імітація відпалу, метод Монте-Карло, тощо.

4. Алгоритми глобальної оптимізації, що вирішуються за допомогою перебору значень змінних, від яких залежить цільова функція (функція похибки).

1.2.2. Алгоритм зворотного поширення похибки.

Це ітеративний градієнтний алгоритм навчання, який використовується з метою мінімізації середньоквадратичного відхилення поточного виходу і бажаного виходу багатосарових НМ.

Цей алгоритм використовується для навчання багатосарових НМ з послідовними зв'язками (див. рис. 1.4). Нейрони в таких мережах діляться на групи (шари) із загальним вхідним сигналом. На кожен нейрон першого шару подаються всі елементи зовнішнього вхідного сигналу, а усі виходи нейронів поточного шару подаються на кожен нейрон наступного шару. Нейрони виконують зважене підсумовування елементів вхідних сигналів. До суми елементів вхідних сигналів, помножених на відповідні синапсичні ваги, додається зсув нейрона. Над результатом підсумовування виконується нелінійне перетворення – функція активації (передавальна функція), значення якої і є виходом нейрона.

В багатосарових мережах оптимальні вихідні значення нейронів усіх шарів, крім останнього, як правило, невідомі, і НМ з трьома або більше шарами

вже неможливо навчити, керуючись тільки величинами похибок на її виходах. Найбільш прийнятним варіантом навчання в таких умовах є градієнтний метод пошуку мінімуму функції похибки з розглядом сигналів похибки від виходів до входів НМ, який й отримав назву *алгоритму зворотного поширення похибки*.

В даному алгоритмі цільовою функцією похибки, що мінімізується є величина:

$$E(w) = \frac{1}{2} \sum_{j,p} (y_{j,p}^{(n)} - d_{j,p})^2 \quad (1.2)$$

де $y_{j,p}^{(n)}$ – реальний вихідний стан нейрону j вихідного шару n НМ при подачі на її входи образу p ; $d_{j,p}$ – ідеальний (бажаний) вихідний стан цього нейрону; $n = \overline{1, N-1}$, $N-1$ – кількість прихованих шарів НМ.

Мінімізація ведеться методом градієнтного спуску, що означає підстроювання вагових коефіцієнтів наступним чином

$$\Delta w_{ij}^{(n)} = -\eta \cdot \frac{\partial E}{\partial w_{ij}}, \quad (1.3)$$

де w_{ij} – ваговий коефіцієнт синапсичного зв'язку, що з'єднує нейрон i шару $n-1$ з нейроном j шару n , η – коефіцієнт швидкості навчання, $0 < \eta < 1$.

Відповідно до правила диференціювання складної функції:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j} \cdot \frac{\partial s_j}{\partial w_{ij}}, \quad (1.4)$$

де s_j – зважена сума вхідних сигналів нейрону j , тобто аргумент активаційної функції.

Оскільки dy_j/ds_j є похідною функції активації по її аргументу, то з цього випливає, що похідна повинна бути визначена на всій осі абсцис. У зв'язку з цим активаційні функції з неоднорідностями (наприклад порогова) не підходять для розглянутих НМ. У них застосовуються такі гладкі функції активації, як гіперболічний тангенс або класичний сигмоїд з експонентою (табл. 1.1). У разі гіперболічного тангенса

$$\frac{dy}{ds} = 1 - s^2 \quad (1.5)$$

Третій множник $\delta_j / \partial w_{ij}$ в (1.4) дорівнює виходу нейрона попереднього шару $y_i^{(n-1)}$. А перший множник в (1.4) легко розкладається наступним чином:

$$\frac{\partial E}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot \frac{\partial s_k}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot w_{jk}^{(n+1)} \quad (1.6)$$

Тут підсумовування по k виконується серед нейронів шару $n+1$.

Шляхом введення нової змінної

$$\delta_j^{(n)} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j} \quad (1.7)$$

було отримано рекурсивну формулу для розрахунків величин $\delta_j^{(n)}$ шару n з величин $\delta_k^{(n+1)}$ більш старшого шару $n+1$

$$\delta_j^{(n)} = \left[\sum_k \delta_k^{(n+1)} \cdot w_{jk}^{(n+1)} \right] \cdot \frac{dy_j}{ds_j}, \quad (1.8)$$

а для вихідного шару l

$$\delta_l^{(N)} = (y_l^{(N)} - d_l) \cdot \frac{dy_l}{ds_l}. \quad (1.9)$$

Тепер (1.3) має вигляд:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \delta_j^{(n)} \cdot y_i^{(n-1)} \quad (1.10)$$

Іноді для додання процесу корекції ваг НМ деякої інерційності, що згладжує різкі скачки при переміщенні по поверхні цільової функції, (1.10) доповнюється значенням зміни ваги на попередній ітерації

$$\Delta w_{ij}^{(n)}(t) = -\eta \cdot (\mu \cdot \Delta w_{ij}^{(n)}(t-1) + (1 - \mu) \cdot \delta_j^{(n)} \cdot y_i^{(n-1)}). \quad (1.11)$$

де μ – коефіцієнт інерційності, t – номер поточної ітерації.

Таким чином, повний алгоритм навчання НМ за допомогою процедури зворотного поширення містить наступні етапи:

1. Подати на входи мережі один з можливих прикладів (образів), і в режимі звичайного функціонування НМ, коли сигнали поширюються від входів до виходів, розрахувати значення останніх. При цьому

$$s_j^{(n)} = \sum_{i=0}^M y_i^{(n-1)} \cdot w_{ij}^{(n)}, \quad (1.12)$$

де M – число нейронів в шарі $k-1$ з урахуванням нейрону з постійним вихідним станом $+1$, що задає зміщення; $y_i^{(n-1)} = x_{ij}^{(n)}$ – вхід і нейрону j шару n .

$$y_j^{(n)} = f(s_j^{(n)}), \quad (1.13)$$

$$y_q^{(0)} = I_q, \quad (1.14)$$

де $f()$ – сигмоїд, I_q – компонента q вектору вхідного прикладу (образу).

2. Розрахувати $\delta^{(N)}$ для вихідного шару за формулою (1.9).

Розрахувати за формулою (1.10) або (1.11) зміни ваг $\Delta w^{(N)}$ шару N .

3. Розрахувати за формулами (1.8) і (1.10) (або (1.8) і (1.11)) відповідно $\delta^{(n)}$ і $\Delta w^{(n)}$ для усіх інших шарів, $n=N-1, \dots, 1$.

4. Скорегувати усі ваги в НМ

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t). \quad (1.15)$$

5. Якщо похибка НМ істотна, перейти до етапу 1. В іншому випадку – кінець алгоритму.

Алгоритм зворотного поширення похибки ілюструється рис. 1.7.

На етапі 1 на входи НМ поперемінно у випадковому порядку пред'являються всі тренувальні приклади (образи), щоб мережа, образно кажучи, не забувала одні при запам'ятовуванні інших.

1.2.3. Навчання нейронних мереж без вчителя.

Навчання без вчителя є більш правдоподібною моделлю поведінки в біологічній системі. Процес навчання, як і при навчанні з вчителем, полягає в підстроювання ваг синапсів НМ. Деякі алгоритми передбачають зміну і структури мережі, тобто кількості нейронів і їх взаємозв'язків, але такі перетворення більш точно назвати широким терміном – самоорганізацією. Очевидно, що під-

строювання синапсів може проводитись тільки на підставі інформації, яка доступна нейрону, тобто його стану і вже наявних вагових коефіцієнтів. Виходячи з цього міркування, й за аналогією з відомими принципами самоорганізації нервових клітин, побудовані алгоритми навчання Хебба.

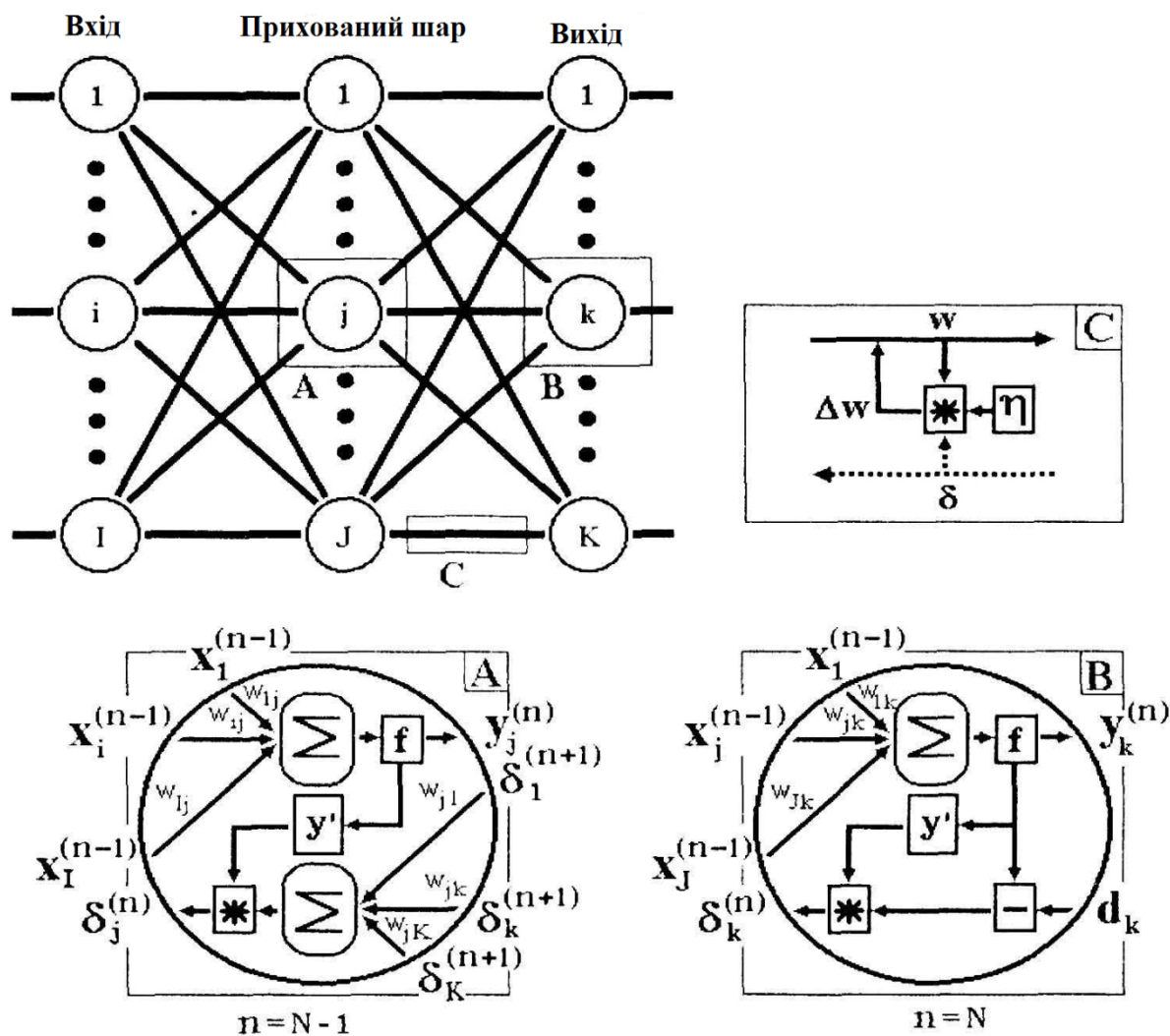


Рис. 1.7. Діаграма сигналів в НМ при навчанні за алгоритмом зворотного поширення похибки

Сигнальний метод навчання Хебба полягає в зміні ваг за наступним правилом:

$$w_{ij}(t) = w_{ij}(t-1) + \alpha \cdot y_i^{(n-1)} \cdot y_j^{(n)}, \quad (1.16)$$

де $y_i^{(n-1)}$ – вихідне значення нейрона i шару $(n-1)$; $y_j^{(n)}$ – вихідне значення нейрона j шару n ; $w_{ij}(t)$ і $w_{ij}(t-1)$ – вагові коефіцієнти синапсів, що з'єднує ці нейрони, на ітераціях t і $t-1$ відповідно; α – коефіцієнт швидкості навчання. Тут і далі, для спільності, під n мається на увазі довільний шар мережі.

При навчанні за цим методом посилюються зв'язки між збудженими нейронами.

Існує також і диференційний метод навчання Хебба.

$$w_{ij}(t) = w_{ij}(t-1) + \alpha \cdot [y_i^{(n-1)}(t) - y_i^{(n-1)}(t-1)] \cdot [y_j^{(n)}(t) - y_j^{(n)}(t-1)], \quad (1.17)$$

де $y_i^{(n-1)}(t)$ і $y_i^{(n-1)}(t-1)$ – вихідне значення нейрона i шару $n-1$ відповідно на ітераціях t і $t-1$; $y_j^{(n)}(t)$ і $y_j^{(n)}(t-1)$ – те ж саме для нейрона j шару n .

Як видно з (1.17), найсильніше навчаються синапси, що з'єднують ті нейрони, виходи яких найбільш динамічно змінилися у бік збільшення.

Повний алгоритм навчання відповідно до методу Хебба може бути представлений у наступному вигляді:

1. На стадії ініціалізації всім ваговим коефіцієнтам привласнюються невеликі випадкові значення.

2. На входи мережі подається вхідний образ (приклад), і сигнали збудження поширюються по всім шарам згідно з принципами класичних НМ прямого розповсюдження, тобто для кожного нейрона розраховується зважена сума його входів, до якої потім застосовується активаційна (передавальна) функція, в результаті чого виходить його вихідне значення $y_i^{(n)}$, $i=0 \dots M_i-1$, де M_i – число нейронів в шарі i ; $n=0, \dots, N-1$, а N – число шарів в мережі.

3. На підставі отриманих вихідних значень нейронів за формулою (1.16) або (1.17) проводиться зміна вагових коефіцієнтів.

4. Цикл з кроку 2, поки вихідні значення мережі не застабілізуються із заданою точністю. Застосування цього нового способу визначення завершення навчання, відмінного від того, що використовувався для мережі зворотного розповсюдження, обумовлено тим, що значення синапсів, які підлаштовуються,

фактично не обмежені. На другому кроці циклу поперемінно пред'являються всі образи (приклади) із вхідного набору.

Слід зазначити, що вид відгуків на кожен клас вхідних образів заздалегідь невідомий і буде являти собою довільне поєднання станів нейронів вихідного шару, обумовлене випадковим розподілом ваг на стадії ініціалізації. Разом з тим, мережа здатна узагальнювати схожі образи, відносячи їх до одного класу. Тестування навченої мережі дозволяє визначити топологію класів у вихідному шарі. Для приведення відгуків навченої мережі до зручного подання можна доповнити НМ одним шаром, який, наприклад, за алгоритмом навчання одношарового перцептрону необхідно змусити відображати вихідні реакції мережі в необхідні образи.

Інший алгоритм навчання без вчителя – алгоритм Кохонена – передбачає підстроювання синапсів на підставі їх значень від попередньої ітерації:

$$w_{ij}(t) = w_{ij}(t-1) + \alpha \cdot [y_i^{(n-1)} - w_{ij}(t-1)]. \quad (1.18)$$

З (1.18) видно, що навчання зводиться до мінімізації різниці між вхідними сигналами нейрону, які надходять з виходів нейронів попереднього шару $y_i^{(n-1)}$, і ваговими коефіцієнтами його синапсів.

Повний алгоритм навчання має приблизно таку ж структуру, як в методах Хебба, але на кроці 3 з усього шару обирається нейрон, значення синапсів якого максимально схожі на вхідний образ, і підстроювання ваг по формулі (1.18) проводиться тільки для нього. Ця, так звана, акредитація може супроводжуватись загальмуванням усіх інших нейронів шару і введенням обраного нейрона в насичення. Вибір такого нейрона може здійснюватися, наприклад, розрахунком скалярного добутку вектору вагових коефіцієнтів з вектором вхідних значень. Максимальний добуток дає виграшний нейрон.

Інший варіант – розрахунок відстані між цими векторами в p -вимірному просторі:

$$D_j = \sqrt{\sum_{i=0}^{p-1} (y_i^{(n-1)} - w_{ij})^2}, \quad (1.19)$$

де p – розмір векторів, j – індекс нейрону в шарі n , i – індекс підсумовування по нейронах шару $(n-1)$, w_{ij} – вага синапсу, що з'єднує нейрони. При цьому виходи нейронів шару $(n-1)$ є вхідними значеннями для шару n . Корінь у (1.19) брати не обов'язково, оскільки важлива лише відносна оцінка різних D_j .

В даному випадку, «перемагає» нейрон з найменшою відстанню. Іноді нейрони, які занадто часто отримують акредитацію, примусово виключаються з розгляду, щоб «зрівняти права» всіх нейронів шару. Найпростіший варіант такого алгоритму полягає в гальмуванні нейрону, який тільки що виграв.

При використанні навчання за алгоритмом Кохонена існує практика нормалізації вхідних образів, а також – на стадії ініціалізації – і нормалізації початкових значень вагових коефіцієнтів

$$x_i = x_i / \sqrt{\sum_{j=0}^{n-1} x_j^2}, \quad (1.20)$$

де x_i – i -та компонента вектора вхідного образу або вектора вагових коефіцієнтів, n – його розмірність. Це дозволяє скоротити тривалість процесу навчання.

Ініціалізація вагових коефіцієнтів випадковими значеннями може привести до того, що різні класи, яким відповідають щільно розподілені вхідні образи, зіллються або, навпаки, подрібняться на додаткові підкласи у разі близьких образів одного й того ж класу. Для уникнення такої ситуації використовується метод опуклої комбінації. Суть його зводиться до того, що вхідні нормалізовані образи піддаються перетворенню:

$$x_i = \alpha(t)x_i + (1 - \alpha(t))\frac{1}{\sqrt{n}}, \quad (1.21)$$

де x_i – i -а компонента вхідного образу, n – загальне число його компонент, $\alpha(t)$ – коефіцієнт, що змінюється в процесі навчання від нуля до одиниці, в результаті чого спочатку на входи мережі подаються практично однакові образи, а з плином часу вони все більше сходяться до вихідних. Вагові коефіцієнти встановлюються на кроці ініціалізації, рівними величині

$$w_o = \frac{1}{\sqrt{n}}, \quad (1.22)$$

де n – розмірність вектора ваг для нейронів шару, який ініціалізується.

На основі розглянутого вище методу будуються НМ особливого типу – так звані структури, що самоорганізуються (self-organizing feature maps). Для них після вибору з шару n нейрона j з мінімальною відстанню D_j (1.19) навчається за формулою (1.18) не тільки цей нейрон, але і його сусіди, розташовані в околиці R . Величина R на перших ітераціях дуже велика, так що навчаються усі нейрони, але з часом вона зменшується до нуля. Таким чином, чим ближче кінець навчання, тим точніше визначається група нейронів, що відповідають кожному класу образів.

1.3. Архітектури нейронних мереж.

1.3.1. Перцептрони.

Ідею *перцептрону* (від латинського «perceptio» – сприйняття), запропонував американський нейрофізіолог Ф. Розенблатт. У 1960 році він представив перший нейрокомп'ютер – «Марк-1», який був здатний розпізнавати деякі букви англійського алфавіту.

У найбільш простому вигляді перцептрон (рис. 1.8) складається із сукупності чутливих (сенсорних) елементів (S -елементів), на які надходять вхідні сигнали. S -елементи випадковим чином пов'язані із сукупністю асоціативних A -елементів, вихід яких відрізняється від нуля тільки тоді, коли порушено досить велике число S -елементів, що впливають на один A -елемент. A -елементи з'єднані з реагуючими R -елементами, зв'язками, коефіцієнти посилення v яких змінюються в процесі навчання. Зважені комбінації виходів R -елементів складають реакцію системи, яка вказує на належність об'єкта, який розпізнається, певному образу.

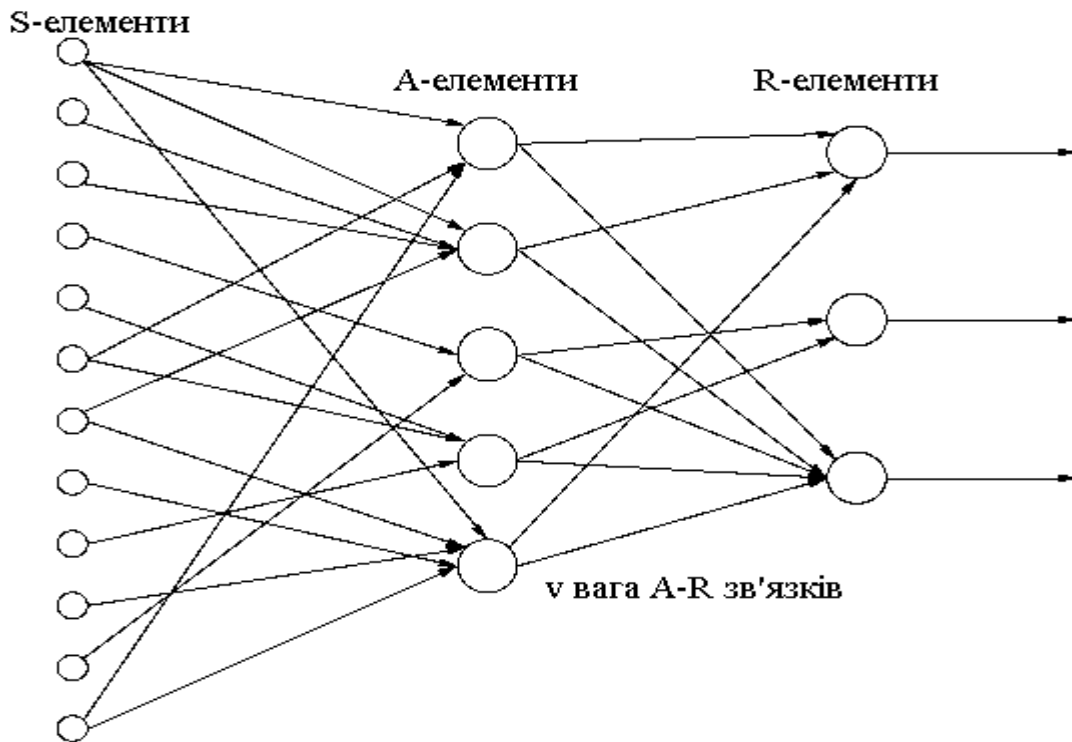


Рис. 1.8. Структурна схема перцептрону

Якщо розпізнаються тільки два образи, то перцептроном встановлюється тільки один R -елемент, який має дві реакції – позитивну і негативну. Якщо образів більше двох, то для кожного образу встановлюють свій R -елемент, вихід якого представляє лінійну комбінацію виходів A -елементів:

$$R_j = \Theta_j + \sum_{i=1}^n v_{ij} x_i, \quad (1.23)$$

де R_j – реакція j -го R -елемента; x_i – реакція i -го A -елемента; v_{ij} – вага зв'язку від i -го A -елемента до j -го R елемента; j – поріг j -го R -елемента.

Аналогічно записується рівняння i -го A -елемента:

$$x_i = \Theta_i + \sum_{k=1}^S y_k. \quad (1.24)$$

Сигнал y_k (1.24) може бути безперервним, але частіше за все він приймає тільки два значення: 0 або 1. Сигнали від S -елементів подаються на входи A -елементів з постійними вагами рівними 1, але кожен A -елемент пов'язаний тільки з групою випадково обраних S - елементів.

Припустимо, що потрібно навчити перцептрон розрізняти два образи V_1 і V_2 . Будемо вважати, що в перцептроні існує два R -елемента, один з яких призначений образу V_1 , а інший – образу V_2 . Перцептрон буде навчений вірно, якщо вихід R_1 перевищує R_2 , коли об'єкт, що розпізнається, належить образу V_1 , і навпаки. Розподіл об'єктів на два образи можна провести і за допомогою тільки одного R -елемента. Тоді об'єкту образу V_1 повинна відповідати позитивна реакція R -елемента, а об'єктам образу V_2 – негативна.

Перцептрон навчається шляхом пред'явлення навчальної послідовності зображень об'єктів, що належать образам V_1 і V_2 . У процесі навчання змінюються ваги v_i A -елементів. Зокрема, якщо застосовується система підкріплення з корекцією похибок, перш за все враховується вірність рішення, прийнятого перцептроном. Якщо рішення вірне, то ваги зв'язків всіх A -елементів, що спрацювали та ведуть до R -елементу, який видав правильне рішення, збільшуються, а ваги A -елементів, що не спрацювали, залишаються незмінними. Можна залишати незмінними ваги A -елементів, що спрацювали, але зменшувати ваги, що не спрацювали, тощо. Після процесу навчання перцептрон сам, без вчителя, починає класифікувати нові об'єкти.

Якщо перцептрон діє за описаною схемою і в ньому допускаються лише зв'язки, що йдуть від бінарних S -елементів до A -елементів та від A -елементів до єдиного R -елементу, то такий перцептрон прийнято називати *елементарним α -перцептроном*. Зазвичай класифікація задається вчителем. Перцептрон повинен виробити в процесі навчання класифікацію, задуману вчителем.

Про перцептрони було сформульовано і доведено кілька фундаментальних теорем, дві з яких, що визначають основні властивості перцептрона, наведені нижче.

Теорема 1. Клас елементарних α -перцептронів, для яких існує рішення для будь-якої задуманої класифікації, не є порожнім.

Ця теорема стверджує, що для будь-якої класифікації навчальної послідовності можна підібрати такий набір (з нескінченного набору) A -елементів, в

якому буде здійснено задуманий поділ навчальної послідовності за допомогою лінійного вирішального правила.

Теорема 2. Якщо для деякої класифікації рішення існує, то в процесі навчання α -перцептронів із корекцією похибок, що починається з довільного початкового стану, це рішення буде досягнуто протягом кінцевого проміжку часу.

Сенс цієї теореми полягає у тому, що якщо щодо задуманої класифікації можна знайти набір A -елементів, в якому існує рішення, то в рамках цього набору воно буде досягнуто в кінцевий проміжок часу.

Зазвичай обговорюють властивості нескінченного перцептрона, тобто перцептронів з нескінченим числом A -елементів із усілякими зв'язками з S -елементами (повний набір A -елементів). В таких перцептронах рішення завжди існує, а раз воно існує, то воно і досягне в α -перцептронах з корекцією похибок.

Дуже цікавою областю досліджень є також багат шарові перцептрони і перцептрони з перехресними зв'язками.

1.3.2. Нейронні мережі зустрічного розповсюдження.

Мережа зустрічного розповсюдження складається із вхідного шару нейронів, а також шарів нейронів Кохонена і Гросберга (рис.1.9). Кожен елемент вхідного сигналу подається на всі нейрони шару Кохонена, кожен нейрон якого з'єднаний з усіма нейронами шару Гросберга. Відмінність НМ зустрічного розповсюдження від інших багат шарових мереж з послідовними зв'язками складається в операціях, які виконуються нейронами Кохонена і Гросберга.

В режимі функціонування мережі пред'являється вхідний сигнал X і формується вихідний сигнал Y . У режимі навчання на вхід мережі подається вхідний сигнал і ваги корегуються, щоб НМ видавала необхідний вихідний сигнал.

НМ зустрічного розповсюдження функціонує наступним чином.

Шар Кохонена у своїй найпростішій формі функціонує за правилом «переможець отримує все». Для даного вхідного вектора один і тільки один нейрон Кохонена видає логічну 1, всі інші видають 0.

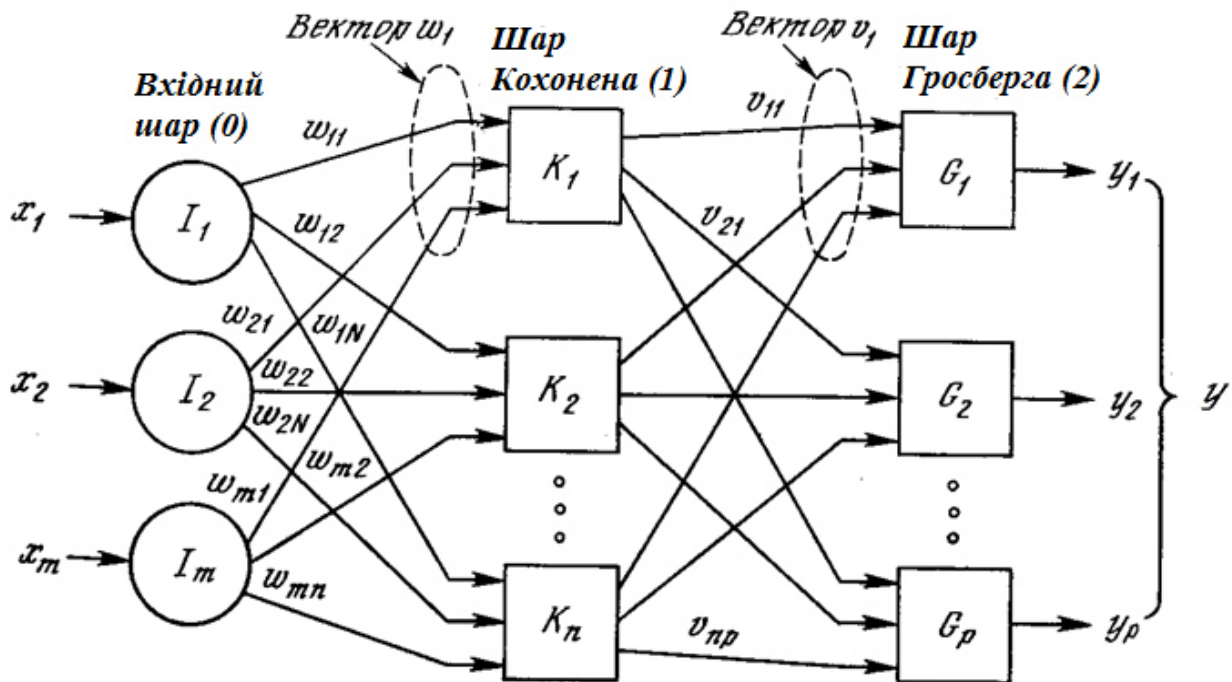


Рис. 1.9. Структурна схема НМ зустрічного розповсюдження

Вихід кожного нейрона Кохонена є сумою зважених елементів вхідних сигналів:

$$s_j = w_{1j}x_1 + w_{2j}x_2 + \dots + w_{nj}x_n = \sum_i x_i w_{ij}, \quad (1.25)$$

де s_j – вихід j -го нейрону Кохонена, $W_j = (w_{1j}, w_{2j}, \dots, w_{nj})$ – вектор ваг j -го нейрону Кохонена, $X = (x_1, x_2, \dots, x_n)$ – вектор вхідного сигналу, або у векторно-матричній формі:

$$S = X \cdot W, \quad (1.26)$$

де S – вектор виходів шару Кохонена.

Нейрон Кохонена з максимальним значенням s_j є «переможцем». Його вихід дорівнює 1, у інших – 0.

Шар Гросберга функціонує у схожій манері. Його вихід є зваженою сумою виходів шару Кохонена (тобто він є шаром нейронів з лінійними функціями активації).

Якщо шар Кохонена функціонує таким чином, що лише один вихід дорівнює 1, а інші – 0, то кожен нейрон шару Гросберга видає величину ваги, яка пов'язує цей нейрон з єдиним нейроном Кохонена, чий вихід відмінний від 0.

НМ зустрічного розповсюдження вимагає попередньої обробки вхідних векторів шляхом їх нормалізації. Така нормалізація виконується шляхом ділення кожної компоненти вхідного вектору на його довжину (квадратний корінь з суми квадратів компонент вектора). Це перетворює вхідний вектор в одиничний вектор з тим же напрямком в n -вимірному просторі.

НМ зустрічного поширення навчається наступним чином.

Шар Кохонена класифікує вхідні вектори у групи схожих. Це досягається за допомогою такого підстроювання ваг шару Кохонена, що близькі вхідні вектори активують один і той же нейрон даного шару. Потім завданням шару Гросберга є отримання необхідних виходів.

Шар Кохонена навчається без вчителя (самонавчається). Тому важко (й не потрібно) передбачати, який саме нейрон Кохонена має використовуватися для заданого вхідного вектора. Необхідно лише гарантовано домогтися, щоб в результаті навчання розділялися несхожі вхідні вектори.

Виходи шару Кохонена подаються на входи нейронів шару Гросберга. Виходи нейронів обчислюються як при нормальному функціонуванні. Далі кожна вага корегується лише у тому випадку, якщо вона з'єднана з нейроном Кохонена, що має ненульовий вихід. Величина корекції ваги пропорційна різниці між вагою і необхідним виходом нейрона Гросберга. Навчання шару Гросберга – це навчання із вчителем, алгоритм використовує задані бажані виходи.

НМ зустрічного розповсюдження застосовуються для розпізнавання та відновлення образів (асоціативна пам'ять), стиснення даних (із втратами), тощо.

У повній моделі НМ зустрічного розповсюдження є можливість отримувати вихідні сигнали по вхідним, і навпаки. Цим діям відповідають пряме і зворотне поширення сигналів.

Переваги НМ зустрічного розповсюдження:

1. Простота. Такі НМ дають можливість отримувати статистичні властивості з множин вхідних сигналів. Кохоненом показано, що для повністю навченої НМ ймовірність того, що випадково обраний вхідний вектор (у відповідності до функції щільності ймовірності вхідної множини) буде найближчим до будь-якого заданого вагового вектора, дорівнює оберненому числу нейронів Кохонена.

2. Швидке навчання. Час навчання таких НМ у порівнянні з НМ зі зворотнім поширенням може бути в 100 разів менше. За своїми можливостями будувати відображення, НМ зустрічного розповсюдження значно перевершує одношарові перцептрони.

3. Корисність для додатків, у яких потрібна швидка початкова апроксимація.

4. Широке застосування при вирішенні практичних завдань, оскільки такі НМ дають можливість будувати функцію і обернену до неї.

До недоліків НМ зустрічного розповсюдження слід віднести те, що мережа не дає можливості будувати точні апроксимації (точні відображення), у чому значно поступається НМ зі зворотнім поширенням похибки. Ще одним недоліком цих мереж є слабе теоретичне опрацювання їх модифікацій. Цими модифікаціями НМ зустрічного розповсюдження є:

1. НМ зустрічного поширення можуть відрізнятися способами визначення початкових значень синапсичних ваг. Так, крім традиційних випадкових значень із заданого діапазону, можуть бути використані значення у відповідності до відомих методів опуклої комбінації.

2. Для підвищення ефективності навчання таких НМ може застосовуватися додавання шуму до вхідних векторів.

3. Надання кожному нейрону «почуттям справедливості»: якщо нейрон стає переможцем частіше, ніж значення оберненого числа нейронів Кохонена, то йому тимчасово збільшують поріг, даючи тим самим навчатися й іншим нейронам.

4. Крім «методу акредитації», при якому для кожного вхідного вектора активується лише один нейрон Кохонена, може бути використаний також «метод інтерполяції». При застосуванні останнього ціла група нейронів Кохонена, що мають найбільші виходи, може передавати свої вихідні сигнали в шар Гросберга. Такий підхід підвищує точність відображень, що реалізуються мережею і реалізований в так званих само організованих картах.

1.3.3. Нейронні мережі з радіальними базисними функціями.

У загальному випадку під терміном НМ з радіальними базисними функціями (елементами РБФ) (Radial Basis Function Network – RBF) розуміється двошарова мережа без зворотних зв'язків, яка містить прихований шар радіально симетричних прихованих нейронів (шаблонний шар) (рис. 1.10).

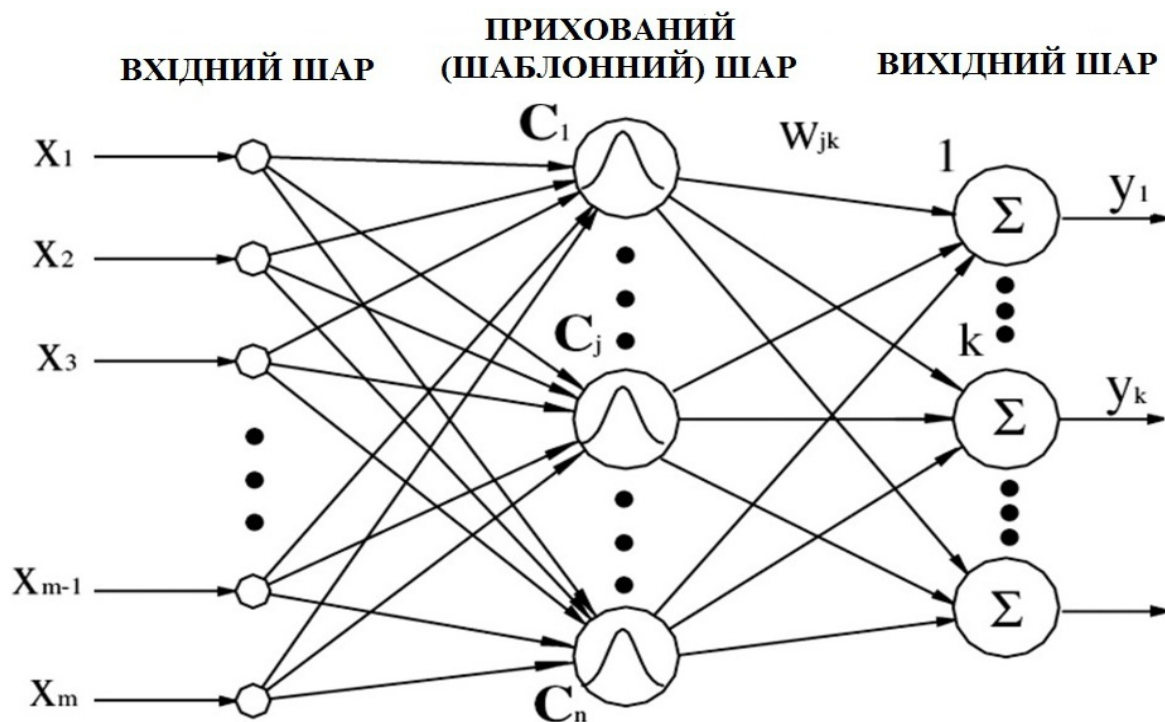


Рис. 1.10. Структурна схема НМ з радіальними базисними функціями

Для того щоб шаблонний шар був радіально-симетричним, необхідне виконання наступних умов:

- наявність центру, представленого у вигляді вектора у вхідному просторі (зазвичай цей вектор зберігається в просторі ваг від вхідного шару до шару шаблонів);
- наявність способу вимірювання відстані вхідного вектора від центру (зазвичай це стандартна евклідова відстань);
- наявність спеціальної функції проходження від одного аргументу, що визначає вихідний сигнал нейрону шляхом відображення функції відстані (зазвичай використовується функція Гауса:

$$\varphi(s) = e^{-s^2}. \quad (1.27)$$

Іншими словами, вихідний сигнал шаблонного нейрона – це функція відстані між вхідним вектором X і збереженим центром C :

$$f(X) = \varphi\left(\frac{\|X - C\|}{\sigma}\right). \quad (1.28)$$

Вихідний шар НМ є лінійним, отже виходи мережі визначаються виразом

$$y_j = \sum_{i=1}^K w_{ij} \varphi\left(\frac{\|X - C_i\|}{\sigma_i}\right), j=1, 2, \dots, m. \quad (1.29)$$

де C_i – центри і σ_i – відхилення радіальних елементів.

Навчання НМ з РБФ відбувається в декілька етапів. Спочатку визначаються центри і відхилення для радіальних елементів, після цього оптимізуються параметри w_{ij} лінійного вихідного шару. При цьому, розташування центрів має відповідати кластерам, які реально присутні у вихідних даних.

Методи, які найчастіше використовуються для навчання НМ з РБФ:

1. *Вибірка з вибірки.* Як центри радіальних елементів беруться декілька випадково обраних точок навчальної множини. Через випадковість вибору вони «представляють» розподіл навчальних даних в статичному сенсі.

2. *Алгоритм С-середніх.* Цей алгоритм прагне вибрати оптимальну множину точок, які є центроїдами кластерів в навчальних даних. При цьому центри радіальних елементів розташовуються таким чином, щоб виконувались наступні умови:

- кожна навчальна точка «відноситься» до одного центру кластера і лежить до нього ближче, ніж до будь-якого іншого центру;
- кожен центр кластеру є центроїдом множини навчальних точок, що належать до цього кластеру.

Далі необхідно знайти відхилення або згладжуючий фактор, величина якого визначає наскільки «гострою» буде функція відстані (функція Гауса). Якщо ці функції обрані занадто гострими, НМ не буде інтерполювати дані між близькими відомими точками, і втратить здатність до узагальнення. Якщо функції взяті надто широкими, мережа не буде сприймати дрібні деталі.

Як правило, відхилення обираються таким чином, щоб «ковпак» кожної функції Гауса захоплював кілька сусідніх центрів. Для цього є кілька методів:

- *явний*, коли відхилення задаються користувачем;
- *ізотропний*, коли відхилення береться однаковим для всіх елементів і визначається евристично з урахуванням кількості радіальних елементів і обсягу простору, що покривається;
- *найближчих сусідів*, коли відхилення кожного елементу встановлюється (індивідуально) рівним середній відстані до його найближчих сусідів (в тих частинах простору, де точки розташовані густо відхилення будуть менше, а там, де точок мало, відхилення будуть великими, і буде проводитися інтерполяція).

Далі параметри вихідного шару оптимізуються за допомогою стандартного методу лінійної оптимізації – алгоритму псевдо обернених матриць (сингулярного розкладання).

Взагалі можуть бути побудовані різні гібридні різновиди НМ з РБФ. Наприклад, вихідний шар може мати нелінійні функції активації, і тоді для його навчання використовується будь-який з алгоритмів навчання багат шарових перцептронів, наприклад, метод зворотного поширення. Можна також навчати радіальний (прихований) шар за допомогою алгоритму навчання Кохонена – розміщувати центри таким чином, щоб вони відображали розташування даних.

НМ з РБФ мають ряд переваг перед багат шаровими мережами прямого розповсюдження. По-перше, вони моделюють довільну нелінійну функцію за

допомогою всього одного проміжного шару, тим самим позбавляючи від необхідності вирішувати питання про кількість шарів. По-друге, параметри лінійної комбінації в вихідному шарі можна повністю оптимізувати за допомогою добре відомих методів лінійної оптимізації, які працюють швидко і не відчують труднощів з локальними мінімумами, які заважають при навчанні з використанням алгоритму зворотного поширення похибки. Тому така мережа навчається дуже швидко.

До недоліків НМ з РБФ слід віднести те, що вони мають погані екстраполюючі властивості, тому виходять досить громіздкими при великій розмірності вектора входів.

1.3.4. Нейронні мережі Хопфілда і Хемінга.

Серед різних конфігурацій штучних НМ зустрічаються такі, при класифікації яких за принципом навчання, напевно, не підходять ні навчання з вчителем, ні навчання без вчителя. У таких мережах вагові коефіцієнти синапсів розраховуються тільки один раз перед початком функціонування мережі на основі інформації про оброблювані дані, і все навчання мережі зводиться саме до цього розрахунку. З одного боку, пред'явлення апріорної інформації можна розцінювати, як допомога вчителя, але з іншого – мережа фактично просто запам'ятовує зразки до того, як на її вхід надходять реальні дані, і не може змінювати свою поведінку, тому говорити про зворотній зв'язок зі «вчителем» не доводиться. Серед НМ з подібною логікою роботи найбільш відомі мережі Хопфілда і Хемінга, які зазвичай використовуються для організації асоціативної пам'яті.

Структурна схема мережі Хопфілда наведена на рис. 1.11. Вона складається з єдиного шару нейронів, число яких є одночасно числом входів і виходів мережі. Кожен нейрон пов'язаний синапсами з усіма іншими нейронами, а також має один вхідний синапс, через який здійснюється введення сигналу. Вихідні сигнали, як зазвичай, утворюються на аксонах.

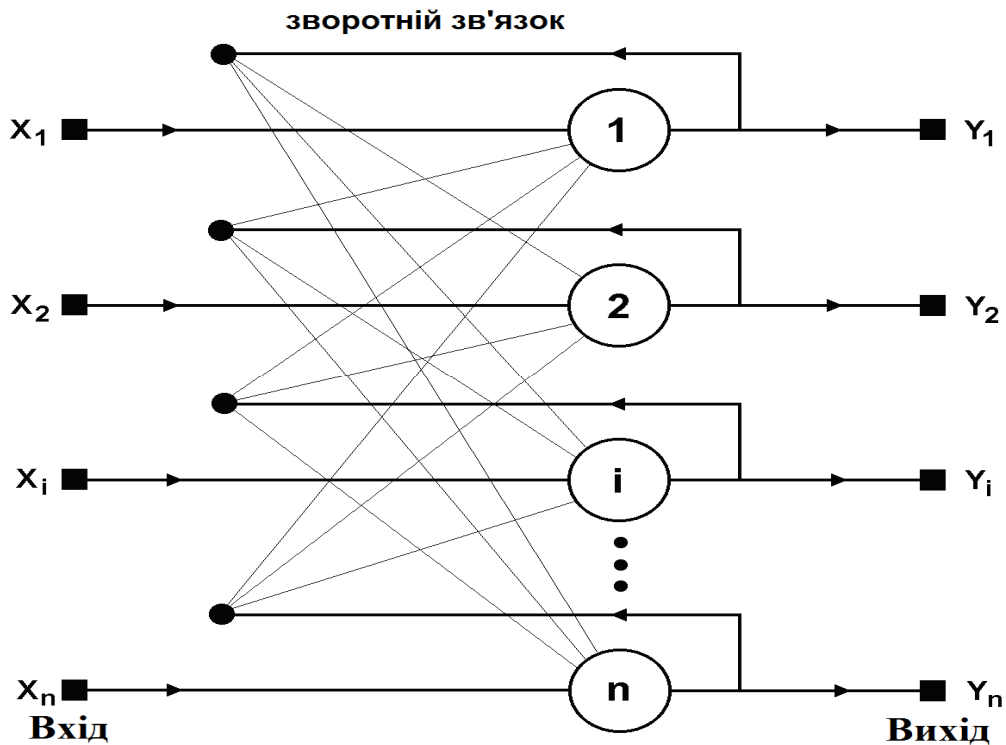


Рис. 1.11. Структурна схема НМ Хопфілда

Завдання, яке вирішується даною мережею в якості асоціативної пам'яті, зазвичай, формулюється у такий спосіб. Відомий деякий набір двійкових сигналів (зображень, звукових та інших даних, що описують якісь об'єкти або характеристики процесів), які вважаються зразковими. Мережа повинна вміти з довільного неідеального сигналу, поданого на її вхід, виділити («згадати» по частковій інформації) відповідний зразок (якщо такий є) або «дати висновок» про те, що вхідні дані не відповідають жодному із зразків. У загальному випадку, будь-який сигнал може бути описаний вектором $X = \{x_i: i=0 \dots n-1\}$, n – число нейронів в мережі та розмірність вхідних і вихідних векторів. Кожен елемент x_i дорівнює або $+1$, або -1 . Позначимо вектор, що описує зразок k , через X^k , а його компоненти, відповідно, $-x_i^k, k=0 \dots m-1, m$ – число зразків. Коли мережа розпізнає (або «згадає») будь-який зразок на основі пред'явлених їй даних, її виходи будуть містити саме його, тобто $Y = X^k$, де Y – вектор вихідних значень мережі: $Y = \{y_i: i=0, \dots, n-1\}$. В іншому випадку, вихідний вектор не співпаде ні з одним зразком.

Якщо, наприклад, сигнали являють собою зображення, то, відобразивши в графічному вигляді дані з виходу мережі, можна буде побачити картинку, яка повністю збігається з однією зі зразкових (у разі успіху) або ж «вільну імпровізацію» мережі (у разі невдачі).

На стадії ініціалізації мережі вагові коефіцієнти синапсів встановлюються таким чином:

$$w_{ij} = \begin{cases} \sum_{k=0}^{m-1} x_i^k x_j^k, i \neq j \\ 0, i = j \end{cases}, \quad (1.30)$$

тут i і j – індекси, відповідно, предсинапсичного і постсинапсичного нейронів; x_i^k, x_j^k – i -ий і j -ий елементи вектора k -го зразка.

Алгоритм функціонування мережі наступний (p – номер ітерації):

1. На входи мережі подається невідомий сигнал. Фактично його введення здійснюється безпосередньою встановленням значень аксонів:

$$y_i(0) = x_i, i=0 \dots n-1, \quad (1.31)$$

тому позначення на схемі мережі вхідних синапсів в явному вигляді носить чисто умовний характер. Нуль в дужках біля y_i означає нульову ітерацію в циклі роботи мережі.

2. Розраховується новий стан нейронів

$$s_j(p+1) = \sum_{i=0}^{n-1} w_{ij} y_i(p), j=0 \dots n-1. \quad (1.32)$$

і нові значення аксонів

$$y_j(p+1) = f[s_j(p+1)], \quad (1.33)$$

де f – активаційна функція у вигляді стрибка, що приведена на рис. 1.12,а.

3. Перевірка, чи змінилися вихідні значення аксонів за останню ітерацію. Якщо так – перехід до пункту 2, інакше (якщо виходи застabilізовано) – кінець алгоритму. При цьому вихідний вектор являє собою зразок, який найкращим чином поєднується з вхідними даними.

Іноді мережа не може провести розпізнавання і видає на виході неіснуючий образ. Це пов'язано із проблемою обмеженості можливостей мережі. Для

НМ Хопфілда число образів, що запам'ятовуються не повинно перевищувати величини, приблизно рівної $0.15 \cdot n$. Крім того, якщо два образи A і B дуже схожі, вони, можливо, будуть викликати у мережі перехресні асоціації, тобто пред'явлення на входи мережі вектора A призведе до появи на її виходах вектору B , і навпаки.

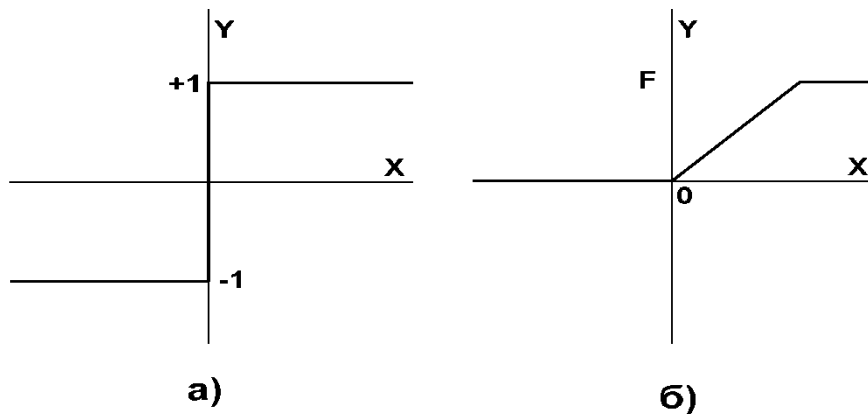


Рис. 1.12. Активаційні функції, які використовуються в НМ Хопфілда (а) та Хемінга (б)

Коли немає необхідності, щоб НМ у явному вигляді видавала зразок, тобто досить, скажімо, отримувати номер зразка, асоціативну пам'ять успішно реалізує мережа Хемінга. У порівнянні з НМ Хопфілда, НМ Хемінга характеризується меншими витратами на пам'ять і обсягом обчислень, що стає очевидним з її структури (рис. 1.13).

Мережа складається з двох шарів, які мають по m нейронів, де m – число зразків. Нейрони першого шару мають по n синапсів, з'єднаних зі входами мережі (утворюють фіктивний нульовий шар). Нейрони другого шару пов'язані між собою інгібіторними (негативними зворотними) синапсичними зв'язками. Єдиний синапс з позитивним зворотнім зв'язком для кожного нейрону з'єднаний з його ж аксоном.

Ідея роботи мережі полягає в знаходженні відстані Хемінга від тестованого образу до всіх зразків. Відстанню Хемінга називається число бітів, що відрі-

зняються, в двох бінарних векторах. Мережа повинна вибрати зразок з мінімальною відстанню Хемінга до невідомого вхідного сигналу, в результаті чого буде активований тільки один вихід мережі, що відповідає цьому зразку.

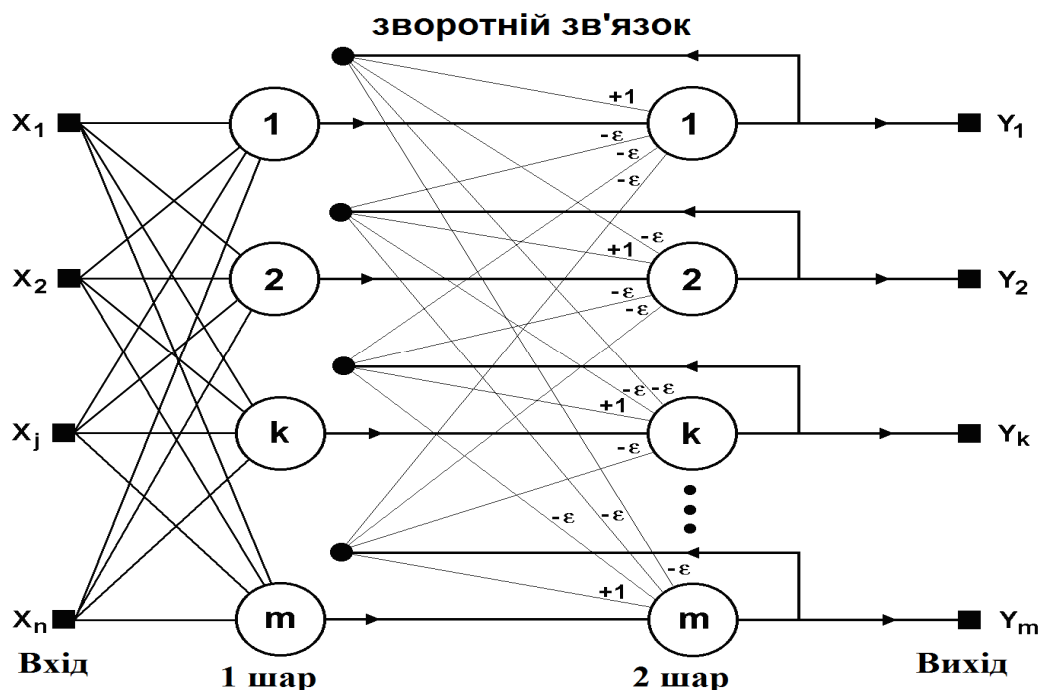


Рис. 1.13. Структурна схема НМ Хемінга

На стадії ініціалізації ваговим коефіцієнтам першого шару і порогу активаційної функції привласнюються наступні значення:

$$w_{ik} = \frac{x_i^k}{2}, \quad i=0 \dots n-1, \quad k=0 \dots m-1, \quad (1.34)$$

$$T_k = n/2, \quad k=0 \dots m-1, \quad (1.35)$$

де x_i^k – елемент i зразка k .

Вагові коефіцієнти гальмуючих синапсів у другому шарі беруть рівними деякій величині $0 < \varepsilon < 1/m$. Синапс нейрону, пов'язаний з його ж аксоном має вагу $+1$.

Алгоритм функціонування НМ Хемінга наступний:

1. На входи мережі подається невідомий вектор $X = \{x_i; i=0 \dots n-1\}$, за яким розраховуються стани нейронів першого шару (верхній індекс у дужках вказує номер шару):

$$y_j^{(1)} = s_j^{(1)} = \sum_{i=0}^{n-1} w_{ij} x_i + T_j, j=0 \dots m-1 \quad (1.36)$$

Після цього отриманими значеннями ініціалізуються значення аксонів другого шару:

$$y_j^{(2)} = y_j^{(1)}, j=0 \dots m-1. \quad (1.37)$$

2. Обчислити нові стани нейронів другого шару:

$$s_j^{(2)}(p+1) = y_j(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), k \neq j, j = 0 \dots m-1, \quad (1.38)$$

і значення їх аксонів:

$$y_j^{(2)}(p+1) = f[s_j^{(2)}(p+1)], j = 0 \dots m-1. \quad (1.39)$$

Активаційна функція f має вигляд порога (рис. 1.12,б), причому величина F повинна бути досить великою, щоб будь-які можливі значення аргументу не призводили до насичення.

3. Перевірити, чи змінилися виходи нейронів другого шару за останню ітерацію. Якщо так – перейди до кроку 2. Інакше – кінець алгоритму.

З оцінки алгоритму видно, що роль першого шару досить умовна: скориставшись значеннями його вагових коефіцієнтів один раз на кроці 1, мережа більше не звертається до нього, тому перший шар може бути взагалі виключений з НМ (замінений на матрицю вагових коефіцієнтів).

Отже, НМ Хопфілда і Хемінга дозволяють просто і ефективно вирішити задачу відтворення образів по неповній та спотвореній інформації. Невисока ємність мереж (число образів, що запам'ятовуються) пояснюється тим, що вони не просто запам'ятовують образи, а дозволяють проводити їх узагальнення, наприклад, за допомогою НМ Хемінга можлива класифікація за критерієм максимальної правдоподібності. Разом з тим, легкість побудови програмних і апаратних моделей роблять ці НМ привабливими для багатьох застосувань.

1.3.5. Ймовірнісні нейронні мережі.

Задача оцінки щільності ймовірності за наявними даними має давню історію в математичній статистиці.

Зазвичай при цьому передбачається, що щільність має деякий певний вид (найчастіше – нормальний розподіл). Після цього оцінюються параметри моделі. Нормальний розподіл часто використовується тому, що тоді параметри моделі (середнє і стандартне відхилення) можна оцінити аналітично.

Слід зазначити, що припущення про нормальність є далеко не завжди виправданим.

Інший підхід до оцінки щільності ймовірності заснований на так званих ядерних оцінках. Той факт, що спостереження розташоване в даній точці простору, свідчить про те, що в цій точці є деяка щільність ймовірності. Кластери з близько лежачими точками вказують на те, що в цьому місці щільність ймовірності велика. Поблизу точки спостереження є більша довіра до рівня щільності, а при віддаленні від нього довіра убуває і прагне до нуля. У методі ядерних оцінок в точці, що відповідає кожному спостереженню, розміщується деяка проста функція, далі всі вони складаються і в результаті виходить оцінка для загальної щільності ймовірності. Найчастіше в якості ядерних функцій беруться дзвіноподібні гаусові функції. Якщо навчальних прикладів достатньо, то цей метод дає досить добре наближення до істинної щільності ймовірності.

Метод апроксимації щільності ймовірності за допомогою ядерних функцій схожий на метод РБФ. Звідси природно впливають поняття ймовірнісної НМ (Probabilistic Neural Networks – PNN). PNN-мережі призначені для задач класифікації, а GRNN – для задач регресії. НМ цих двох типів є реалізацією методів ядерної апроксимації, оформлених у вигляді НМ.

НМ PNN має щонайменше три шари: вхідний, радіальний і вихідний. Радіальні елементи беруться по одному на кожне навчальне спостереження, й кожен з них є функцією Гауса з центром в цьому спостереженні. Кожному класу відповідає один вихідний елемент, який, у свою чергу, з'єднаний з усіма радіальними елементами, що відносяться до його класу, а з усіма іншими радіальни-

ми елементами він має нульове з'єднання. Таким чином, вихідний елемент просто складає відгуки всіх елементів, що належать до його класу. Значення вихідних сигналів виходять пропорційними ядерним оцінкам ймовірності належності відповідним класам, та після нормування їх до одиниці, отримують остаточні оцінки ймовірності належності класам.

Вихід даної мережі, що відповідає класу, описується виразом

$$y = \frac{1}{N\sigma^n} \sum_{k=1}^N \varphi\left(\frac{\|X - X^k\|}{\sigma}\right), \quad (1.40)$$

де n – розмірність вхідного вектора, N – обсяг навчальної вибірки, φ – радіальна функція (Гауса); X^k – елемент (вектор) цієї вибірки, відповідний зазначеному класу.

НМ PNN має єдиний керуючий параметр навчання, значення якого повинно обиратися користувачем, – відхилення функції Гауса σ (параметр згладжування). Цей параметр, як і для НМ RBF, обирається з тих міркувань, щоб «шапки» певну кількість раз перекривались: вибір занадто маленьких відхилень призведе до «гострих» апроксимуючих функцій і нездатності НМ до узагальнення, а при занадто великих відхиленнях будуть губитися деталі. Необхідне значення нескладно знайти дослідним шляхом, підбираючи його таким чином, щоб контрольна похибка була якомога менше. На щастя, НМ PNN не дуже чутливі до вибору параметра згладжування.

Перевагою НМ PNN є швидке навчання. При цьому час витрачається тільки на те, щоб подавати на вхід мережі навчальні спостереження, і мережа працює настільки швидко, наскільки це взагалі можливо. Інша перевага – вихідне значення НМ має ймовірнісний сенс, і тому його легше інтерпретувати.

Істотним недоліком таких мереж є їх обсяг. НМ PNN фактично вміщує в себе все навчальні дані, тому вона вимагає багато пам'яті і може повільно працювати на фон-неймановських машинах.

PNN-мережі особливо корисні при пробних експериментах (наприклад, коли потрібно вирішити, які з вхідних змінних використовувати), так як завдя-

ки короткому часу навчання можна швидко виконати велику кількість пробних тестів.

1.3.6. Узагальнено-регресійні нейронні мережі.

Узагальнено-регресійна НМ (Generalized Regression Neural Network – GRNN) влаштована аналогічно НМ PNN: в точку розташування кожного навчального спостереження поміщається гаусівська ядерна функція і вважається, що кожне спостереження свідчить про деяку впевненість в тому, що поверхня відгуку в даній точці має певну висоту, і ця впевненість убуває при відході в сторону від точки. Мережа GRNN копіює всередину себе всі навчальні спостереження та використовує їх для оцінки відгуку в довільній точці. Остаточна вихідна оцінка мережі виходить як зважене середнє виходів по всім навчальним спостереженням:

$$y = \frac{\sum_{k=1}^N y^k \varphi(\|X - X^k\|/\sigma)}{\sum_{k=1}^N \varphi(\|X - X^k\|/\sigma)}, \quad (1.41)$$

де X^k, y^k – точки навчальної вибірки. Інші позначення – як в формулі (1.40).

Перший проміжний шар НМ GRNN складається з радіальних елементів, а другий – з елементів, які допомагають оцінити зважене середнє. При цьому, кожен вихід має в вихідному шарі свій елемент, що формує для нього зважену суму. Щоб отримати зі зваженої суми зважене середнє, цю суму потрібно поділити на суму вагових коефіцієнтів. Останню суму обчислює спеціальний елемент другого шару. Далі в вихідному шарі проводиться власне ділення (за допомогою спеціальних елементів «ділення»). Таким чином, число елементів в другому проміжному шарі на одиницю більше, ніж у вихідному. Як правило, в задачах регресії потрібно оцінити одне вихідне значення, і, відповідно, другий проміжний шар містить два елементи.

Для зменшення розмірів НМ GRNN і збільшення швидкості навчання, можна модифікувати мережу таким чином, щоб радіальні елементи відповідали

не окремим навчальним випадків, і їх кластерам. Центри для таких елементів можна обирати за допомогою будь-якого призначеного для цієї мети алгоритму (вибірки з вибірки, *C*-середніх або Кохонена).

Переваги та недоліки у мережі GRNN в основному такі ж, як і у НМ PNN: НМ GRNN навчається майже миттєво, але може вийти великою і повільною (хоча в ній не обов'язково мати по одному радіальному елементу на кожен навчальний приклад, їх число все одно буде великим), а також НМ GRNN не має здатності екстраполювати дані.

1.3.7. Лінійні нейронні мережі.

Згідно із загальноприйнятим у науці принципом, якщо більш складна модель не дає кращих результатів, ніж простіша, то з них слід віддати перевагу другій. У термінах апроксимації відображень найпростішою моделлю буде лінійна, в якій апроксимуюча функція визначається гіперплощиною. У задачі класифікації гіперплощина розміщується таким чином, щоб вона розділяла собою два класи (лінійна дискримінантна функція). Лінійна модель зазвичай задається рівнянням

$$Y = XW + B, \quad (1.42)$$

де W – матриця ваг мережі, B – вектор зсувів.

Лінійна НМ є мережею без проміжних шарів, яка в вихідному шарі містить тільки лінійні елементи (тобто елементи з лінійною функцією активації). Ваги відповідають елементам матриці, пороги – компонентам вектора зсуву. Під час роботи мережа фактично помножує вектор входів на матрицю ваг, а потім до отриманого вектору додає вектор зсуву.

Одним з найбільш поширених методів навчання лінійної НМ є стандартний алгоритм лінійної оптимізації, заснований на псевдо обернених матрицях. Лінійна НМ є точкою відліку для оцінки якості побудованих НМ. Може виявитися так, що задачу, яка вважалася дуже складною, можна успішно вирішити не тільки НМ, а й простим лінійним методом. Якщо ж в задачі не так багато на-

вчальних даних, то, ймовірно, просто немає підстав використовувати більш складні моделі.

1.4. Ефективність нейронних мереж

НМ є універсальними структурами, що дозволяють реалізувати будь-який обчислювальний алгоритм. Ця ефективність НМ впливає з ряду так званих теорем про повноту.

Теорема про повноту. Будь-яка безперервна функція на замкнутій обмеженій множині може бути рівномірно наближена функціями, що обчислюються нейронними мережами, якщо функція активації нейрону двічі безперервно диференційована і безперервна.

У 1989 р. К. Фунахаші (Funahashi) показав, що нескінченно велика НМ з єдиним прихованим шаром здатна апроксимувати будь-яку безперервну функцію, та сформулював дане твердження у формі наступної теореми.

Теорема Фунахаші. Нехай $\phi(x)$ – непостійна, обмежена і монотонно зростаюча безперервна функція. Нехай, $U \subset \mathbb{R}^n$ – обмежена множина і $f : U \rightarrow \mathbb{R}$ – речова безперервна функція, що визначена на U . Тоді для довільного $\varepsilon > 0$ існує ціле L і речові константи w_i, w_{ij} , такі, що апроксимація

$$\tilde{f}(x_1, x_2, \dots, x_n) = \sum_{i=1}^L w_i \phi\left(\sum_{j=1}^n w_{ij} x_j\right) \quad (1.43)$$

задовольняє нерівності

$$\|f - \tilde{f}\|_{\infty} = \sup_{x \in U} |f(x) - \tilde{f}(x)| \leq \varepsilon. \quad (1.44)$$

Іншими словами, будь-яке безперервне відображення може бути апроксимоване в сенсі однорідної топології на U двошаровою НМ з активаційними функціями $\phi(x)$ для нейронів прихованого шару і лінійними активаційними функціями для нейронів вихідного шару.

При цьому, основними недоліками технології НМ є:

- відсутність суворої теорії з вибору структури НМ;
- практична неможливість вилучення набутих знань з навченої НМ (оскільки НМ практично завжди – «річ у собі», тобто «чорний ящик» для дослідника).

Контрольні запитання

1. Розкрийте суть наступних підходів до побудови інтелектуальних систем: структурного, логічного, еволюційного та імітаційного.
2. Наведіть області використання нейронних мереж.
3. Наведіть структуру і властивості штучного нейрона.
4. Розкрийте типи нейронних мереж з точки зору їх топології.
5. В чому суть навчання нейронних мереж?
6. Розкрийте суть алгоритму зворотного поширення похибки.
7. Наведіть методи навчання нейронних мереж без вчителя.
8. В чому особливості нейронних мереж зустрічного розповсюдження?
9. В чому особливості нейронних мереж Хопфілда і Хемінга?
10. З яких теорем впливає ефективність нейронних мереж?

Глава 2. НЕЧІТКА ЛОГІКА І СИСТЕМИ НЕЧІТКОГО ВИСНОВКУ

2.1. Вступ

Теорія нечітких множин (fuzzy sets theory) веде свій початок з 1965 р., коли професор Лотфі Заде (Lotfi Zadeh) з університету Берклі опублікував свою основну роботу «Fuzzy Sets» в журналі «Information and Control». Ця робота заклала основи моделювання інтелектуальної діяльності людини і стала початковим поштовхом у розвитку нової математичної теорії.

Прикметник «fuzzy», який можна перекласти як «нечіткий», «розмитий», «пухнастий», введено в назву нової теорії з метою дистанціювання від традиційної чіткої математики і аристотелевої логіки, що оперують з чіткими поняттями: «належить – не належить», «істина – неправда». Концепція нечіткої множини зародилася у Заде як «незадоволеність математичними методами класичної теорії систем, яка змушувала домагатися штучної точності, недоречної в багатьох системах реального світу, особливо в так званих гуманістичних системах, що включають людей».

Л. Заде розширив класичне поняття множини (по Г. Кантору), допустивши, що характеристична функція (*функція належності* елемента множині) може приймати будь-які значення в інтервалі $[0; 1]$, а не тільки значення 0 або 1. Такі множини були названі їм нечіткими (fuzzy). Він визначив також ряд операцій над нечіткими множинами і запропонував узагальнення відомих методів логічного висновку *modus ponens* і *modus tollens*. Ввівши поняття лінгвістичної змінної, і допустивши, що в якості її значень (термів) виступають нечіткі множини, Л. Заде створив апарат для опису процесів інтелектуальної діяльності, включаючи нечіткість і невизначеність виразів.

Подальші роботи професора Л. Заде і його послідовників заклали міцний фундамент нової теорії і створили передумови для впровадження методів нечіткого керування в інженерну практику.

Прийнято виділяти три періоди в розвитку теорії нечіткої логіки і нечітких систем. Перший період (кінець 60-х – початок 70 рр. ХХ ст.) Характеризується розвитком теоретичного апарату нечітких множин (Заде, Мамдані, Беллман). У другому періоді (70-80-ті рр. ХХ ст.) з'являються перші практичні результати в області нечіткого керування технічними системами (поршневий двигун). Одночасно вчені колективи стали приділяти увагу питанням побудови експертних систем на основі нечіткої логіки, розробці нечітких контролерів. Нарешті, в третьому періоді, який триває з кінця 80-х років ХХ ст. по теперішній час, з'являються пакети програм для побудови нечітких експертних систем, а області застосування нечіткої логіки помітно розширюються. До початку 90-х рр. ХХ ст. більша частина досліджень велась на Сході (Японія, Китай).

Наразі, системи з нечіткою логікою успішно впроваджені в таких областях, як керування технологічними процесами, керування транспортом, медична діагностика, технічна діагностика, фінансовий менеджмент, біржове прогнозування, розпізнавання образів, тощо. Спектр додатків дуже широкий – від відеокамер і побутових пральних машин до засобів наведення ракет протиповітряної оборони і керування бойовими вертольотами. Практичний досвід розробки систем нечіткого логічного висновку свідчить про те, що терміни і вартість їх проектування значно менше, ніж при використанні традиційного математичного апарату; при цьому забезпечується необхідний рівень робастності і прозорості моделей.

Поняття нечіткої множини – це спроба математичної формалізації нечіткої інформації для побудови математичних моделей. В основі цього поняття лежить уявлення про те, що елементи, які складають дану множину та володіють загальною властивістю, можуть володіти цією властивістю у різній мірі й, отже, належати до даної множини із різною мірою. У разі такого підходу вислови про те, що «елемент належить даній множині» втрачають сенс, оскільки необхідно вказати «наскільки сильно» цей елемент задовольняє властивостям даної множини.

2.2. Нечіткі множини

2.2.1. Визначення нечітких множин.

Підхід до формалізації поняття нечіткої множини складається в узагальненні поняття належності. В теорії класичних множин існує кілька способів завдання множини. Одним з них вважається завдання за допомогою характеристичної функції, яка визначається наступним чином (рис. 2.1):

$$\chi(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \quad (2.1)$$

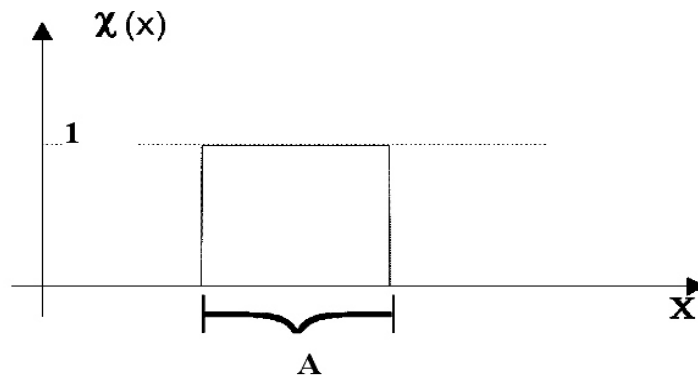


Рис. 2.1. Характеристична функція чіткої множини

Нехай E – універсальна множина, з елементів якого утворюються інші множини, що розглядаються в даному класі задач, наприклад, множини всіх цілих чисел, множини всіх гладких функцій, тощо. Характеристична функція множини $A \subseteq E$ – це функція, значення якої вказують, чи є елемент $x \in E$ елементом множини A . Особливістю цієї функції – в бінарному характері її значень.

Приклад. Для множини A чисел $2 \leq x \leq 4$ характеристична функція має вигляд, представлений рис. 2.2,а.

З точки зору характеристичної функції нечіткої множини є природним узагальненням звичайних множин, коли ми відмовляємося від бінарного характеру цієї функції і припускаємо, що вона може приймати будь-які значення на відрі-

зку $[0,1]$. Множина значень x , на якому визначена функція належності, отримало назву *нечіткої множини*.

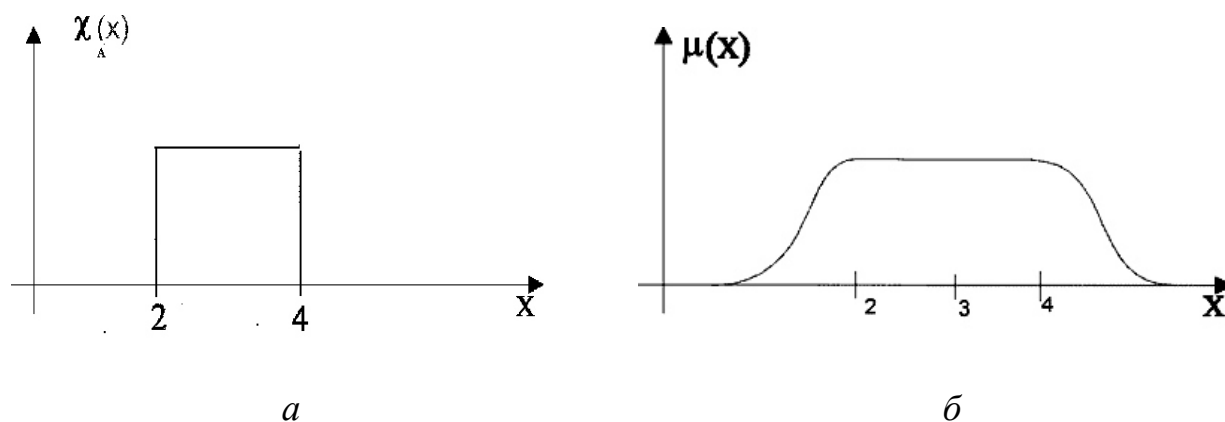


Рис. 2.2. Приклад характеристичної функції чіткої (а) і нечіткої (б) множини

Нехай E – універсальна множина, x – елемент E , а R – деяка властивість. Нечітка підмножина відрізняється від звичайної тим, що для елементів x з E немає однозначної відповіді «так-ні» щодо властивості R . У зв'язку з цим, *нечітка підмножина* A універсальної множини E , елементи якої задовольняють властивості R , визначається як множина впорядкованих пар

$$A = \{\mu_A(x)/x\}, \tag{2.2}$$

де $\mu_A(x)$ – характеристична *функція належності* (або просто функція належності), що приймає значення в деякій цілком впорядкованій множині M (наприклад, $M=[0,1]$).

Функція належності вказує ступінь (або рівень) належності елемента x підмножині A . Множина M є *множиною належностей*. Якщо $M=\{0,1\}$, то нечітка підмножина A може розглядатися як звичайна або чітка множина.

Тепер припущення, що « x приблизно лежить в межах від 2 до 4» (рис. 2.2,а), може бути представлено відповідною функцією належності (рис. 2.2,б).

2.2.2. Основні характеристики нечітких множин.

Нехай $M=[0,1]$ і A – нечітка множина з елементами з універсальної множини E і множиною належностей M .

Величина $\sup_{x \in E} \mu_A(x)$ називається *висотою* нечіткої множини A . Нечітка множина A *нормальна*, якщо її висота дорівнює 1, тобто верхня межа її функції належності дорівнює 1 ($\sup_{x \in E} \mu_A(x) = 1$). При $\sup_{x \in E} \mu_A(x) < 1$ нечітка множина є *субнормальною*.

Нечітка множина *порожня*, якщо $\forall x \in E \mu_A(x) = 0$. Непорожню субнормальну множину можна нормалізувати за формулою:

$$\mu_A(x) := \frac{\mu_A(x)}{\sup_{x \in E} \mu_A(x)} \quad (2.3)$$

Нечітка множина *унімодальна*, якщо $\mu_A(x) = 1$ тільки на одному x з E .

Носієм нечіткої множини A є звичайна підмножина з властивістю $\mu_A(x) > 0$, тобто носій $A = \{x/x \in E, \mu_A(x) > 0\}$.

Елементи $x \in E$, для яких $\mu_A(x) = 0,5$ називаються *точками переходу* множини A .

2.2.3. Функції належності нечітких множин.

Функції належності зручно задавати у параметричній формі. У цьому випадку задача побудови функції належності зводиться до визначення її параметрів. Зазвичай функції належності мають від 2 до 4 параметрів. Найбільш відомі функції належності наведені в таблиці 2.1 і на рис. 2.3.

Популярні функції належності

Таблиця 2.1

Назва функції належності	Аналітична формула	Параметри
Трикутна	$\mu(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases}$	a, c – координати нечіткої множини; b – координата максимуму
Сигмоїдальна	$\mu(x) = \frac{1}{1 + e^{-a(x-c)}}$	a – коефіцієнт крутизни; c – координата переходу через 0,5

Назва функції належності	Аналітична формула	Параметри
Трапецієподібна	$\mu(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}$	a, d – координати нечіткої множини; b, c – координати максимуму
Симетрична Гауса	$\mu(x) = e^{-\frac{(x-b)^2}{2c^2}}$	b – координата максимуму; c – коефіцієнт концентрації
Дзвіноподібна	$\mu(x) = \frac{1}{1 + \left \frac{x-c}{a} \right ^{2b}}$	a – коефіцієнт концентрації; b – коефіцієнт крутизни (обов'язково позитивний, a < b < c); c – координата максимуму;
Z-образна	$\mu(x) = \begin{cases} 1, & x \leq a \\ \text{нелінійна апроксимація, } a < x < b \\ 0, & x \geq b \end{cases}$	a – координата нечіткої множини; b – ядро нечіткої множини
S-образна	$\mu(x) = \begin{cases} 1, & x \leq a \\ \text{нелінійна апроксимація, } a < x < b \\ 0, & x \geq b \end{cases}$	a – координата нечіткої множини; b – ядро нечіткої множини
Двостороння Гауса	<p>якщо $c_1 < c_2$, то</p> $\mu(x) = \begin{cases} \exp\left(\frac{(x-c_1)^2}{-2a_1^2}\right), & x < c_1; \\ 1, & c_1 \leq x \leq c_2; \\ \exp\left(\frac{(x-c_2)^2}{-2a_2^2}\right), & x > c_2; \end{cases}$ <p>якщо $c_1 > c_2$, то</p> $\mu(x) = \begin{cases} \exp\left(\frac{(x-c_1)^2}{-2a_1^2}\right), & x < c_2; \\ 1, & c_2 \leq x \leq c_1; \\ \exp\left(\frac{(x-c_2)^2}{-2a_2^2}\right), & x > c_1; \end{cases}$	c ₁ (c ₂) – мінімальне (максимальне) значення ядра нечіткої множини; a ₁ (a ₂) – коефіцієнт концентрації лівої (правої) частини функції належності
Π-образна	Добуток Z-образної та S-образної функції	a, d – координати нечіткої множини;

Назва функції належності	Аналітична формула	Параметри
		b, c – ядра нечіткої множини
Різниця між двома сигмоїдами	$\mu(x) = \frac{1}{1 + e^{-a_1(x-c_1)}} - \frac{1}{1 + e^{-a_2(x-c_2)}}$	a_1, a_2 – коефіцієнти крутизни; c_1, c_2 – координати переходу через 0,5
Добуток двох сигмоїдів	$\mu(x) = \frac{1}{1 + e^{-a_1(x-c_1)}} \cdot \frac{1}{1 + e^{-a_2(x-c_2)}}$	a_1, a_2 – коефіцієнти крутизни; c_1, c_2 – координати переходу через 0,5

Трикутні (рис. 2.3,а) і трапецієподібні (рис. 2.3,б) функції належності використовуються для завдання таких властивостей множин, які характеризують невизначеність типу: «приблизно дорівнює», «середнє значення», «розташований в інтервалі», «подібний до об'єкту», «схожий на предмет», тощо. Вони також служать для запису нечітких чисел і інтервалів.

Z-образні функції належності (рис. 2.3,к) використовуються для представлення таких властивостей нечітких множин, які характеризуються невизначеністю типу: «мала кількість», «невелике значення», «незначна величина», «низька собівартість продукції», «низький рівень цін або доходів», «низька процентна ставка» і багатьох інших. Загальним для всіх таких ситуацій є слабка ступінь прояву тієї чи іншої якісної або кількісної ознаки. Особливість нечіткого моделювання при цьому полягає в поданні відповідних нечітких множин за допомогою не зростаючих (монотонно убуючих) функцій належності.

S-образні функції належності (рис. 2.3,м) використовуються для подання таких нечітких множин, які характеризуються невизначеністю типу: «велика кількість», «велике значення», «значна величина», «високий рівень доходів і цін», «висока норма прибутку», «висока якість послуг», «високий сервіс обслуговування» і багатьох інших. Загальним для всіх таких ситуацій є висока ступінь прояву тієї чи іншої якісної або кількісної ознаки. Особливість нечіткого

моделювання при цьому полягає в поданні відповідних нечітких множин за допомогою неубуваючих (монотонно зростаючих) функцій належності.

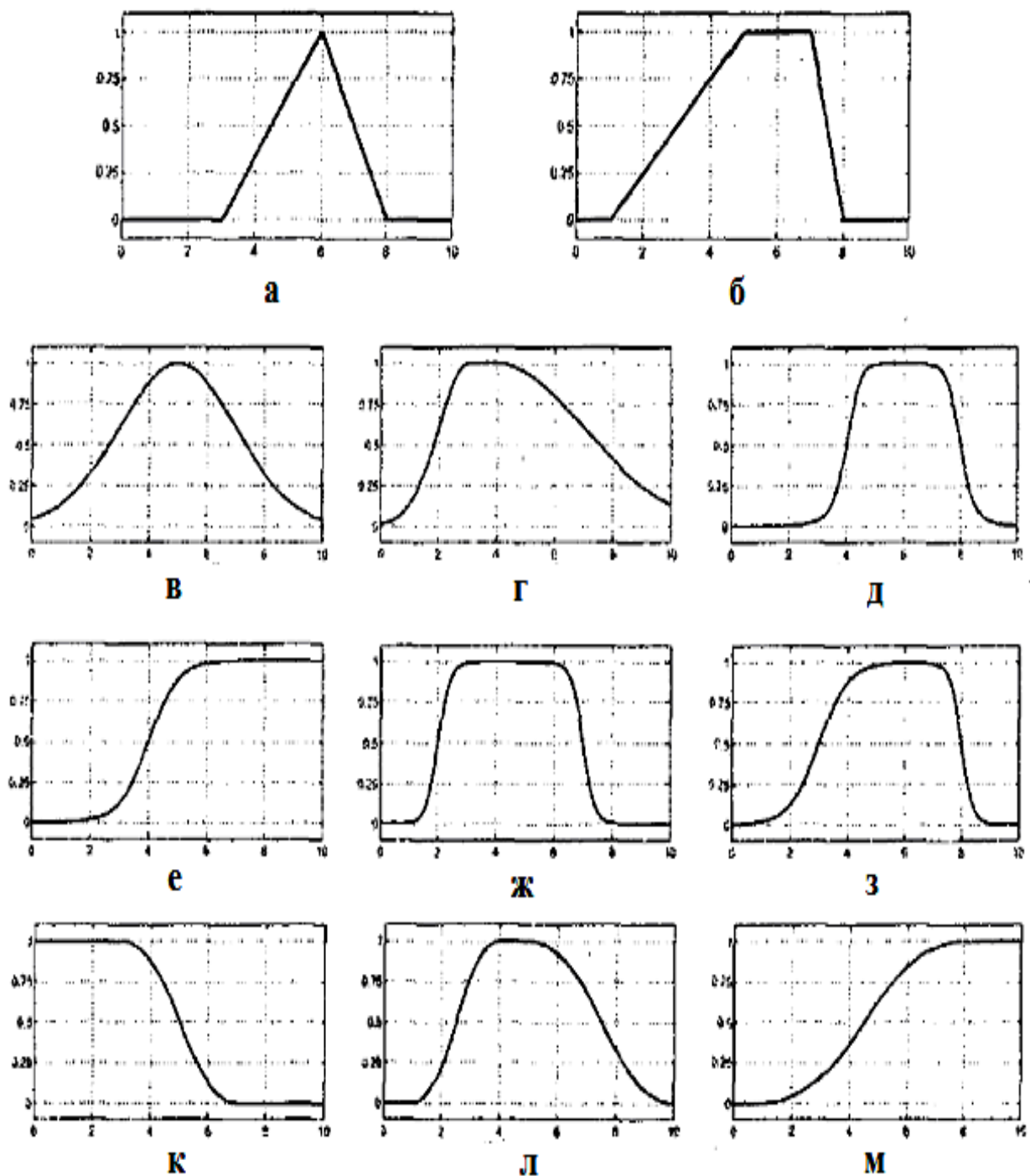


Рис. 2.3. Функції належності нечітких множин: *a* – трикутна, *б* – трапецієподібна, *в* – симетрична Гаусова, *г* – двостороння Гаусова, *д* – дзвіноподібна, *е* – сигмоїдальна, *ж* – різниця між двома сигмоїдами, *з* – добуток двох сигмоїдів, *к* – *Z*-образна, *л* – *Π*-образна, *м* – *S*-образна

2.2.4. Методи побудови функцій належності.

При побудові функцій належності для нечітких множин слід дотримуватись деяких правил, які визначаються характером невизначеності, що має місце при побудові конкретних нечітких моделей. Найбільшого поширення при побудові функцій належності нечітких множин отримали прямі і непрямі методи.

У *прямих методах* експерт або група експертів задають для кожного $x \in X$ значення функції належності $\mu_{\tilde{A}}(x)$. Як правило, прямі методи побудови функцій належності використовуються для таких властивостей, що можуть бути виміряні в деякій кількісній шкалі. Наприклад, такі фізичні величини, як швидкість, час, відстань, тиск, температура та інші мають відповідні одиниці і еталони для свого виміру. При цьому доцільно обмежити розгляд тільки тими значеннями величин, які мають фізичний зміст в контексті розв'язуваної задачі.

При *прямій побудові* функцій належності слід враховувати ту обставину, що теорія нечітких множин не вимагає абсолютно точного завдання функцій належності. Найчастіше буває досить зафіксувати лише найбільш характерні значення і вид (тип) функції належності.

Так, наприклад, якщо необхідно побудувати нечітку множину, яка представляє властивість «швидкість руху автомобіля близько 50 км/год», на початковому етапі може виявитися достатнім представити відповідну нечітку множину трикутною функцією належності f_{Δ} з параметрами $a=40$ км/год, $b=50$ км/год і $c=60$ км/год. Аналогічно, у разі побудови нечіткої множини для подання властивості «швидкість руху автомобіля знаходиться приблизно в межах 50-60 км/год», на початковому етапі може виявитися достатнім представити відповідну нечітку множину трапецієподібною функцією належності f_T з параметрами $a=45$ км/год, $b=50$ км/год, $c=60$ км/год і $d=65$ км/год. У подальшому функція належності може бути уточнена дослідним шляхом на основі аналізу результатів вирішення конкретних завдань.

Процес побудови або завдання нечіткої множини на основі деякого відомого заздалегідь кількісного значення вимірюваної ознаки отримав спеціальну назву – *фазифікація* або приведення до нечіткості. Йдеться про те, що хоча іноді деяке значення вимірної величини є відомим, ми визнаємо той факт, що це значення відомо неточно, можливо з похибкою або випадковою помилкою. При цьому, чим менше ми впевнені в точності вимірювання ознаки, тим більшим буде інтервал носія відповідної нечіткої множини. Слід пам'ятати, що у більшості практичних випадків абсолютна точність вимірювання є лише зручною абстракцією для побудови математичних моделей. Саме з цієї причини фазифікація дозволяє більш адекватно представити об'єктивно присутню неточність результатів фізичних вимірювань.

Непрямі методи визначення значень функції належності, зазвичай, використовуються в тих випадках, коли відсутні очевидні вимірювані властивості, які можуть бути використані для побудови нечітких моделей розглянутої предметної області.

Серед непрямих методів найбільш відомий так званий метод попарних порівнянь. Цей метод використовується для кінцевих нечітких множин і заснований на наступному припущенні. Якби значення шуканої функції належності були відомі і дорівнювали значенням $\mu_{\tilde{A}}(x)$ для $i \in \{1, 2, \dots, n\}$, то попарні порівняння відповідних елементів носія нечіткої множини \tilde{A} можна було б представити у вигляді матриці A з елементами $a_{ij} = \mu_{\tilde{A}}(x_i) / \mu_{\tilde{A}}(x_j)$, де символ "/" позначає операцію ділення.

На практиці буває простіше спочатку побудувати матрицю A в припущенні, що її діагональні елементи повинні бути рівні 1, а симетричні відносно головної діагоналі елементи повинні бути взаємно зворотними, тобто $a_{ij} = 1/a_{ji}$. Остання умова означає, що якщо ступінь належності одного з елементів оцінюється в a разів сильніше ступеня належності іншого, то ступінь належності другого елемента повинна бути в $1/a$ раз сильніше ступеня належності першого елемента.

У цьому випадку задача побудови функції належності зводиться до знаходження такого вектора w , що є рішенням наступного рівняння:

$$A \cdot w = \lambda_{\max} w, \quad (2.4)$$

де λ_{\max} – найбільше власне значення матриці A . Оскільки всі значення елементів матриці A з побудови позитивні, рішення даного рівняння існує і є позитивним.

Власне процес попарного порівняння елементів може бути заснований на суб'єктивній інтуїції або на виконанні певної послідовності алгоритмічних або логічних дій. З алгоритмічних процедур найбільшу популярність здобули методи ітеративного уточнення значень функцій належності, засновані на НМ та генетичних алгоритмах. Логічні процедури використовують методи індуктивного навчання і побудови нечітких метаправил.

Слід зазначити, що у разі нестачі інформації про особливості функцій належності нечітких змінних рекомендується починати побудову нечіткої моделі з використання найбільш простих форм функції належності, а саме – кусочно-лінійних функцій (трикутної, трапецієподібної). Надалі їх характер може бути уточнений і врахований на етапі корекції нечіткої моделі.

2.2.5. Операції над нечіткими множинами.

Перш ніж розглядати операції над нечіткими множинами необхідно зробити ряд зауважень, які будуть сприяти як розумінню суті цих операцій, так і коректному їх застосуванню.

По-перше, необхідно мати на увазі, що нечіткі множини – це узагальнення класичних множин. Оскільки можна допустити найрізноманітніші варіанти подібного узагальнення, то це означає принципову можливість неоднозначності різних визначень, що мають аналогії в класичній теорії множин і представляють практичний інтерес. Стосовно до операцій над нечіткими множинами це означає, що будь-яке визначення тієї чи іншої операції повинно бути справедливо в тому окремому випадку, коли замість нечітких множин використову-

ються класичні. Інакше кажучи, ці визначення повинні перетворюватися в відомі визначення теоретико-множинних операцій, якщо використовувати в них функції належності заміни на характеристичні функції.

По-друге, слід мати на увазі, що порівняння нечітких множин і виконання над ним різних операцій буде можливо тільки тоді, коли відповідні нечіткі множини визначені на одному і тому ж універсумі (універсальній множині E).

По-третє, кажучи про відповідність нечітких множин і функцій належності, слід розуміти цю відповідність у формі математичного ізоморфізму, оскільки одна й та ж функція належності може описувати різні якісні поняття. При цьому, хоча одна й та ж нечітка множина (точніше та чи інша властивість у формі нечіткої множини) може бути представлена різними функціями, що відображають суб'єктивні переваги, з формальної точки зору говорять про різні нечіткі множини.

А. Логічні операції.

Вміщення (домінування). Нехай A і B – нечіткі множини на універсальній множині E . Кажуть, що A міститься в B , якщо $\forall x \in E \mu_A(x) \leq \mu_B(x)$. Операція вміщення позначається таким чином: $A \subset B$.

Якщо використовують термін *домінування*, то у випадку, коли $A \subset B$, кажуть, що « B домінує A ».

Рівність. A та B рівні, якщо $\forall x \in E \mu_A(x) = \mu_B(x)$. Операція рівності позначається таким чином: $A = B$.

Доповнення. Нехай $M = [0, 1]$, A і B – нечіткі множини, що задані на E . A і B доповнюють одна одну, якщо $\forall x \in E \mu_A(x) = 1 \Leftrightarrow \mu_B(x) = 0$. Операція доповнення позначається таким чином: $B = \bar{A}$ або $A = \bar{B}$.

Очевидно, що $\bar{\bar{A}} = A$ (доповнення визначено для $M = [0, 1]$, але очевидно, що його можна визначити для будь-якого впорядкованого M).

Перетин. $A \cap B$ – найбільша нечітка підмножина, що міститься одночасно в A і B :

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)). \quad (2.5)$$

Об'єднання. $A \cup B$ – найменша нечітка підмножина, що включає як A , так і B з функцією належності:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)). \quad (2.6)$$

Різниця. $A \ominus B = A \cap \bar{B}$ з функцією належності:

$$\mu_{A \ominus B}(x) = \mu_{A \cap \bar{B}}(x) = \min(\mu_A(x), 1 \ominus \mu_B(x)). \quad (2.7)$$

Диз'юнктивна сума. $A \oplus B = (A \ominus B) \cup (B \ominus A) = (A \cap \bar{B}) \cup (B \cap \bar{A})$ з функцією належності:

$$\mu_{A \oplus B}(x) = \max(\min(\mu_A(x), 1 \ominus \mu_B(x)), \min(1 \ominus \mu_A(x), \mu_B(x))). \quad (2.8)$$

Для нечітких множин можна будувати візуальне представлення. Розглянемо прямокутну систему координат, на осі ординат якої відкладаються значення $\mu_A(x)$, а на осі абсцис в довільному порядку розташовані елементи E . Якщо E за своєю природою впорядковано, то цей порядок бажано зберегти в розташуванні елементів на осі абсцис. Таке представлення робить наочними прості логічні операції над нечіткими множинами (рис. 2.4).

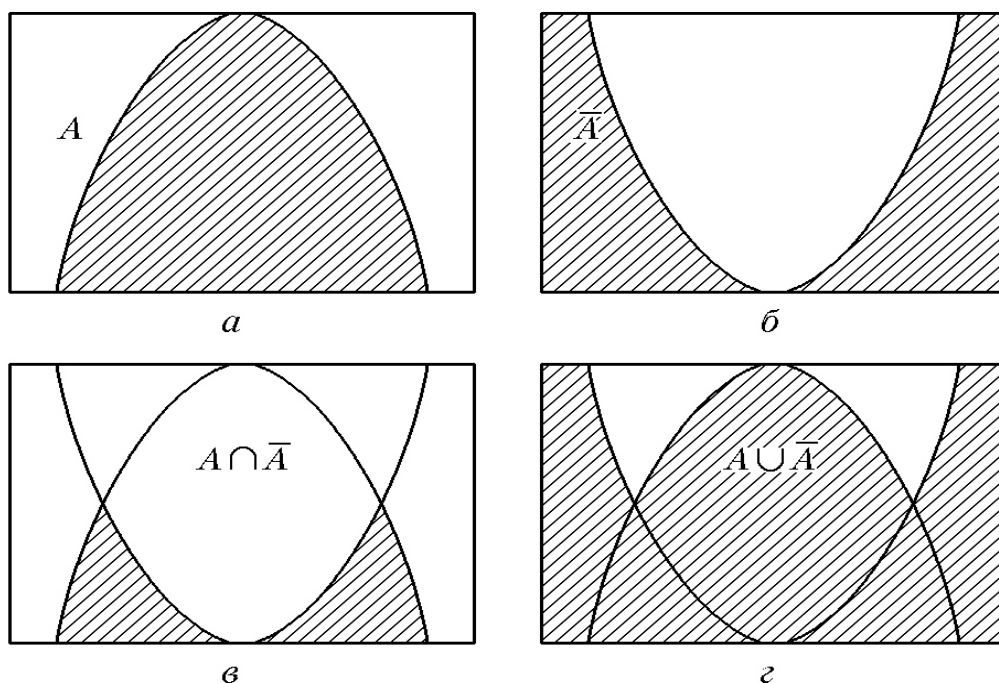


Рис. 2.4. Графічне представлення логічних операцій:

a – нечітка множина A ; $б$ – нечітка множина \bar{A} , $в$ – $A \cap \bar{A}$, $г$ – $A \cup \bar{A}$

Один з підходів до операторів перетину та об'єднання полягає в їх визначенні в класі трикутних норм і конорм.

Трикутною нормою (t-нормою) називається двомісна дійсна функція $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$ з властивостями:

- $T(0, 0)=0$; $T(\mu_A, 1)=\mu_A$; $T(1, \mu_A)=\mu_A$ – обмеженість;
- $T(\mu_A, \mu_B) \leq T(\mu_C, \mu_D)$; якщо $\mu_A \leq \mu_C, \mu_D$; $\mu_B \leq \mu_D$ – монотонність;
- $T(\mu_A, \mu_B)=T(\mu_B, \mu_A)$ – комутативність;
- $T(\mu_A, T(\mu_B, \mu_C))=T(T(\mu_A, \mu_B), \mu_C)$ – асоціативність.

Трикутною конормою (t-конормою) називається двомісна дійсна функція $S: [0, 1] \times [0, 1] \rightarrow [0, 1]$, яка задовольняє таким умовам:

- $S(1, 1)=1$; $S(\mu_A, 0)=\mu_A$; $S(0, \mu_A)=\mu_A$ – обмеженість;
- $S(\mu_A, \mu_B) \geq S(\mu_C, \mu_D)$; якщо $\mu_A \geq \mu_C, \mu_D$; $\mu_B \geq \mu_D$ – монотонність;
- $S(\mu_A, \mu_B)=S(\mu_B, \mu_A)$ – комутативність;
- $S(\mu_A, S(\mu_B, \mu_C))=S(S(\mu_A, \mu_B), \mu_C)$ – асоціативність.

Б. Алгебраїчні операції.

Алгебраїчний добуток множин A і B позначається як $A \cdot B$, і визначається таким чином: $\forall x \in E \mu_{A \cdot B}(x) = \mu_A(x) \cdot \mu_B(x)$.

Алгебраїчна сума множин A і B позначається як $A \hat{+} B$ і визначається таким чином: $\forall x \in E \mu_{A \hat{+} B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x)$.

Зведення в ступінь α нечіткої множини A визначається на основі операції алгебраїчного добутку. Тут α – позитивне число. Нечітка множина A^α визначається функцією належності $\mu_{A^\alpha} = \mu_A^\alpha(x)$.

Окремим випадком зведення в ступінь є операції, що використовуються при роботі з лінгвістичними невизначеностями:

- $\text{CON}(A) = A^2$ – концентрування;
- $\text{DIL}(A) = A^{0,5}$ – розтягування.

Множення на число. Якщо α – позитивне число, таке, що $\alpha \max_{x \in A} \mu_A(x) \leq 1$, то нечітка множина αA має функцію належності: $\mu_{\alpha A}(x) = \alpha \mu_A(x)$.

Опукла комбінація нечітких множин. Нехай A_1, A_2, \dots, A_n – нечіткі множини універсальної множини E , а $\omega_1, \omega_2, \dots, \omega_n$ – ненегативні числа, сума яких дорівнює 1. Опуклою комбінацією A_1, A_2, \dots, A_n називається нечітка множина A з функцією належності $\forall x \in E$:

$$\mu_A(x_1, x_2, \dots, x_n) = \omega_1 \mu_{A_1}(x) + \omega_2 \mu_{A_2}(x) + \dots + \omega_n \mu_{A_n}(x). \quad (2.9)$$

Декартовий добуток нечітких множин. Нехай A_1, A_2, \dots, A_n – нечіткі підмножини відповідних універсальних множин E_1, E_2, \dots, E_n . Декартовий або прямий добуток $A = A_1 \times A_2 \times \dots \times A_n$ є нечіткою підмножиною множини $E = E_1 \times E_2 \times \dots \times E_n$ з функцією належності:

$$\mu_A(x_1, x_2, \dots, x_n) = \min(\mu_{A_1}(x) + \mu_{A_2}(x) + \dots + \mu_{A_n}(x)). \quad (2.10)$$

Оператор збільшення нечіткості використовується для перетворення чітких множин в нечіткі і для збільшення нечіткості нечіткої множини. Нехай A – нечітка множина, E – універсальна множина і для усіх $x \in E$ визначені нечіткі множини $K(x)$. Сукупність усіх $K(x)$ є ядром оператора збільшення нечіткості Φ , результатом дії якого на нечітку множину A є нечітка множина вигляду:

$$\Phi(A, K) = \bigcup_{x \in E} \mu_A(x) K(x), \quad (2.11)$$

де $\mu_A(x)K(x)$ – добуток числа на нечітку множину.

Чітка множина α -рівня (або рівня α). Множиною α -рівня нечіткої множини A універсальної множини E називається чітка підмножина A_α універсальної множини E , яка визначається у вигляді

$$A_\alpha = \{x / \mu_A(x) \geq \alpha\}, \quad \alpha \leq 1. \quad (2.12)$$

2.3. Нечіткі змінні та відношення

2.3.1. Нечіткі та лінгвістичні змінні.

При описі об'єктів і явищ за допомогою нечітких множин використовуються поняття нечіткої та лінгвістичної змінних.

Нечітка змінна характеризується набором (α, X, A) , де

α – найменування змінної;

X – універсальна множина (область визначення α);

A – нечітка множина на X , що описує обмеження (тобто $\mu_A(x)$) на значення нечіткої змінної α .

Лінгвістична змінна характеризується набором (β, T, X, G, M) , де

β – найменування лінгвістичної змінної;

T – множина її значень (*терм-множина*), що являє собою найменування нечітких змінних, областю визначення кожної з яких є множина X . Множина T є базовою *терм-множиною* лінгвістичної змінної;

G – синтаксична процедура, що дозволяє оперувати елементами терм-множини T , зокрема, генерувати нові терми (значення). Множина $T \cup G(T)$, де $G(T)$ – множина згенерованих термів, є розширеною терм-множиною лінгвістичної змінної;

M – семантична процедура, що дозволяє перетворити кожне нове значення лінгвістичної змінної, утворене процедурою G , в нечітку змінну, тобто сформувати відповідну нечітку множину.

Слід зазначити, що для того, щоб уникнути великої кількості символів:

1) символ β використовують як для назви самої змінної, так і для всіх її значень;

2) користуються одним і тим же символом для позначення нечіткої множини і його назви, наприклад терм «Молодий», який є значенням лінгвістичної змінної $\beta =$ «вік», одночасно є і нечіткою множиною M («Молодий»).

Присвоєння кількох значень символам передбачає, що контекст дозволяє вирішити можливі невизначеності.

2.3.2. Нечіткі числа.

Нечіткі числа – це нечіткі змінні, які визначені на числовій вісі. Тобто нечітке число визначається як нечітка множина A на множині дійсних чисел R з функцією належності $\mu_A(x) \in [0,1]$, де x – дійсне число, тобто $x \in R$.

Нечітке число A нормальне, якщо $\max_x \mu_A(x)=1$, опукле, якщо для будь-яких $x \leq y \leq z$ виконується

$$\mu_A(x) \geq \mu_A(y) \wedge \mu_A(z). \quad (2.13)$$

Множина α -рівня нечіткого числа A визначається як

$$A_\alpha = \{x / \mu_A(x) \geq \alpha\}. \quad (2.14)$$

Підмножина $S_A \subset R$ є носієм нечіткого числа A , якщо

$$S_A = \{x / \mu_A(x) > 0\}. \quad (2.15)$$

Нечітке число A унімодальне, якщо умова $\mu_A(x)=1$ справедливо тільки для однієї точки дійсної вісі.

Опукле нечітке число A є нечітким нулем, якщо

$$\mu_A(0) = \sup_x (\mu_A). \quad (2.16)$$

Нечітке число A позитивне, якщо $\forall x \in S_A, x > 0$ і негативне при $x < 0$.

2.3.3. Нечіткі відношення та операції над ними.

Нехай $E = E_1 \times E_2 \times \dots \times E_n$ – прямий добуток універсальних множин і M – деяка множина належностей (наприклад, $M=[0, 1]$). Нечітке n -не відношення визначається як нечітка підмножина R на E , що приймає свої значення в M .

Якщо $n=2$ і $M=[0, 1]$ нечітким відношенням R між множинами $X=E_1$ і $Y=E_2$ буде називатися функція $R: (X, Y) \rightarrow [0, 1]$, яка ставить у відповідність кожній парі елементів $(x, y) \in X \times Y$ величину $\mu_R(x, y) \in [0, 1]$.

Нечітке відношення на $X \times Y$ позначається у вигляді $x \in X, y \in Y: xRy$.

Якщо $X=Y$, тобто X і Y збігаються, нечітке відношення $R: X \times X \rightarrow [0, 1]$ є нечітким відношенням на множині X .

Взагалі, нечіткі відношення дозволяють не тільки зафіксувати сам факт відношення, а й вказувати ступінь його виконання, що є важливим для багатьох практичних задач.

Розглянемо операції над нечіткими відношеннями.

Об'єднання двох відношень R_1 і R_2 позначається як $R_1 \cup R_2$ і визначається наступним виразом:

$$\mu_{R_1 \cup R_2}(x, y) = \mu_{R_1}(x, y) \vee \mu_{R_2}(x, y). \quad (2.17)$$

Перетин двох відношень R_1 і R_2 позначається як $R_1 \cap R_2$ і визначається наступним виразом:

$$\mu_{R_1 \cap R_2}(x, y) = \mu_{R_1}(x, y) \wedge \mu_{R_2}(x, y). \quad (2.18)$$

Алгебраїчний добуток двох відношень R_1 і R_2 позначається як $R_1 \cdot R_2$ і визначається наступним виразом:

$$\mu_{R_1 \cdot R_2}(x, y) = \mu_{R_1}(x, y) \cdot \mu_{R_2}(x, y). \quad (2.19)$$

Алгебраїчна сума двох відношень R_1 і R_2 позначається як $R_1 \hat{+} R_2$ і визначається наступним виразом:

$$\mu_{R_1 \hat{+} R_2}(x, y) = \mu_{R_1}(x, y) + \mu_{R_2}(x, y) \Leftrightarrow \mu_{R_1}(x, y) \cdot \mu_{R_2}(x, y). \quad (2.20)$$

Доповнення відношення R позначається як \bar{R} і визначається наступною функцією належності:

$$\mu_{\bar{R}}(x, y) = 1 \Leftrightarrow \mu_R(x, y). \quad (2.21)$$

Диз'юнктивна сума двох відношень R_1 і R_2 позначається як $R_1 \oplus R_2$ і визначається наступним виразом:

$$R_1 \oplus R_2 = (R_1 \cap \bar{R}_2) \cup (\bar{R}_1 \cap R_2). \quad (2.22)$$

Звичайне відношення, найближче до нечіткого. Нехай R – нечітке відношення з функцією належності $\mu_R(x, y)$. Звичайне відношення, найближче до нечіткого, позначається \underline{R} і визначається наступним виразом:

$$\mu_{\underline{R}}(x, y) = \begin{cases} 0, & \text{якщо } \mu_R(x, y) < 0,5, \\ 1, & \text{якщо } \mu_R(x, y) > 0,5, \\ 0 \text{ або } 1, & \text{якщо } \mu_R(x, y) = 0,5. \end{cases} \quad (2.23)$$

За домовленістю приймають $\mu_{\underline{R}}(x, y) = 0$ при $\mu_R(x, y) = 0,5$.

Композиція (згортка) двох нечітких відношень. Нехай R_1 – нечітке відношення $R_1: (X \times Y) \rightarrow [0, 1]$ між X і Y , і R_2 – нечітке відношення $R_2: (Y \times Z) \rightarrow [0, 1]$ між Y і Z . Нечітке відношення між X і Z , що позначається $R_2 \square R_1$, визначається через R_2 і R_1 наступним виразом:

$$\mu_{R_1 \circ R_2}(x, z) = \bigvee_y (\mu_{R_1}(x, y) \wedge \mu_{R_2}(y, z)) \quad (2.24)$$

та називається (max-min)-композицією ((max-min)-згорткою) відношень R_1 і R_2 .

2.4. Системи нечіткого висновку

2.4.1. Нечіткі висновки.

Механізм нечітких висновків зазвичай використовується в експертних і керуючих системах. У своїй основі він має базу знань, що формується фахівцями-експертами предметної області у вигляді сукупності нечітких предикатних правил виду:

Правило 1: якщо $x \in A_1$, тоді $y \in B_1$,

Правило 2: якщо $x \in A_2$, тоді $y \in B_2$,

...

Правило N : якщо $x \in A_n$, тоді $y \in B_n$,

(2.25)

де x – вхідна змінна, y – змінна висновку (ім'я для значення даних, яке буде обчислено); A і B – функції належності, визначені відповідно на x і y .

Знання експерта $A \rightarrow B$, що відбиває нечітке причинне відношення передумови і висновків, можна назвати нечітким відношенням R :

$$R = A \rightarrow B, \quad (2.26)$$

де " \rightarrow " називають нечіткою імплікацією.

Відношення R можна розглядати як нечітку підмножину прямого добутку $X \times Y$ повної множини передумов X і висновків Y . Таким чином, процес отримання нечіткого результату виведення B' з використанням спостереження A' і знання $A \rightarrow B$ можна відобразити наступним чином:

$$B' = A' R = A' (A \rightarrow B). \quad (2.27)$$

Таким чином, у загальному випадку нечіткий логічний висновок здійснюється в наступні етапи.

1. *Нечіткість* (введення нечіткості, фаззифікація, fuzzification). Функції належності, що визначені на вхідних змінних, застосовуються до їх фактичних значень, для того щоб визначити ступінь істинності кожної передумови кожного правила.

2. *Логічний висновок*. Обчислення значення істинності для передумов кожного правила застосовується до висновків кожного правила. Це призводить до однієї нечіткої підмножини, що буде призначена кожній змінній висновку для кожного правила. За правила логічного висновку, зазвичай, використовуються тільки операції *min* (мінімум) або *prod* (множення). При операції *min* функція належності виведення «відсікається» по висоті, що відповідає обчисленому ступеню істинності передумови правила (нечітка логічна операція «І»). При операції *prod* функція належності виведення масштабується за обчисленим ступенем істинності передумови правила.

3. *Композиція*. Всі нечіткі підмножини, призначені до кожної змінної висновку (у всіх правилах), об'єднуються разом, щоб сформувати одну нечітку підмножину для кожної змінної виводу. При подібному об'єднанні зазвичай використовуються операції *max* (максимум) або *sum* (сума). При операції *max* комбінований висновок нечіткої підмножини конструюється як поточний максимум за всіма нечіткими підмножинами (нечітка логічна операція «АБО»). У разі операції *sum* комбінований висновок нечіткої підмножини конструюється як поточна сума за всіма нечіткими підмножинами, призначеними змінній виводу правилами логічного висновку.

4. *Приведення до чіткості* (дефаззифікація, defuzzification) використовується для перетворення нечіткого набору висновків в чітке число.

Слід зазначити, що вибір конкретних способів реалізації окремих етапів нечіткого висновку визначається тим або іншим алгоритмом нечіткого висновку.

2.4.2. Алгоритм Мамдані (Mamdani).

Одним із перших алгоритмів, який знайшов застосування в системах нечіткого висновку є алгоритм Мамдані (Mamdani), що був запропонований у 1975 р. англійським математиком Е. Мамдані (E. Mamdani) як метод для керування поршневим двигуном. Цей алгоритм, заснований на нечіткому логічному висновку, дозволив уникнути надмірно великого обсягу обчислень і був позитивно оцінений фахівцями.

Суть алгоритму Мамдані полягає у наступному. Припустимо, що в базі знань містяться тільки два нечітких правила виду:

Правило 1: якщо $x \in A_1$ і $y \in B_1$, тоді $z \in C_1$,

Правило 2: якщо $x \in A_2$ і $y \in B_2$, тоді $z \in C_2$, (2.28)

де x, y – вхідні змінні, z – змінної виводу, $A_1, A_2, B_1, B_2, C_1, C_2$ – деякі задані функції належності. При цьому, на основі наведеної інформації та чітких значень x_0, y_0 необхідно визначити чітке значення z_0 .

Алгоритм Мамдані може бути описаний наступним чином (рис. 2.5):

Етап 1. *Нечіткість*: знаходяться ступені істинності для передумов кожного правила: $A_1(x_0), A_2(x_0), B_1(y_0), B_2(y_0)$.

Етап 2. *Нечіткий висновок*. Знаходяться рівні «відсікання» для передумов кожного з правил (з використанням операції *min*): $\alpha_1 = A_1(x_0) \wedge B_1(y_0)$, $\alpha_2 = A_2(x_0) \wedge B_2(y_0)$, де \wedge – операція логічного мінімуму (*min*).

Далі знаходяться «усічені» функції належності: $C'_1(z) = (\alpha_1 \wedge C_1(z))$, $C'_2(z) = (\alpha_2 \wedge C_2(z))$.

Етап 3. *Композиція*: з використання операції максимум (*max*) проводиться об'єднання знайдених усічених функцій, що призводить до отримання підсумкової нечіткої підмножини для змінної виходу з функцією належності:

$$\mu_z(z) = C(z) = C'_1(z) \vee C'_2(z) = (\alpha_1 \wedge C_1(z)) \vee (\alpha_2 \wedge C_2(z)), \quad (2.29)$$

де \vee – операція логічного максимуму (*max*).

Етап 4. *Приведення до чіткості* (для знаходження z_0) проводиться, як правило, центроїдним методом: чітке значення вихідної змінної визначається як центр ваги для кривої $\mu_\Sigma(z)$:

$$z_0 = \frac{\int_{\Omega} z \mu_\Sigma(z) dz}{\int_{\Omega} \mu_\Sigma(z) dz} \quad (2.30)$$

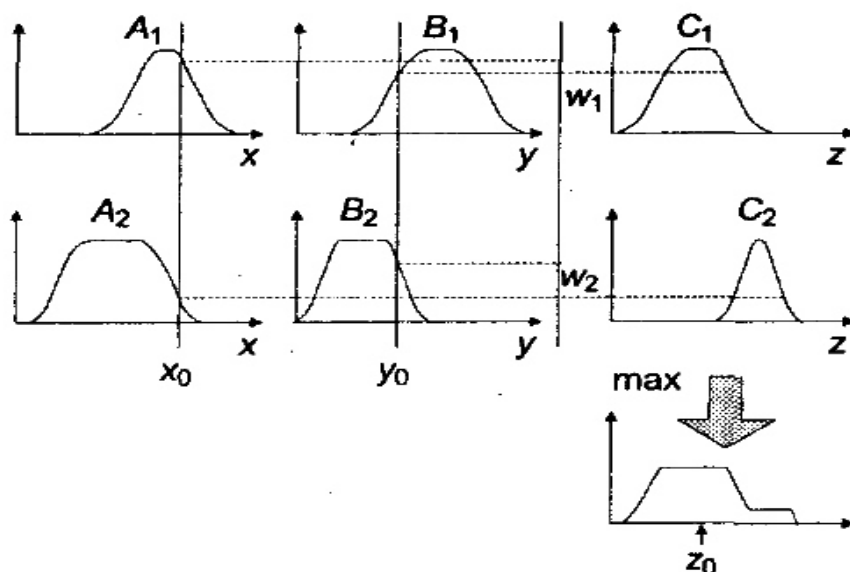


Рис. 2.5. Ілюстрація нечіткого висновку Мамдані

2.4.3. Алгоритм Сугено (Sugeno).

Суть алгоритму Сугено полягає у наступному. Припустимо, що в базі знань містяться тільки два нечітких правила виду:

Правило 1: якщо $x \in A_1$ і $y \in B_1$, тоді $z_1 = a_1 x + b_1 y$,

Правило 2: якщо $x \in A_2$ і $y \in B_2$, тоді $z_2 = a_2 x + b_2 y$. (2.31)

Алгоритм Сугено може бути описаний наступним чином (рис. 2.6):

Етап 1. Знаходяться ступені істинності для передумов кожного правила:

$A_1(x_0), A_2(x_0), B_1(y_0), B_2(y_0)$.

Етап 2. Знаходяться рівні «відсікання» для передумов кожного з правил (з

використанням операції *min*): $\alpha_1 = A_1(x_0) \wedge B_1(y_0)$, $\alpha_2 = A_2(x_0) \wedge B_2(y_0)$,

а також індивідуальні виходи правил: $z_1^* = a_1 x_0 + b_1 y_0$, $z_2^* = a_2 x_0 + b_2 y_0$.

Етап 3. Визначається чітке значення вихідної змінної:

$$z_0 = \frac{\alpha_1 z_1^* + \alpha_2 z_2^*}{\alpha_1 + \alpha_2}. \quad (2.32)$$

Наведене представлення відноситься до алгоритму Сугено 1-го порядку.

Разом з вищеописаним використовується також алгоритм Сугено 0-го порядку, у якому правила записані у формі:

Правило 1: якщо $x \in A_1$ і $y \in B_1$, тоді $z_1 = c_1$,

Правило 2: якщо $x \in A_2$ і $y \in B_2$, тоді $z_2 = c_2$. (2.33)

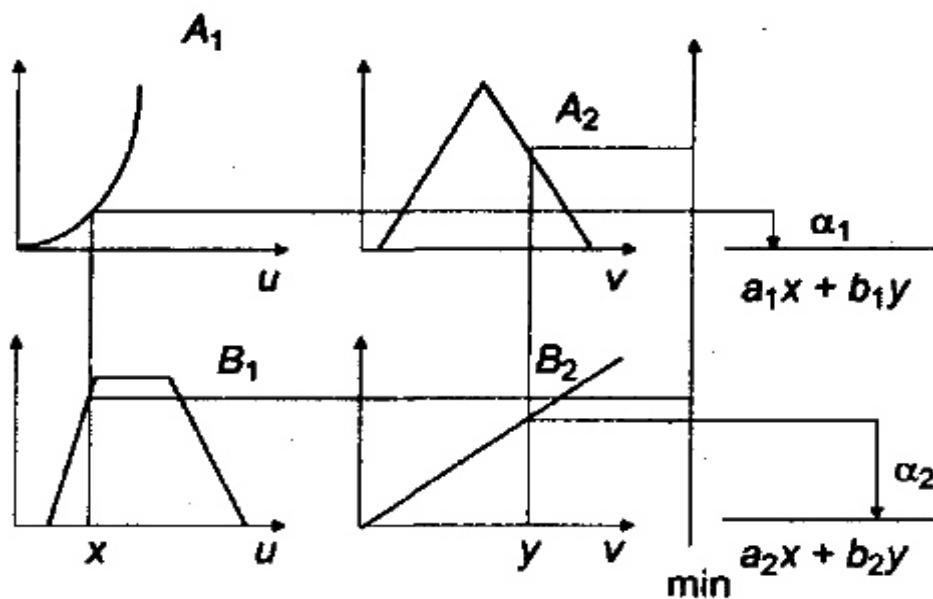


Рис. 2.6. Ілюстрація нечіткого висновку Сугено 1-го порядку

2.4.4. Алгоритм Цукамото (Tsukamoto).

Суть алгоритму Цукамото полягає у наступному. Припустимо, що в базі знань містяться тільки два нечітких правила виду

Правило 1: якщо $x \in A_1$ і $y \in B_1$, тоді $z \in C_1$,

Правило 2: якщо $x \in A_2$ і $y \in B_2$, тоді $z \in C_2$, (2.34)

але, на відміну від алгоритму Мамдані, тут функції $C_1(z)$, $C_2(z)$ є монотонними.

Алгоритм Цукамото може бути описаний наступним чином (рис.2.7):

Етап 1. Знаходяться ступені істинності для передумов кожного правила:

$A_1(x_0)$, $A_2(x_0)$, $B_1(y_0)$, $B_2(y_0)$.

Етап 2. Знаходяться рівні «відсікання» для передумов кожного з правил (з використанням операції \min): $\alpha_1 = A_1(x_0) \wedge B_1(y_0)$, $\alpha_2 = A_2(x_0) \wedge B_2(y_0)$, а потім – за допомогою вирішення рівнянь $\alpha_1=C_1(z)$, $\alpha_2=C_2(z)$ – чіткі значення (z_1 і z_2) для кожного з вихідних правил.

3. Визначається чітке значення змінної виводу (як зважене середнє z_1 і z_2):

$$z_0 = \frac{\alpha_1 z_1 + \alpha_2 z_2}{\alpha_1 + \alpha_2}; \quad (2.35)$$

або у загальному випадку (дискретний варіант центроїдного методу):

$$z_0 = \frac{\sum_{i=1}^n \alpha_i z_i}{\sum_{i=1}^n \alpha_i}. \quad (2.36)$$

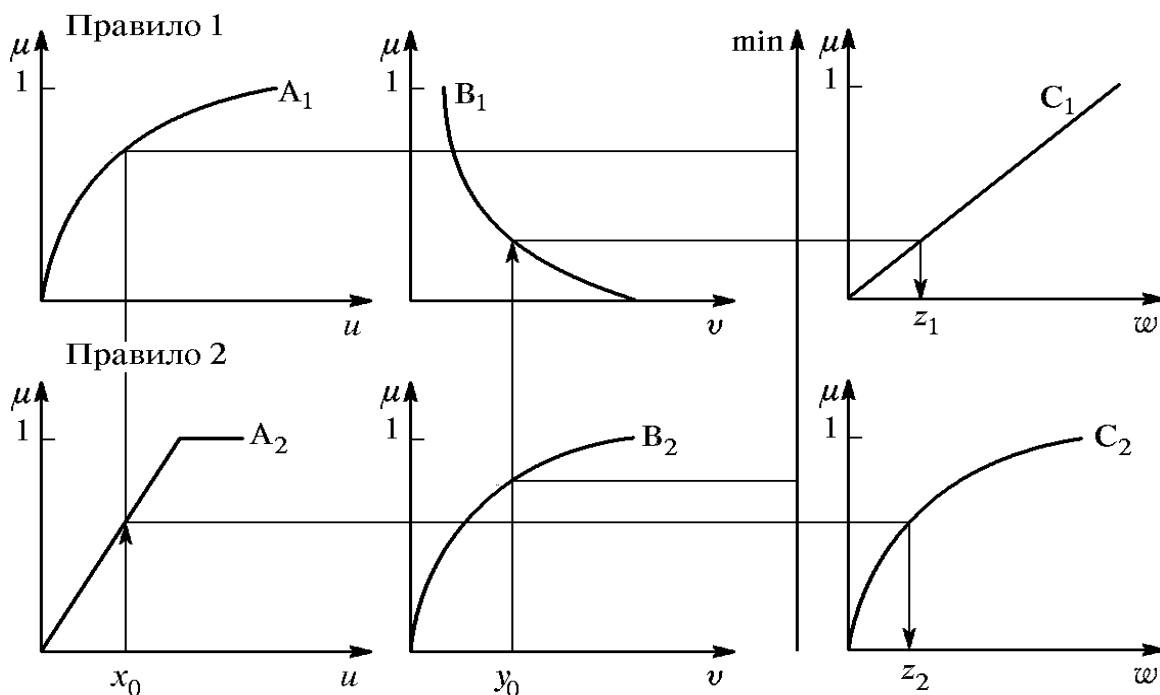


Рис. 2.7. Ілюстрація нечіткого висновку Цукамото

2.4.5. Алгоритм Ларсена (Larsen).

В алгоритмі Ларсена нечітка імплікація моделюється з використанням оператора множення.

Алгоритм Ларсена може бути описаний наступним чином (рис. 2.8):

Етап 1. Знаходяться ступені істинності для передумов кожного правила:
 $A_1(x_0), A_2(x_0), B_1(y_0), B_2(y_0)$.

Етап 2. Знаходяться рівні «відсікання» для передумов кожного з правил:
 $\alpha_1 = A_1(x_0) \wedge B_1(y_0), \alpha_2 = A_2(x_0) \wedge B_2(y_0)$, а потім – приватні нечіткі підмножини
 $\alpha_1 C_1(z)$ та $\alpha_2 C_2(z)$.

Етап 3. Знаходиться підсумкова нечітка підмножина з функцією належності

$$\mu_{\Sigma}(z) = C(z) = (\alpha_1 C_1(z)) \vee (\alpha_2 C_2(z)), \quad (2.37)$$

(у загальному випадку n правил $\mu_{\Sigma}(z) = C(z) = \bigvee_{i=1}^n (\alpha_i C_i(z))$).

Етап 4. Приведення до чіткості (аналогічно до розглянутих раніше алгоритмів).

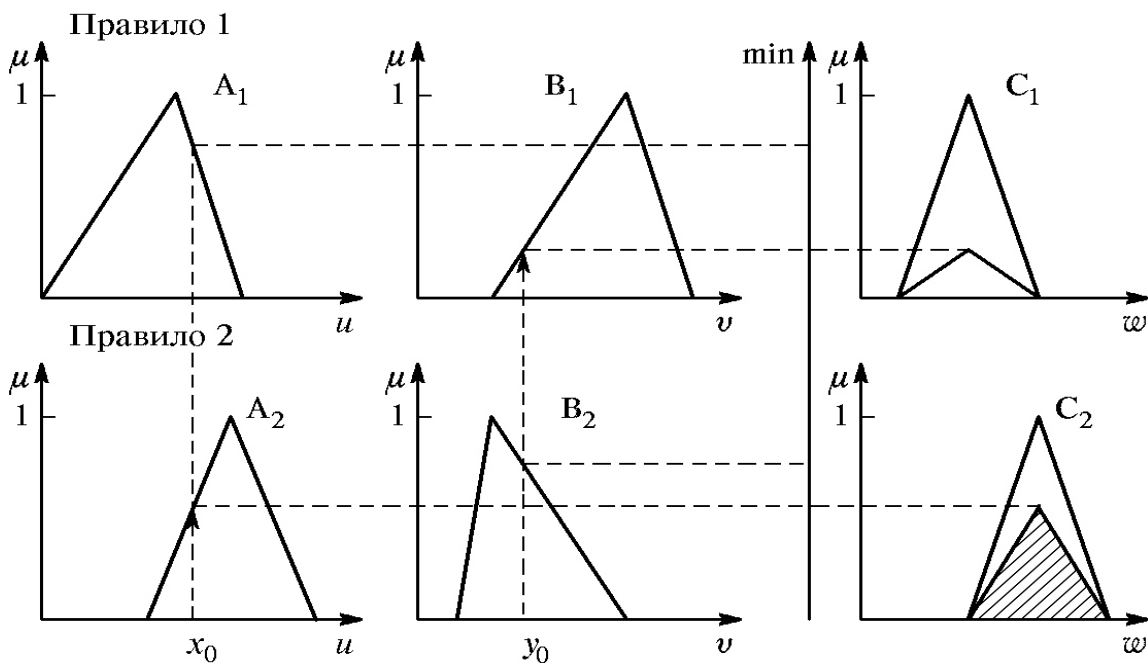


Рис. 2.8. Ілюстрація нечіткого висновку Ларсена

Слід зазначити, що результати нечіткого висновку, які отримані за допомогою різних алгоритмів, можуть відрізнятися. Певної переваги того чи іншого методу дослідники в області нечітких систем не виділяють. Подібне можна сказати й про способи приведення до чіткості.

2.4.6. Методи приведення до чіткості.

1. *Центроїдний метод* (використовується в алгоритмі Мамдані) для безперервного варіанту має вигляд:

$$z_0 = \frac{\int_{\Omega} z \cdot C(z) dz}{\int_{\Omega} C(z) dz}; \quad (2.38)$$

а для дискретного варіанту:

$$z_0 = \frac{\sum_{i=1}^n \alpha_i z_i}{\sum_{i=1}^n \alpha_i}. \quad (2.39)$$

2. *Перший максимум*. Чітка величина змінної висновку знаходиться як найменше значення, при якому досягається максимум підсумкової нечіткої множини (рис. 2.9,а):

$$z_0 = \min \{z \mid C(z) = \max_u C(u)\}. \quad (2.40)$$

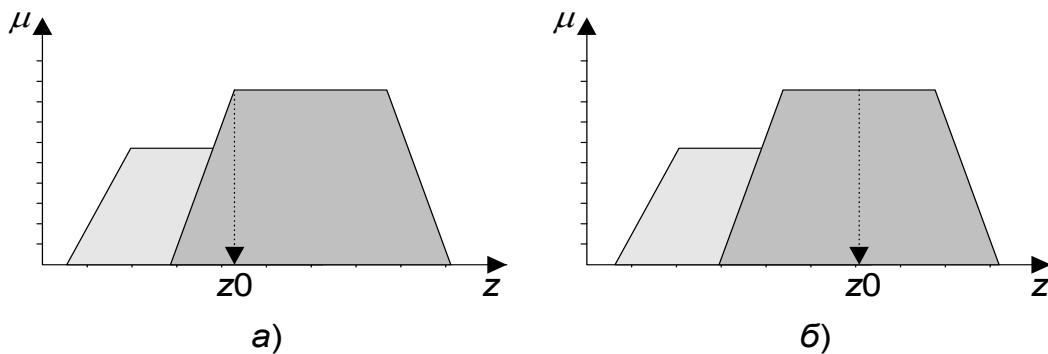


Рис. 2.9. Ілюстрація методів приведення до чіткості:

a – перший максимум; *б* – середній максимум

3. *Середній максимум*. Чітке значення знаходиться за формулою

$$z_0 = \frac{\int_G z dz}{\int_G dz}, \quad (2.41)$$

де G – підмножина елементів, що максимізують C (рис. 2.9,б).

Дискретний варіант (якщо C – дискретно):

$$z_0 = \frac{1}{N} \sum_{j=1}^N z_j. \quad (2.42)$$

4. *Критерій максимуму.* Чітке значення обирається довільно серед множини елементів, що доставляють максимум C , тобто

$$z_0 \in \{z \mid C(z) = \max_u C(u)\}. \quad (2.43)$$

5. *Висотна дефазифікація.* Елементи області визначення Ω , для яких значення функції належності менше, ніж деякий рівень α , в розрахунок не приймаються і чітке значення розраховується за формулою:

$$z_0 = \frac{\int_{C_\alpha} z \cdot C(z) dz}{\int_{C_\alpha} C(z) dz}, \quad (2.44)$$

де C_α – нечітка множина α -рівня.

2.5. Нечітка кластеризація

2.5.1. Кластерний аналіз.

Кластерний аналіз призначений для розбиття множини об'єктів на задане або невідоме число кластерів на підставі деякого математичного критерію якості класифікації (від англ. «cluster» – пучок, скупчення, група елементів, що характеризуються будь-якою загальною властивістю). Критерій якості кластеризації в тій чи іншій мірі відображає наступні неформальні вимоги:

- а) в середині кластера об'єкти повинні бути тісно пов'язані між собою;
- б) об'єкти різних кластерів повинні бути далекі один від одного;

в) за інших рівних умов розподілу об'єктів по кластерам повинні бути рівномірними.

Вимоги а) і б) відображають стандартну концепцію компактності класів розбиття; вимога в) полягає у тому, щоб критерій не нав'язував об'єднання окремих груп об'єктів.

Вузловим моментом в кластерному аналізі вважається вибір метрики (або міри близькості об'єктів), від якого залежить остаточний варіант розбиття об'єктів на групи (кластери) при заданому алгоритмі розбиття. У кожній конкретній задачі цей вибір проводиться по різному, з урахуванням головних цілей дослідження, фізичної та статистичної природи використовуваної інформації тощо.

Іншою важливою величиною в кластерному аналізі є відстань між цілими групами об'єктів, що характеризують взаємне розташування окремих груп об'єктів. Нехай w_i – i -а група (клас, кластер) об'єктів, N_i – число об'єктів, що утворюють групу w_i , вектор μ_i – середнє арифметичне об'єктів, що входять в w_i (або μ_i – «центр ваги» i -ї групи), $q(w_l, w_m)$ – відстань між групами w_l і w_m .

Найбільш поширеними відстанями між групами об'єктів є (рис. 2.10):

- *відстань центрів тяжіння* – відстань між центральними точками кластерів:

$$q(w_l, w_m) = d(\mu_l, \mu_m); \quad (2.45)$$

- *відстань найближчого сусіда* – відстань між найближчими об'єктами кластерів:

$$q_{\min}(w_l, w_m) = \min_{x_i \in w_l, x_j \in w_m} d(x_i, x_j); \quad (2.46)$$

- *відстань дальнього сусіда* – відстань між найбільш далекими об'єктами кластерів:

$$q_{\max}(w_l, w_m) = \max_{x_i \in w_l, x_j \in w_m} d(x_i, x_j). \quad (2.47)$$

Узагальнена (за Колмогоровим) відстань між класами, або узагальнена К-відстань, обчислюється наступним чином:

$$q_{\tau}^{(K)}(w_l, w_m) = \left[\frac{1}{N_l N_m} \sum_{x_i \in w_l} \sum_{x_j \in w_m} d^{\tau}(x_i, x_j) \right]^{\frac{1}{\tau}}. \quad (2.48)$$

Зокрема, при $\tau \rightarrow \infty$ і при $\tau \rightarrow -\infty$ маємо:

$$q_{\infty}^{(K)}(w_l, w_m) = q_{\max}(w_l, w_m); \quad (2.49)$$

$$q_{-\infty}^{(K)}(w_l, w_m) = q_{\min}(w_l, w_m). \quad (2.50)$$

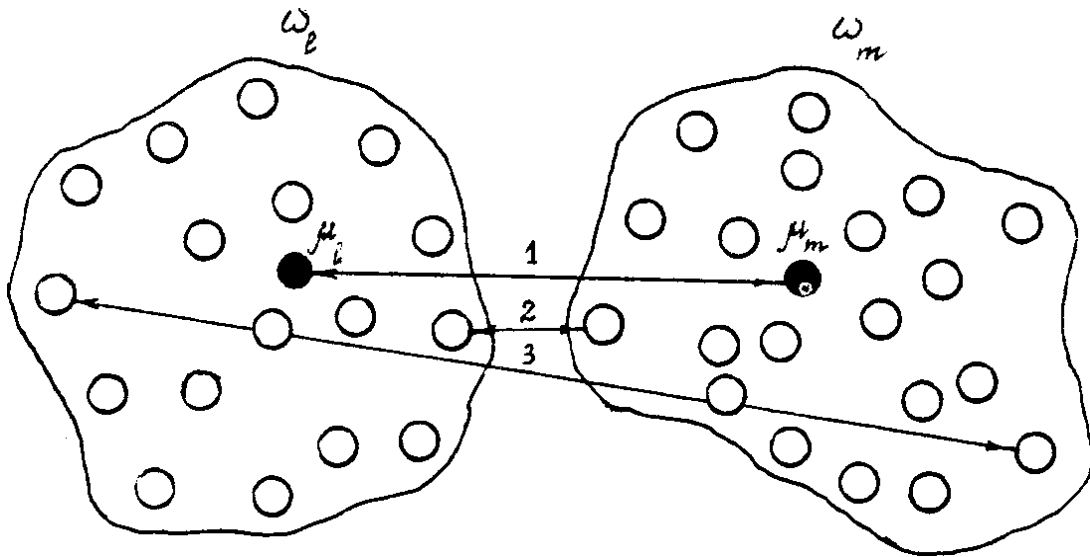


Рис. 2.10. Ілюстрація способів визначення відстані між кластерами w_l і w_m :

1 – по центрам тяжіння, 2 – по найближчим об'єктам, 3 – по далеким об'єктам

Вибір тієї чи іншої міри відстані між кластерами впливає, головним чином, на вигляд геометричних угруповань об'єктів в просторі ознак, які виділяються алгоритмами кластерного аналізу. Так, алгоритми, засновані на відстані найближчого сусіда, добре працюють в разі угруповань, що мають складну, зокрема, ланцюгову структуру. Відстань далекого сусіда застосовується, коли шукані угруповання утворюють в просторі ознак кулясті хмари. Щодо алгоритмів, які використовують відстані центрів тяжіння і середнього зв'язку, вони найкраще працюють у разі угруповань еліпсоїдної форми.

Існує багато методів кластеризації, які можна класифікувати на чіткі й нечіткі. Чіткі методи кластеризації розбивають початкову множину об'єктів X на кілька підмножин, що не перетинаються. При цьому будь-який об'єкт з X на-

лежить тільки одному кластеру. Нечіткі методи кластеризації дозволяють одному й тому ж об'єкту належати одночасно декільком (або навіть усім) кластерам, але з різним ступенем зв'язку. Таким чином, нечітка кластеризація в багатьох ситуаціях більш «природна», ніж чітка, наприклад, для об'єктів, розташованих на кордоні кластерів.

Методи кластеризації також класифікуються за тим, чи визначено кількість кластерів заздалегідь чи ні. В останньому випадку кількість кластерів визначається в ході виконання алгоритму на основі розподілу вхідних даних.

Найбільш відомими методами нечіткої кластеризації є: субтрактивна кластеризація (Subtractive Clustering) – поліпшена версія методу гірської кластеризації та нечітка кластеризація *C*-середніх (Fuzzy *C*-means).

2.5.2. Субтрактивна кластеризація.

В основі методу субтрактивної кластеризації лежить припущення, що кожна експериментальна точка може бути центром кластеру.

При субтрактивній кластеризації генерується система нечіткого логічного висновку типу Сугено. Екстракція правил з даних відбувається в два етапи. Спочатку визначається кількість правил і потужностей терм-множин вихідних змінних. Далі за допомогою методу найменших квадратів визначається «то-» частина кожного правила. Результатом є система нечіткого логічного висновку з базою правил, що охоплюють всю предметну область.

Алгоритм субтрактивної кластеризації може бути представлений наступним чином:

1. Розрахувати потенціалу кожної точки x_k (як міри просторової близькості між нею та іншими):

$$E(x_k) = \frac{1}{K} \sum_{i=1}^K e^{-\frac{\|x_k - x_i\|}{(R_c/2)^2}}, \quad (2.51)$$

де R_c – позитивне число, яке представляє собою радіус центру кластера, K – число точок даних в навчальній послідовності.

2. Встановити кількість кластерів $k_c=0$.

3. Виявити точку даних з найвищим потенціалом $E(x_p)$, x_p :

$$p = \arg \max_{i=1}^K E(x_i). \quad (2.52)$$

4. Встановити j -й центр кластера:

$$k_{c_j} = x_p, \quad (2.53)$$

при цьому $E(j_1)$ – його потенціал, $j=j+1$ – приріст.

5. Знизити потенціал всіх точок:

$$E(x_i) = E(x_i) - E(k_{c_j}) e^{-\frac{\|k_{c_j} - x_i\|}{(r/2)^2}}, \quad (2.54)$$

де $r=[1,1.5] R_c$ – позитивна постійна, що визначає діапазон впливу одного кластера; $i=1, \dots, K$.

6. Перевірити значення потенціалу точок відносно встановленого порогу thr :

$$\max_{i=1}^K E(x_i) < thr. \quad (2.55)$$

Якщо умова (2.55) виконується, то кінець алгоритму, інакше – перехід до пункту 3.

В алгоритмі субтрактивної кластеризації радіуси кластерів визначають наскільки далеко від центру кластера можуть бути його елементи. Слід зазначити, що вибір радіусу може сильно вплинути на результат. Якщо задати невелике значення радіусу, то база буде повнішою, але чутливою до викидів і неточностей вимірів. Якщо задати радіус занадто великим, то можна втратити деякі правила при синтезі моделі.

2.5.3. Нечітка кластеризація C -середніх.

В основі алгоритму нечіткої кластеризації C -середніх лежить метод невідзначених множників Лагранжа, який дозволяє задачі знаходження умовного екстремуму цільової функції на множині допустимих значень перетворитись на задачу безумовної оптимізації функції.

Алгоритм нечіткої кластеризації C -середніх – це ітеративна процедура, в якій виконуються наступні кроки:

1. Завдання нечітких кластерів матрицею розбиття:

$$M_D = [\mu_{\theta i}], \mu_{\theta i} \in [0,1], \theta = \overline{1, \Theta}, i = \overline{1, k_c}; \quad (2.56)$$

при цьому

$$\sum_{i=1}^{k_c} \mu_{\theta i} = 1, \quad 0 < \sum_{\theta=1}^{\Theta} \mu_{\theta i} < \Theta; \quad (2.57)$$

де $\mu_{\theta i}$ – ступінь належності об'єкта θ до кластеру i , k_c – кількість кластерів, Θ – кількість елементів.

2. Установка параметрів алгоритму: k_c – кількість кластерів, ϖ – експоненційна вага, яка визначає нечіткість, розмазаність кластерів ($\varpi \in [1, \infty]$), ε – параметр зупинки алгоритму.

3. Генерація випадковим чином матриці нечіткого розбиття з урахуванням умов (2.57).

4. Розрахунок центрів кластерів Ω_i :

$$\Omega_i = \frac{\sum_{\theta=1}^{\Theta} \mu_{\theta i}^{\varpi} * |X_{\theta}|}{\sum_{\theta=1}^{\Theta} \mu_{\theta i}^{\varpi}}, \quad i = \overline{1, k_c}. \quad (2.58)$$

5. Розрахунок відстані між об'єктами з матриці спостережень і центрами кластерів:

$$D_{\theta i} = \sqrt{\|X_{\theta} - \Omega_i\|^2}. \quad (2.59)$$

6. Перерахунок елементів матриці розбиття.

$$\text{- якщо } D_{\theta i} > 0, \text{ то } \mu_{\theta i} = 1 / \left(D_{j\theta}^2 * \sum_{j=1}^{k_c} \frac{1}{D_{j\theta}^2} \right)^{1/(\varpi-1)}. \quad (2.60)$$

$$\text{- якщо } D_{\theta i} = 0, \text{ то } \mu_{\theta i} = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}, \quad j = \overline{1, k_c}. \quad (2.61)$$

6. Перевірка умови (якщо вона виконується, то кінець алгоритму, інакше – перехід до пункту 4):

$$\|M_D - M_D^*\| < \varepsilon, \quad (2.62)$$

де M_D^* – матриця нечіткого розбиття на попередній ітерації алгоритму.

У наведеному алгоритмі найважливішим параметром, який може сильно вплинути на результат, є число кластерів k_c . Правильно вибрати кількість кластерів для реальних завдань без будь-якої апіорної інформації про структури даних досить складно, й наразі існують два підходи до цього. Перший підхід заснований на критерії компактності і роздільності отриманих кластерів (використовуючи, наприклад, індекс Хей-Бені). Другий підхід пропонує починати кластеризацію при досить великому числі кластерів, а потім послідовно об'єднувати схожі суміжні кластери. При цьому використовуються різні формальні критерії схожості кластерів.

Також на результат алгоритму нечіткої кластеризації С-середніх може вплинути такий параметр, як експоненційна вага (ϖ), яка задає рівень нечіткості отриманих кластерів. Наразі не існує обґрунтованого правила вибору значення експоненціального ваги, і зазвичай її встановлюють рівною 2.

2.6. Ефективність систем з нечіткою логікою

Ефективність використання апарату нечіткої логіки базується на наступних результатах.

1. У 1992 р. Ванг (Wang) показав, що нечітка система, яка використовує набір правил виду:

$$\text{Правило } i: \text{ якщо } x_i \in A_i \text{ і } y_i \in B_i, \text{ то } z_i \in C_i \quad i=1,2,\dots,n \quad (2.63)$$

при гаусівських функціях належності, композиції у вигляді добутку, імплікації у формі Ларсена, а також центроїдного методу приведення до чіткості є універсальним апроксиматором, тобто може апроксимувати будь-яку безперервну функцію з довільною точністю (звісно, при $n \rightarrow \infty$).

Інакше кажучи, Ванг довів *теорему*:

для кожної речової безперервної функції g , заданої на компактї U і для довільного $\varepsilon > 0$ існує нечітка експертна система, що формує вихідну функцію $f(x)$ таку, що

$$\sup_{x \in U} \|g(x) - f(x)\| \leq \varepsilon, \quad (2.64)$$

де $\|\cdot\|$ – символ прийнятої відстані між функціями.

2. У 1995 р. Кастро (Castro) показав, що логічний контролер Мамдані при симетричних трикутних функціях належності, композиції з використанням операції \min , імплікації у формі Мамдані, а також центроїдного методу приведення до чіткості також є універсальним апроксиматором.

Взагалі, системи з нечіткою логікою доцільно застосовувати для складних процесів, коли немає простої математичної моделі, а також якщо експертні знання про об'єкт або про процес можна сформулювати тільки в лінгвістичній формі.

Системи з нечіткою логікою застосовувати недоцільно у випадках, коли необхідний результат може бути достатньо просто отриманий будь-яким іншим (стандартним) шляхом, а також коли для об'єкта або процесу вже знайдена адекватна й легко досліджувана математична модель.

Основні недоліки систем з нечіткою логікою:

1. Вихідний набір нечітких правил-постулатів формулюється експертом-людиною і може виявитися неповним або суперечливим.

2. Вид і параметри функцій належності, що описують вхідні і вихідні змінні системи, обираються суб'єктивно і можуть виявитися такими, що не цілком відображають реальну дійсність.

2.7. Нейронечіткі мережі

Кожна з систем ШІ має свої особливості, що робить їх найбільш придатними для вирішення одних задач і менш придатними – для інших. Взагалі, сис-

теми з нечіткою логікою і штучні НМ еквівалентні один одному, проте, на практиці у них є свої власні переваги і недоліки.

Так, НМ ефективні для задач розпізнавання образів, але дуже незручні для з'ясування питання, як вони таке розпізнавання здійснюють. Вони можуть автоматично здобувати знання, але процес їх навчання може відбуватися досить повільно, а аналіз навченої мережі досить складний (навчена мережа зазвичай – «чорний ящик» для користувача). При цьому будь-яку апіорну інформацію (знання експерта) для прискорення процесу її навчання в НМ ввести неможливо.

Системи ж з нечіткою логікою, навпаки, ефективні для пояснення отриманих з їх допомогою висновків, але вони не можуть автоматично здобувати знання для використання їх в механізмах висновків. Необхідність розбивки універсальних множин на окремі області, зазвичай, обмежує кількість вхідних змінних в таких системах невеликим значенням.

Для усунення недоліків НМ і систем з нечіткою логікою запропоновані гібридні мережі, в яких висновки робляться на основі апарату нечіткої логіки, але відповідні функції належності підлаштовуються із використанням алгоритмів навчання НМ, наприклад, алгоритму зворотного поширення похибки. Такі системи не тільки використовують апіорну інформацію, але й можуть набувати нових знань, а для користувача є логічно прозорими.

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

Одним з перших варіантів гібридних мереж є Anfis (Adaptive Neuro Fuzzy Inference System) – адаптивна мережа нечіткого висновку, яка реалізує систему нечіткого висновку Сугено у вигляді п'ятишарової НМ прямого розповсюдження. Призначення шарів наступне:

- перший шар – терми вхідних змінних;

- другий шар – антецеденти (посилки) нечітких правил;
- третій шар – нормалізація ступенів виконання правил;
- четвертий шар – укладення правил;
- п'ятий шар – агрегування результату, отриманого за різними правилами.

Входи мережі в окремий шар не виділяються. На рис. 2.11 зображено Anfis -мережу з двома вхідними змінними (x_1 і x_2) і чотирма нечіткими правилами.

Для лінгвістичної оцінки вхідної змінної x_1 використовується три терми, для змінної x_2 – два терми.

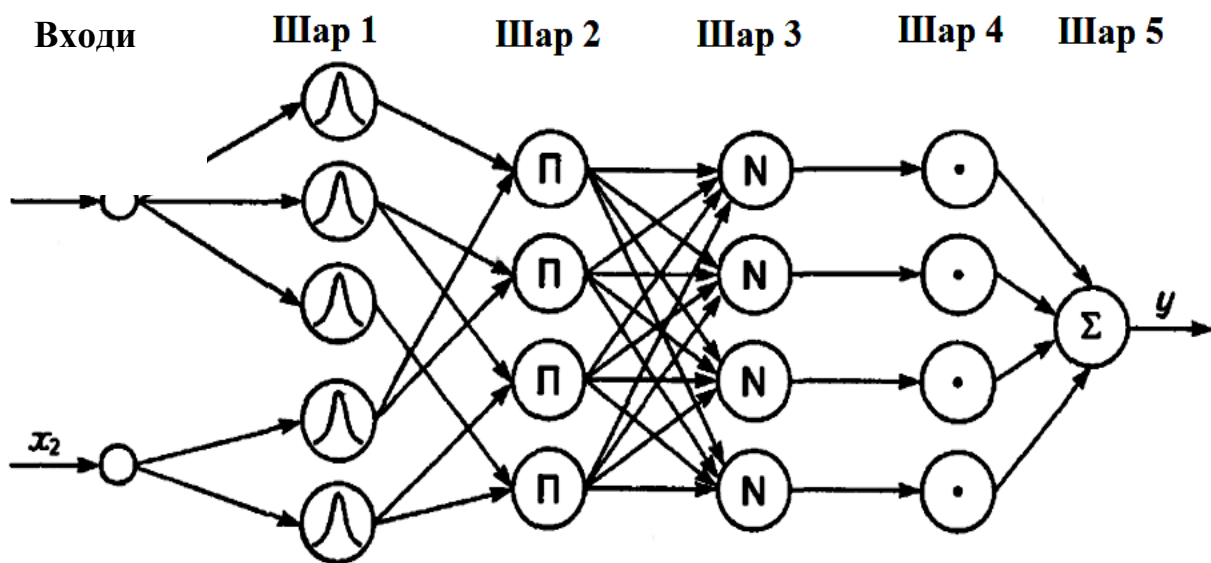


Рис. 2.11. Структура гібридної мережі Anfis

Адаптивна система нечіткого висновку Anfis функціонує наступним чином.

Шар 1. Кожен вузол першого шару представляє один терм з функцією належності. Входи мережі x_1, x_2, \dots, x_n з'єднані тільки зі своїми термами. Кількість вузлів першого шару дорівнює сумі потужностей терм-множин вхідних змінних. Виходом вузла є ступінь належності значення вхідної змінної відповідного нечіткого терму:

$$\mu_r(x_i) = \frac{1}{1 + \left| \frac{x_i - c}{a} \right|^{2b}} \quad (2.65)$$

де a , b і c – параметри функції належності, які налаштовуються.

Шар 2. Кількість вузлів другого шару дорівнює m . Кожен вузол цього шару відповідає одному нечіткому правилу. Вузол другого шару з'єднаний з тими вузлами першого шару, які формують антецеденти відповідного правила. Отже, кожен вузол цього шару може приймати від 1 до n вхідних сигналів. Виходом вузла τ_r ($r = \overline{1, m}$) є ступінь виконання правила, яка розраховується як добуток вхідних сигналів.

Шар 3. Кількість вузлів третього шару також дорівнює m . Кожен вузол цього шару розраховує відносну ступінь виконання нечіткого правила:

$$\tau_r^* = \frac{\tau_r}{\sum_{j=1, m} \tau_j}. \quad (2.66)$$

Шар 4. Кількість вузлів цього шару також дорівнює m . Кожен вузол з'єднаний з одним вузлом третього шару, а також із усіма входами мережі (на рис. 2.11 зв'язки зі входами не показані). Вузол четвертого шару розраховує внесок одного нечіткого правила у вихід мережі:

$$y_r = \tau_r^* (b_{0,r} + b_{1,r}x_1 + \dots + b_{n,r}x_n). \quad (2.67)$$

Шар 5. Єдиний вузол цього шару підсумовує вклади усіх правил:

$$y = y_1 + \dots + y_r + \dots + y_m. \quad (2.68)$$

Типові процедури навчання НМ можуть бути застосовані для настройки мережі Anfis, оскільки в ній використовуються тільки функції, які диференціюються. Зазвичай застосовується комбінація градієнтного спуску у вигляді алгоритму зворотного поширення похибки і методу найменших квадратів. Алгоритм зворотного поширення похибки налаштовує параметри антецедентів правил, тобто функцій належності. Метод найменших квадратів оцінює коефіцієнти висновків правил, оскільки вони лінійно пов'язані із виходом мережі.

Кожна ітерація процедури налаштування виконується у два етапи. На першому етапі на входи подається навчальна вибірка й по розбіжності між бажаною і дійсною поведінкою мережі ітераційним методом найменших квадратів знаходяться оптимальні параметри вузлів шару 4. На другому етапі залиш-

кова розбіжність передається з виходу мережі на входи і методом зворотного поширення похибки модифікуються параметри вузлів шару 1. При цьому знайдені на першому етапі коефіцієнти висновків правил не змінюються. Ітераційна процедура настройки триває поки розбіжність перевищує заздалегідь встановлене значення. Для налаштування функцій належності поряд з методом зворотного поширення похибки можуть використовуватись й інші алгоритми оптимізації (метод Левенберга-Марквардта та інші).

Контрольні запитання

1. Дайте визначення нечітких множин.
2. Наведіть функції належності нечітких множин.
3. Наведіть логічні операції над нечіткими множинами.
4. Наведіть алгебраїчні операції над нечіткими множинами.
5. Дайте визначення нечітких відношень та розкрийте основні операції над ними.
6. Наведіть схему алгоритму Мамдані (Mamdani).
7. Наведіть схему алгоритму Сугено (Sugeno).
8. У чому полягає суть методу субтрактивної кластеризації?
9. На чому базується ефективність використання систем з нечіткою логікою?
10. Наведіть схему та принцип роботи адаптивної нейронної системи нечіткого висновку Anfis.

Глава 3. ЕВОЛЮЦІЙНЕ МОДЕЛЮВАННЯ

3.1. Вступ

Для розв'язання задач оптимізації складних систем різного походження широко використовують еволюційні алгоритми (ЕА), які включають в себе *генетичні алгоритми* (ГА – Genetic Algorithm), *еволюційну стратегію* (ЕС – Evolution Strategy), *еволюційне програмування* (ЕП – Evolutionary Programming), *алгоритми диференціальної еволюції* (ДЕ – Differential Evolution), а також *генетичне програмування* (ГП – Genetic Programming). Останній клас алгоритмів не використовується для вирішення задач безперервної оптимізації, тому далі він не розглядається.

Суть парадигми ЕА полягає у використанні базових принципів теорії біологічної еволюції – відбору, мутації і відтворення осіб. ЕА є частиною більш широкої технології так званих *м'яких обчислень* (Soft Computing), що включають в себе ще НМ, нечітку логіку, ймовірнісні міркування і мережі довіри. Дані технології доповнюють одна одну та використовуються у різних комбінаціях або самостійно для створення інтелектуальних систем.

Найбільш розвинений клас еволюційних алгоритмів – ГА. Також великий розвиток отримали ЕС та ЕП. У порівнянні із зазначеними алгоритмами, алгоритм ДЕ розроблений та використовується в істотно меншій мірі.

Алгоритми ЕС та ЕП мають ряд важливих загальних властивостей, так що існує навіть точка зору, відповідно до якої обидва ці алгоритми, по суті, однакові і являють собою модифікації ГА. При розробці алгоритмів ЕС та ЕП в якості базової прийнято таку концепція: еволюція популяції полягає перш за все в пристосуванні її осіб до навколишнього середовища на поведінковому рівні, а не на генетичному, як у ГА. Наслідком цієї концепції є відсутність в зазначених методах рекомбінації осіб, тобто оператора кросовера. Єдиним оператором у цих методах, що забезпечує пошук альтернативних рішень, є оператор мутації.

Перевагами ЕА є:

- широка область застосування;
- можливість проблемно-орієнтованого кодування рішень;
- підбір початкової популяції, комбінування еволюційних обчислень з не еволюційними алгоритмами, продовження процесу еволюції до тих пір, поки є необхідні ресурси;
- придатність для пошуку в складному просторі рішень великої розмірності;
- відсутність обмежень на вид цільової функції;
- ясність схеми та базових принципів еволюційних обчислень;
- інтегрованість еволюційних обчислень з іншими неklasичними парадигмами ШІ, такими як НМ і нечітка логіка.

До недоліків ЕА слід віднести наступні:

- евристичний характер еволюційних обчислень не гарантує оптимальності отриманого рішення (правда, на практиці, найчастіше, важливо за заданий час отримати одне або кілька субоптимальних альтернативних рішень, тим більше що початкові дані в завданні можуть динамічно змінюватися, бути неточними або неповними);
- відносно висока обчислювальна трудомісткість, яка проте долається за рахунок розпаралелювання на рівні організації еволюційних обчислень і на рівні їх безпосередньої реалізації в обчислювальній системі;
- відносно невисока ефективність на заключних фазах моделювання еволюції (оператори пошуку в ЕА не орієнтовані на швидке потрапляння в локальний оптимум);
- невирішеність питань само адаптації.

Слід зазначити, що до ЕА відносять також *метод групового урахування аргументів* (МГУА – Group Method of Data Handling, GMDH). Це типовий метод індуктивного моделювання і один з найбільш ефективних методів структурно-параметричної ідентифікації складних об'єктів, процесів і систем за даними спостережень в умовах неповноти інформації.

МГУА відрізняється активним застосуванням принципів автоматичної генерації варіантів, послідовної селекції моделей і зовнішніх критеріїв для побудови моделей оптимальної складності. Він має оригінальну процедуру багаторядної автоматичної генерації структур моделей, що імітує процес біологічної селекції з попарним урахуванням послідовних ознак. Цей метод є оригінальним і ефективним засобом розв'язання широкого спектру задач штучного інтелекту, в тому числі ідентифікації та прогнозування, розпізнавання образів і кластеризації, інтелектуального аналізу даних і пошуку закономірностей.

3.2. Біологічні передумови й загальна схема еволюційних алгоритмів

Перші еволюційні алгоритми – ГА – запропонував професор Мічиганського університету Холланд (J. Holland) в 60 рр. XX століття. Елементи цих алгоритмів досліджені раніше в роботах інших авторів. ГА отримали загальне визнання після виходу у 1975 р. в світ книги Холланда, що стала класикою в цій області «Адаптація в природних і штучних системах». Холланд відзначав, що розробку ним теорії ГА стимулювало читання біологічної літератури по дарвінівській теорії природного відбору і селекції сільськогосподарських культур.

Поряд з моделлю еволюції Дарвіна, покладеної в основу канонічного і сучасних ЕА, відомо також значне число інших моделей еволюції – модель Ламарка, модель Фріза, модель Гаулда і Елдріджа, тощо. На основі принципів, закладених в цих моделях, побудовані відповідні варіанти ЕА. Деякі ідеї зазначених моделей можуть бути використані і використовуються також для модифікації класичних ГА.

Чарльз Дарвін в його знаменитій праці «Походження видів» (1859 р.) показав, що еволюційний розвиток земної флори і фауни відбувається під впливом навколишнього середовища на основі наступних принципів: спадковість (нащадки зберігають властивості батьків), мінливість (нащадки майже завжди неідентичні), природний відбір (виживають найбільш пристосовані нащадки).

У 1944 р. Евері (O. Avery), Маклауд (C. MacLeod) і Маккарті (M. McCarty) довели, що дезоксирибонуклеїнова кислота (ДНК) є речовиною, що визначає спадкові процеси. У 1953 р. Крик (F. Crick) і Уотсон (J. D. Watson) розшифрували структуру ДНК у вигляді подвійної спіралі. Таким чином, на початку 60-х рр. ХХ століття стали відомі молекулярно-біологічні основи спадковості і мінливості видів, які й лягли в основу теорії ГА, а пізніше – в основу інших ЕА.

У кожній клітині будь-якої живої істоти міститься уся її генетична інформація, записана в молекулах ДНК, що представляють собою ланцюжок з молекул нуклеотидів чотирьох типів, які позначаються А, Т, С, G. Генетичну інформацію несе порядок слідування нуклеотидів у ДНК. Тобто генетичний код індивідуума – це дуже довге слово, алфавіт якого містить всього 4 букви.

Структуру, яка містить ДНК, називають *хромосомою* (chromosome). Кожна вроджена якість особи (колір волосся, очей, спадкові хвороби тощо) кодує певна частина хромосоми – *ген* (gene) цієї властивості. Місце гена в хромосомі називають *локусом* (locus), а функціональне значення гена – його *алелем* (allele). Сукупність генів даної особи є *генотипом* (genotype). На основі генотипу формується *фенотип* (phenotype) – сукупність характеристик, притаманних особі на певній стадії її розвитку.

При розмноженні живих істот відбувається злиття двох батьківських клітин, і їх ДНК взаємодіють, утворюючи ДНК нащадка. Основним способом взаємодії ДНК є їх *схрещування* (mating, crossing), при якому ДНК батьків діляться на дві частини, а потім обмінюються ними. Як синоніми використовують також терміни *кросовер* (crossover) і *рекомбінація* (recombination).

Як наслідок певних факторів, деякі гени в клітинах батьків можуть змінюватися – відбуваються *мутації* (mutation) генів. В результаті схрещування змінені гени можуть передаватися нащадку, надаючи йому властивості, які відсутні у батьків. Якщо ці нові властивості корисні, то вони, швидше за все, збережуться в даному виді і приведуть до підвищення його пристосованості.

Схема канонічного ГА має такий вигляд.

1. Представлення агентів (осіб, індивідуумів) у вигляді хромосом.
2. Створення випадковим чином деякого числа початкових осіб – *початкової популяції*.
3. Оцінка осіб за допомогою фітнес функції (або цільової функції) – кожній особі ставлять у відповідність певне значення *пристосованості* (fitness), яке визначає ймовірність її виживання.
4. Відбір осіб для схрещування на основі пристосованості – етап *селекції* (selection). Застосування до хромосом цих осіб *генетичних операторів* схрещування і мутації, тобто створення таким чином наступного *покоління* осіб.
5. Оцінка осіб наступного покоління, їх селекція, застосування до них генетичних операторів і так далі до тих пір, поки не буде виконано критерій зупинки алгоритму.

Суть ЕА, як і популяційних алгоритмів в цілому, полягає в забезпеченні більш високої середньої пристосованості нового покоління у порівнянні з такою ж пристосованістю попереднього покоління.

Наразі існує велика кількість варіантів розглянутої схеми ГА, що відрізняються способами кодування хромосом, набором генетичних операторів, структурою і параметрами алгоритму тощо.

Розглянемо задачу глобальної умовної оптимізації

$$\max_{X \in D \subset R^{|X|}} f(X) = f(X^*) = f^* \quad (3.1)$$

з обмеженнями загального вигляду D . Вважаємо, що більшим значенням цільової функції $f(X)$ відповідають більші значення фітнес функції $\varphi(X)$.

Закодований вектор змінних параметрів X , тобто хромосому відповідної особи s , позначимо як $H = (H_i^s, i \in [1:|X|])$. Тут ген H_i^s являє собою код компоненти x_i вектору X ; i – номер локусу, в якому знаходиться цей ген. Алель гену H_i^s є сукупністю усіх можливих значень величини x_i .

З іншої сторони хромосома H є сукупністю $h_1, h_2, \dots, h_{|H|}$ символів деякого алфавіту, за допомогою якого виводимо кодування вектору X , таким чином,

що $H=(h_1, h_2, \dots, h_{|H|})$, а набори $(h_1, h_2, \dots, h_{|H_1^s|})$, $(h_{|H_1^s|+1}, h_{|H_1^s|+2}, \dots, h_{|H_1^s|+|H_2^s|})$ є кодами генів H_1^s, H_2^s, \dots відповідно. Тут $|H|$ – довжина хромосоми, $|H_i^s|$ – розмірність i -го гену, $i \in [1:|X|]$. Легко побачити, що має місце співвідношення

$$\sum_{i=1}^{|X|} |H_i^s| = |H|. \quad (3.2)$$

Отже, особа s визначається таким чином

$$s = \langle X, H, \varphi(X) \rangle, \quad (3.3)$$

де X – фенотип особи, H – її генотип, $\varphi(X)$ – пристосованість цієї особи. Поряд із $\varphi(X)$ у тому ж сенсі позначається $\varphi(H)$ і $\varphi(s)$.

Далі для простоти запису у ситуаціях під фразою «ген H_i^s особи s », мається на увазі ген H_i^s хромосоми H , за допомогою якої виконано кодування фенотипу цієї особи. Аналогічно під фразою «розряд h_i особи s », мається на увазі розряд h_i хромосоми H .

Основними є наступні генетичні оператори:

- оператор мутації, який реалізує довільну зміну одного або декількох генів хромосоми;
- оператор схрещування, призначений для створення осіб-нащадків шляхом рекомбінації хромосом двох або більше батьківських осіб;
- оператор управління популяцією, який формує на основі популяції $S=S(t)$ поточного покоління t проміжну популяцію $S' = (s_{i_j}, i_j \in [1:|S|], j \in [1:|S'|])$, де $|S'|$ – число осіб у цій популяції;
- оператор селекції, який здійснює вибір з проміжної популяції S' осіб для схрещування, і формування на цій основі популяції $S(t+1)$ нового покоління $(t+1)$.

Слід зазначити, що у найпростішому випадку проміжна популяція може співпадати із поточною.

Зазначені оператори використовують, зазвичай, певну кількість вільних параметрів. Ці параметри можуть змінювати свої значення в процесі обчислень або в функції числа поколінь алгоритму, або адаптивно – на основі використання зворотного зв'язку їх значень з ефективністю процесу пошуку. Оператори з такими параметрами називають *нерівномірними* і *адаптивними операторами*, відповідно.

Для ЕА взагалі, й зокрема для ГА, гостро стоїть питання збіжності. У деякій мірі відповідь на це питання дає наступна теорема.

Теорема. Припустимо, що послідовність популяцій $S(t)$, $t=0,1,\dots$, яка генерується ГА, монотонна, тобто для всіх t задовольняється умова

$$\max(\varphi(s_i) \mid s_i \in S(t+1)) \geq \max(\varphi(s_j) \mid s_j \in S(t)). \quad (3.4)$$

Нехай також для будь-яких двох допустимих осіб s , s' можливий перехід від першої з них до другої за допомогою одноразового або багаторазового застосування операторів мутації і кросовера. Тоді ГА відшукує глобально оптимальну особу s^* , що доставляє максимум функції $\varphi(s)$ з ймовірністю ξ , яка дорівнює одиниці, тобто

$$\lim_{t \rightarrow \infty} (\xi(s^* \in S(t))) = 1. \quad (3.5)$$

Перша із зазначених в теоремі умов є досить сильною, і в реальних ГА без додаткових зусиль зазвичай не виконується. Друга ж умова виконується завжди. В цілому доказ збіжності різних комбінацій генетичних операторів є сучасною актуальною науковою задачею.

3.3. Генетичні алгоритми

Мета при оптимізації за допомогою ГА полягає у тому, щоб знайти найкраще можливе рішення або рішення задачі по одному або декільком критеріям. Щоб реалізувати ГА, потрібно спочатку вибрати відповідну структуру для представлення цих рішень. У постановці задачі пошуку об'єкт цієї структури даних представляє точку в просторі пошуку усіх можливих рішень. Властивості

об'єктів представлені значеннями параметрів, що об'єднуються в хромосоми. У генетичних методах оперують хромосомами, що відносяться до множини об'єктів популяції. Імітація генетичних принципів веде до еволюційного поліпшення значень цільової функції (функції пристосованості) від покоління до покоління.

Найчастіше хромосома – це бітовий рядок. Однак ГА не обмежені бінарним представленням даних. Деякі реалізації використовують цілочисельне або дійсне кодування. Незважаючи на те, що для багатьох реальних задач, мабуть, більше підходять рядки змінної довжини, в даний час структури фіксованої довжини найбільш поширені й вивчені. Тому далі розглядаються лише структури, які є поодинокими рядками по n біт.

Кожна хромосома (рядок) є конкатенацією ряду підкомпонентів, званих генами. Як було вже сказано раніше, гени розташовуються в різних позиціях або локусах хромосоми й приймають значення, звані алелями. В уявленнях з бінарними рядками ген – це біт, локус – його позиція в рядку і алель – його значення (0 або 1). Генотип відноситься до повної генетичної моделі особи і відповідає структурі в ГА, а фенотип відноситься до зовнішніх спостережуваних ознак і відповідає вектору в просторі параметрів.

Приклад використання ГА – задача максимізації наступної функції двох змінних: $f(x_1, x_2) = x_1 x_2$, де $0 \leq x_1 \leq 1$ і $0 \leq x_2 \leq 1$.

Зазвичай методика кодування реальних змінних x_1 і x_2 полягає в їх перетворенні в двійкові цілочисельні рядки достатньої довжини (достатньої для забезпечення бажаної точності). Якщо припустити, що 10-розрядне кодування достатньо й для x_1 , й для x_2 , то встановити відповідність між генотипом і фенотипом закодованих осіб можна, розділивши відповідне бінарному представленню ціле число на значення $2^{10} - 1$. Наприклад, код [0000000000] відповідає дійсному значенню $0/1023$ або 0, тоді як код [1111111111] відповідає $1023/1023$ або 1. Структура даних, яка оптимізується – 20-бітний рядок, що представляє конкатенацію кодувань x_1 і x_2 . Змінна x_1 розміщується в крайніх лівих 10-розрядах, тоді як x_2 розміщується у правій частині генотипу особи (20-бітовому рядку).

При цьому, генотип – точка в 20-вимірному бінарному просторі, який досліджується ГА. Фенотип – точка в двовимірному просторі параметрів.

Кодування рішень.

Після того як обрані параметри, їх число і розрядність, необхідно вирішити, як безпосередньо записувати дані. Можна використовувати звичайне кодування, коли, наприклад, $1011_2=11_{10}$, або *коди Грея*, коли $1011_G=1110_2=14_{10}$. Незважаючи на те, що використання кодів Грея тягнуть неминуче кодування/декодування даних, вони дозволяють уникнути деяких проблем, які є нормальним результатом кодування. Перевага коду Грея у тому, що якщо два числа є послідовними при кодуванні, то і їх двійкові коди розрізняються тільки на один розряд, а в двійкових кодах це не так. Слід зазначити, що кодувати й декодувати в коді Грея можна таким чином: спочатку копіюється найстарший розряд, потім:

- з двійкового коду в код Грея: $G[i] = \text{XOR}(B[i+1], B[i])$;
- з коду Грея в двійковий код: $B[i] = \text{XOR}(B[i+1], G[i])$.

Тут $G[i]$ i -й розряд коду Грея, а $B[i]$ – i -й розряд бінарного коду. Наприклад, послідовність чисел від 0 до 7 в двійковому коді: {000, 001, 010, 011, 100, 101, 110, 111}, а в кодах Грея: {000, 001, 011, 010, 110, 111, 101, 100}.

Загальна схема ГА була представлена у попередньому розділі. Зупинимося нижче докладніше на усіх етапах цього алгоритму.

3.3.1. Формування початкової популяції.

Стандартний ГА починає свою роботу з формування початкової популяції I_0 – кінцевого набору допустимих рішень задачі. Ці рішення можуть бути обрані випадковим чином або отримані за допомогою простих наближених алгоритмів. Вибір початкової популяції не має значення для збіжності процесу в асимптотиці, проте формування «гарної» початкової популяції (наприклад, із множини локальних оптимумів) може помітно скоротити час досягнення глобального оптимуму. Якщо відсутні припущення про місцезнаходження глобального

оптимуму, то індивіди з початкової популяції бажано розподілити рівномірно по всьому простору пошуку рішення.

3.3.2. Оцінка осіб популяції.

Щоб оптимізувати будь-яку структуру з використанням ГА, потрібно задати міру якості для кожного індивіда в просторі пошуку. Для цієї мети використовується *функція пристосованості*. У задачах максимізації цільова функція часто сама виступає як функція пристосованості; для задач мінімізації цільову функцію слід інвертувати. Якщо задача, яку необхідно вирішити, має обмеження, виконання яких неможливо контролювати алгоритмічно, то функція пристосованості, як правило, включає також штрафи за невиконання обмежень (вони зменшують її значення).

3.3.3. Відбір (селекція).

На кожному кроці еволюції за допомогою ймовірнісного оператора селекції (відбору) обираються два рішення-батька для їх подальшого схрещування. Серед операторів селекції найбільш поширеними є два ймовірнісних оператора пропорційної і турнірної селекції. У деяких випадках також застосовується відбір урізанням.

Пропорційний відбір (Proportional Selection).

При пропорційній селекції ймовірність на k -му кроці вибрати i -те рішення в якості одного з батьків задається формулою:

$$P \{i - \text{обране}\} = \frac{f(i)}{\sum_{j \in I_k} f(j)} \quad (3.6)$$

в припущенні, що $f(i) > 0$ для всіх $i \in I_k$ (тут I_k – популяція на кроці k).

Найпростіший пропорційний відбір – рулетка – відбирає осіб за допомогою n «запусків» рулетки. Колесо рулетки містить по одному сектору для кожного i -го члена популяції. Розмір i -го сектора пропорційний відповідній величині $P(i)$. При такому відборі члени популяції з більш високою пристосованіс-

ттю з більшою ймовірністю будуть частіше вибиратися, ніж особи з низькою пристосованістю.

Турнірний відбір (Tournament Selection).

Турнірний відбір може бути описаний таким чином: з популяції, що містить m рядків (осіб), вибирається випадковим чином t рядків й найкращий рядок записується в проміжний масив (між обраними рядками проводиться турнір). Ця операція повторюється m раз. Рядки в отриманому проміжному масиві потім використовуються для схрещування (також випадковим чином). Розмір групи рядків, що відбираються для турніру, часто дорівнює 2. У цьому випадку говорять про двійковий/парний турнір. Взагалі ж t – чисельність турніру.

Відбір урізанням (Truncation Selection).

Дана стратегія використовує відсортовану по спадаючій популяцію. Число осіб для схрещування вибирається відповідно до порогу $T \in [0; 1]$. Поріг означає, яка частка осіб, починаючи з найпершої (самої пристосованої), братиме участь у відборі. В принципі, поріг можна задати й рівним 1, тоді всі особи поточної популяції будуть допущені до відбору. Серед осіб, допущених до схрещування випадковим чином $m/2$ раз вибираються батьківські пари, нащадки яких утворюють нову популяцію.

3.3.4. Схрещування.

Як тільки два рішення-батька обрані, до них застосовується ймовірнісний оператор схрещування (Crossover), який буде на їх основі нові (1 або 2) рішення-нащадка. Відібрані особи піддаються кросоверу (іноді званого рекомбінацією) із заданою вірогідністю P_c . Якщо кожна пара батьків породжує двох нащадків, для відтворення популяції необхідно схрестити $m/2$ пари. Для кожної пари з ймовірністю P_c застосовується кросовер. Відповідно, з ймовірністю $1-P_c$ кросовер не відбувається і тоді незмінені особи переходять на наступну стадію (мутації).

Існує велика кількість різновидів оператора схрещування. Найпростіший *одноточковий кросовер* працює наступним чином (рис. 3.1). Спочатку випадко-

вим чином вибирається одна з можливих точок розриву (точка розриву – ділянка між сусідніми бітами в рядку.) Обидві батьківські структури розриваються на два сегменти по цій точці. Далі відповідні сегменти різних батьків склеюються і виходять два генотипу нащадків.

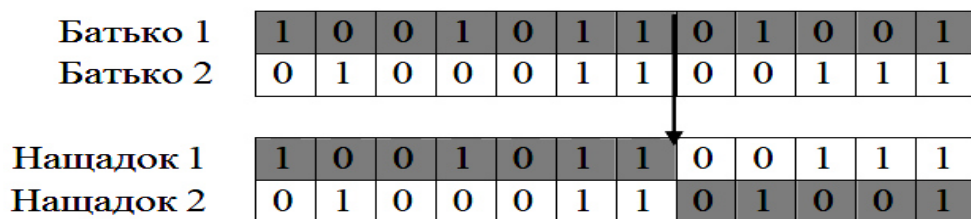


Рис. 3.1. Приклад роботи одноточкового кросовера

Наразі дослідники ГА пропонують багато інших операторів схрещування. Двоточковий і рівномірний кросовер цілком гідні альтернативи одноточковому. У двоточковому кросовері обираються дві точки розриву, й батьківські хромосоми обмінюються сегментом, який знаходиться між двома цими точками. У рівномірному кросовері кожен біт першого нащадка випадковим чином успадковується від одного з батьків; другому нащадку дістається біт другого з батьків.

3.3.5. Мутація.

Після того як закінчиться стадія кросовера, нащадки можуть піддаватися випадковим модифікаціям, званим мутаціями. У найпростішому випадку в кожній хромосомі, яка піддається мутації, кожен біт з ймовірністю P_m змінюється на протилежний – це так звана *одноточкова мутація* (рис. 3.2).

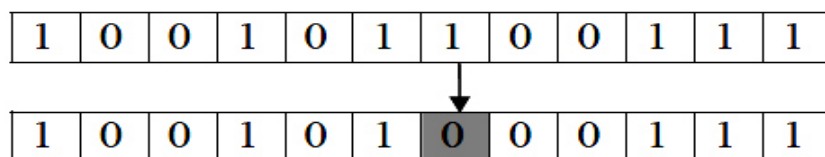


Рис. 3.2. Приклад виконання одноточкової мутації

Складнішою різновидом мутації є оператори інверсії і транслокації. *Інверсія* це перестановка генів у зворотному порядку всередині навмання вибраної ділянки хромосоми (рис. 3.3).



Рис. 3.3. Приклад виконання інверсії

Транслокація – це перенесення будь-якої ділянки хромосоми, в інший сегмент цієї ж хромосоми (рис. 3.4).

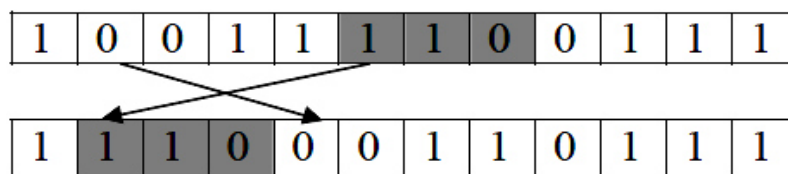


Рис. 3.4. Приклад виконання транслокації

Слід зазначити, що всі розглянуті вище генетичні оператори (одноточковий і багатоточковий кросовер, одноточкова мутація, інверсія, транслокація) мають схожі біологічні аналоги.

У деяких роботах пропонується використовувати стратегію *інцесту* як механізму само адаптації оператора мутації. Вона полягає у тому, що ймовірність мутації кожного гена P_m визначається для кожного нащадка на підставі генетичної близькості його батьків. Наприклад, це може бути відношення числа співпадаючих генів батьків до загальної кількості генів хромосоми. Це призводить до цікавого ефекту – при високій різноманітності генофонду популяції наслідки мутації будуть мінімальними, що дозволяє оператору схрещування працювати без стороннього втручання. У разі ж зниження різноманітності, що виникає в основному при «застряганні» ГА в локальному оптимумі, наслідки мутації стають більш відчутними, а при повному стиску популяції ГА просто стає

стохастичним, що збільшує ймовірність виходу популяції з локального оптимуму.

Іноді (з метою підвищення середньої пристосованості популяції) допустимо здійснювати *спрямовані мутації*, тобто після кожної зміни хромосоми перевіряти, чи підвищилася в результаті цієї мутації її пристосованість і, якщо ні, повертати хромосому до початкового стану.

3.3.6. Формування нового покоління.

Після схрещування і мутації осіб необхідно вирішити проблему про те, які з нових осіб увійдуть в наступне покоління, а які ні, а також що робити з їхніми батьками. Є два найпоширеніші способи.

1. Нові особи (нащадки) займають місця своїх батьків. Після чого настає наступний етап, у якому нащадки оцінюються, відбираються, дають потомство і поступаються місцем своїм «дітям».

2. Наступна популяція включає в себе як батьків, так і їх нащадків.

У другому випадку необхідно додатково визначити, які з осіб батьків і нащадків потраплять в нове покоління. У найпростішому випадку, в нього після кожного схрещування включаються дві кращі особи з четвірки батьків та їхніх нащадків. Більш ефективним є механізм витиснення, який реалізується таким чином, що прагне видаляти «схожі» хромосоми із популяції та залишати такі, що відрізняються.

При розгляданні питання формування нового покоління слід виділити окремо принцип «елітизму». Суть цього принципу полягає у тому, що в нове покоління завжди включаються кращі батьківські особи. Їх число може бути від 1 і більше. Використання цього принципу дозволяє не втратити гарне проміжне рішення, але із-за цього алгоритм може «застрягти» у локальному екстремумі. У більшості випадків «елітизм» не шкодить пошуку рішення, і головне – надає алгоритму можливість аналізувати різні хромосоми з простору пошуку.

3.3.7. Зупинка алгоритму.

Робота ГА являє собою ітераційний процес, який триває до тих пір, поки не пройде задане число поколінь або не виконається будь-який інший критерій зупинки. В оптимізаційних задачах традиційними критеріями зупинки алгоритму є, наприклад, тривала відсутність прогресу в сенсі поліпшення значення середньої (або кращої) пристосованості популяції, мала різниця між кращим і гіршим значенням пристосованості для поточної популяції і тому подібне.

3.4. Типові генетичні алгоритми

3.4.1. Канонічний генетичний алгоритм.

Канонічний (Canonical) ГА є класичним алгоритмом. Ця еволюційна модель була запропонована Дж. Холландом в його знаменитій праці «Адаптація в природних і штучних середовищах» (1975). Часто можна зустріти опис *простого* ГА (Simple GA) Голдберга (D. Goldberg), він відрізняється від канонічного тим, що використовує замість рулеточного, турнірний відбір.

Модель канонічного ГА має наступні характеристики.

- фіксований розмір популяції;
- фіксовану розрядність генів;
- пропорційний відбір;
- одноточковий кросовер і одноточкову мутацію;
- формування наступного покоління з нащадків поточного покоління без «елітизму».

Алгоритм роботи ГА (репродуктивний план Холланда) складається в даному випадку з наступних етапів.

Етап 1. Ініціалізація початкової популяції. Покласти номер епохи $t=0$. Ініціалізувати випадковим чином m генотипів осіб і сформуванати з них випадкову популяцію. Обчислити пристосованість осіб популяції $F(0)=(f_1(0), \dots, f_m(0))$, а потім середню пристосованість популяції

$$f_{\tilde{n}\delta}(0) = \sum_{i=1}^m f_i(0) / m. \quad (3.7)$$

Етап 2. Вибір батьків для схрещування. Збільшити номер епохи на одиницю: $t=t+1$. Визначити випадковим чином номер першого з батьків $l \in \{1 \dots m\}$, призначивши ймовірність випадання будь-якого номера h пропорційною величині $f_h(t)/f_{cp}(t)$. Повторним випробуванням визначити номер другого з батьків k .

Етап 3. Формування генотипу нащадків. З заданою вірогідністю p_c провести над генотипами обраних батьків одноточковий кросовер. Далі до кожного з отриманих нащадків з імовірністю p_m застосувати оператор мутації.

Етап 4. Оновлення популяції. Помістити нащадків в популяцію, попередньо видаливши з неї батьків. Обчислити пристосованості нащадків й оновити значення середньої пристосованості популяції $f_{cp}(t)$.

Якщо формування популяції не завершено, перейти до етапу 2.

Слід зауважити, що у деяких модифікаціях цього алгоритму нащадки заміняють в популяції не своїх батьків, а дві випадково вибрані особи.

3.4.2. Генетичний алгоритм Genitor.

У алгоритмі Genitor (D. Whitley) використовується специфічна стратегія відбору. Спочатку, як і належить, популяція ініціалізується і її особи оцінюються. Далі вибираються випадковим чином дві особи, схрещуються, причому виходить тільки один нащадок, який оцінюється й займає місце найменш пристосованої особи. Після цього знову випадковим чином вибираються 2 особи, і їх нащадок займає місце особи з найнижчою пристосованістю. Таким чином, на кожному кроці в популяції оновлюється тільки одна особи. Отже, ГА Genitor має наступні характерні особливості:

- фіксований розмір популяції;
- фіксовану розрядність генів;
- вибір осіб для схрещування випадковим чином;
- відсутність обмежень на тип кросовера і мутації;

- в результаті схрещування осіб виходить один нащадок, який займає місце найменш пристосованої особи.

Таким чином, ГА Genitor є прикладом реалізації підходу, у якому популяція оновлюється частинами, а не вся відразу.

Одним з переваг ГА Genitor є те, що він вимагає в два рази менше пам'яті комп'ютера для зберігання інформації у порівнянні з будь-яким ГА, що використовує проміжну популяцію. З іншого боку, ГА Genitor має у порівнянні з канонічним ГА значно гірші диверсифікаційні властивості, що обумовлено потенційно швидким збідненням різноманітності популяції за рахунок можливого тривалого збереження в ній кращих осіб і швидкого виключення з неї найменш пристосованих осіб.

3.4.3. Гібридні генетичні алгоритми.

Використання гібридних (Hybrid) ГА дозволяє об'єднати переваги ГА з перевагами класичних методів. Справа в тому, що ГА є робастними алгоритмами, тобто вони дозволяють знаходити гарне рішення, але знаходження оптимального рішення найчастіше виявляється набагато більш важким завданням в силу стохастичності принципів роботи алгоритму. Тому виникла ідея використовувати ГА на початковому етапі для ефективного звуження простору пошуку навколо глобального екстремуму, а потім, взявши кращу особу, застосувати один з «класичних» методів оптимізації. Характеристиками гібридних ГА є:

- фіксований розмір популяції;
- фіксована розрядність генів;
- будь-які комбінації стратегій відбору та формування наступного покоління;
- відсутність обмежень на тип кросовера і мутації;
- застосування ГА на початковому етапі, а потім – класичного методу оптимізації.

3.4.4. Острівний генетичний алгоритм.

Острівна модель ГА (Island Model GA) є найбільш відомою моделлю паралельних генетичних алгоритмів. Такі ГА використовують так звану «модель островів», де кожна підпопуляція розвивається на своєму «острові». Між островами проводиться (достатньо рідко) обмін кращими особами. Дана модель ГА має такі властивості:

- наявність декількох популяцій, як правило, однакового фіксованого розміру;
- фіксована розрядність генів;
- будь-які комбінації стратегій відбору та формування наступного покоління в кожній популяції (можна зробити так, що в різних популяціях будуть використовуватися різні комбінації стратегій, хоча навіть один варіант дає різноманітні рішення на різних «островах»);
- відсутність обмежень на тип кросовера і мутації;
- випадковий обмін особами між «островами». Якщо міграція буде занадто активною, то особливості острівної моделі будуть згладжені, й вона буде не дуже сильно відрізнятися від моделей ГА без паралелізму.

3.4.5. Генетичний алгоритм перехресного відбору.

ГА перехресного відбору (СНС) Ешельмана (Cross-population selection, Heterogenous recombination and Cataclysmic mutation, Ch. L. Eshelman) досить швидко сходиться, оскільки в ньому немає мутацій, що використовуються популяції невеликого розміру, а відбір осіб у наступне покоління ведеться як між батьківськими особами, так й між їхніми нащадками. Після знаходження деякого рішення алгоритм перезавантажується, причому найкраща особа копіюється у нову популяцію, а особи, які залишилися, піддаються сильній мутації (мутує приблизно третина бітів в хромосомі); далі пошук повторюється. Ще однією специфічною рисою є стратегія схрещування: всі особи розбиваються на пари, причому схрещуються лише ті пари, в яких хромосоми осіб істотно різні (хемінгова відстань більше деякого порогу, плюс можливі обмеження на максималь-

ну довжину ланцюжка однакових біт). При схрещуванні використовується HUX-оператор (Half Uniform Crossover) – різновид однорідного кросовера, але в ньому до кожного нащадку потрапляє рівно половина бітів хромосоми від кожного з батьків. Отже, алгоритм СНС має наступні властивості:

- фіксований розмір популяції;
- фіксована розрядність генів;
- перезапуск алгоритму після знаходження рішення;
- невелика популяція;
- особи для схрещування розбиваються на пари і схрещуються за умови істотних відмінностей;
- відбір в наступне покоління проводиться між батьківськими особами і нащадками;
- використовується половинний однорідний кросовер (HUX);
- макромутація при перезапуску.

Слід зазначити, що ГА СНС використовує мутацію тільки для подолання стагнації. В основному циклі алгоритм використовує тільки кросовер.

3.4.6. Теорема схем.

Схемою (або *шимою* – schema) називається рядок виду $(s_1, s_2, \dots, s_i, \dots, s_n)$, $s_i \in \{0, 1, *\}$. Символ «*» позначає невизначене значення (як 0, так і 1).

Наприклад, для двох бінарних рядків «111000111000» і «110011001100» схема буде виглядати наступним чином: «11*0***1*00». Тобто за допомогою схем можна як би виділяти загальні ділянки двійкових рядків і маскувати відмінності. Маючи в складі схеми m символів «*» можна закодувати (узагальнити) $2m$ двійкових рядків. Так наприклад, схема «01*0*1» описує набір рядків {«010001», «010011», «011001», «011011»}.

Визначаюча довжина схеми – відстань між двома крайніми символами «0» і/або «1». Для схеми «01*0*1» визначаюча довжина дорівнює 5, а для схеми «**0**1*» – 3.

Порядок схеми – ще одна її характеристика, вона дорівнює числу фіксованих позицій в рядку, тобто загальній кількості «0» і «1». Для схем «01*0*1» і «**0**1*» порядки рівні 4 і 2, відповідно.

Хоча зовні здається, що ГА обробляє рядки, насправді при цьому неявно відбувається обробка схем, які представляють шаблони подібності між рядками (Гольдберг, 1989; Холланд, 1992). ГА практично не може займатися повним перебором всіх представлень в просторі пошуку. Однак він може робити вибірку значного числа гіперплощин в областях пошуку з високою пристосованістю. Кожна така гіперплощина відповідає множині схожих рядків з високою пристосованістю.

Будуючі блоки (Гольдберг, 1989) – це схеми, які мають високу пристосованість, низький порядок та коротку визначаючу довжину.

Пристосованість схеми визначається як середня пристосованість рядків, які її містять. Після процедури відбору залишаються тільки рядки з більш високою пристосованістю. Отже, рядки, які є прикладами схем з високою пристосованістю, вибираються частіше. Кросовер рідше руйнує схеми з коротшою певною довжиною, а мутація рідше руйнує схеми з низьким порядком. Тому такі схеми мають більше шансів переходити із покоління в покоління. Холланд у 1992 р. показав, що, у той час як ГА явним чином обробляє n рядків на кожному поколінні, в той же час неявно обробляються близько n^3 таких коротких схем низького порядку і з високою пристосованістю. Це явище було названо як неявний паралелізм. Для вирішення реальних задач присутність неявного паралелізму означає, що велика популяція має більше можливостей локалізувати рішення експоненційно швидше популяції із меншим числом осіб.

Теорема схем (The schema theorem) є однією з основних теорем теорії ГА. Вона отримана і уточнена для канонічного ГА та полягає в наступному.

Нехай існує популяція двійкових рядків довжини n . Вірогідність проведення кросовера дорівнює P_c , тип кросовера одноточковий. A визначаюча довжина схеми S дорівнює $\delta(S)$, її пристосованість $f(S)$. Частка рядків, відповідних схемі S в поточному поколінні t , дорівнює $m(H, t)$. Нас цікавить, яка частка ряд-

ків, відповідних схемі S буде присутня в популяції у наступному поколінні $m(S, t+1)$. Обчислимо, з якою ймовірністю кросовер зруйнує вже наявну схему. Очевидно, що якщо точка розриву не потрапляє всередину вже наявної схеми, то схема не буде зруйнована. Тобто якщо $P_c * \delta(S)/(n-1)$ – ймовірність того, що точка розриву потрапить всередину схеми, то $1 - P_c * \delta(H)/(L-1)$ – ймовірність того, що кросовер не зруйнує схему. Згідно стратегії відбору канонічного ГА шанси особи взяти участь у схрещуванні обчислюються відповідно до співвідношення f/f_{cp} , де f – значення пристосованості даної особи, f_{cp} – середня пристосованість. Таким чином, ймовірність того, що рядок, який відповідає схемі H братиме участь у схрещуванні дорівнює $m(S, t) \cdot f(S, t) / f_{cp}(t)$.

Беручи до уваги ймовірність руйнування схеми кросовером, *первісна теорема схем* (Холланд, 1975) формулюється наступним чином:

$$m(S, t+1) \geq m(S, t) \frac{f(S, t)}{f_{cp}(t)} \left[1 - P_c \frac{\delta(S)}{n-1} \right]. \quad (3.8)$$

Одним із недоліків даної теореми є те, що в ній відсутній вплив мутації на створення і руйнування схем. Якщо вважати, що ймовірність мутації дорівнює P_m , а порядок схем H дорівнює $o(S)$, то ймовірність того, що мутація не зруйнує схему дорівнює $(1-P_m)^{o(S)}$. Тобто якщо мутуючий розряд не потрапляє ні на одну фіксовану позицію всередині схеми, то вона не змінюється.

З огляду на це *виправлена теорема схем* виглядає наступним чином (Холланд, 1992):

$$m(S, t+1) \geq m(S, t) \frac{f(S, t)}{f_{cp}(t)} \left[1 - P_c \frac{\delta(S)}{n-1} \right] (1 - P_m)^{o(S)}. \quad (3.9)$$

У той час як теорема схем пророкує зростання прикладів хороших схем, сама теорема вельми спрощено описує поведінку ГА. По-перше, $f(S)$ і f_{cp} не залишаються постійними від покоління до покоління. Пристосованості членів популяції знаменно змінюються вже після кількох перших поколінь. По-друге, теорема схем пояснює втрати схем, але не появу нових. Нові схеми часто створюються кросовером і мутацією. Крім того, у міру еволюції, члени популяції

стають все більше схожими один на одного таким чином, що зруйновані схеми будуть відразу ж відновлені.

Однак, незважаючи на простоту, теорема схем описує кілька важливих аспектів поведінки ГА. Мутації з більшою ймовірністю руйнують схеми високого порядку, у той час як кросовери з більшою ймовірністю руйнують схеми з більшою визначальною довжиною. Коли відбувається відбір, популяція сходиться пропорційно відношенню пристосованості кращої особи до середньої пристосованості в популяції; це відношення – міра тиску відбору. Збільшення або P_c , або P_m або зменшення тиску відбору веде до збільшення різноманітності вибірки або дослідженню простору пошуку, але не дозволяє використовувати всі хороші схеми, якими володіє ГА. Зменшення або P_c , або P_m або збільшення тиску вибору веде до поліпшення використання знайдених схем, але гальмує дослідження простору у пошуках нових хороших схем. ГА повинен підтримувати тонку рівновагу між тим і іншим, що наразі відоме як проблема «балансу дослідження і використання».

3.5. Еволюційна стратегія

Розглянемо задачу (3.1) глобальної умовної оптимізації цільової і фітнес-функції з обмеженнями типу нерівностей, коли шукається мінімум (а не максимум). Алгоритм еволюційної стратегії рішення цієї задачі розроблений у 1964 р. Рехенбергом (I. Rechenberg). Його суть передає наступна послідовність основних кроків:

- 1) проводиться мутація і схрещування осіб поточної популяції;
- 2) батьківські особи і отримані таким чином особи-нащадки включаються у проміжну популяцію;
- 3) на основі цієї популяції формується наступна популяція шляхом детермінованого *відбору без повторень* найбільш пристосованих осіб.

Основні відмінності алгоритму ЕС від класичного ГА полягають у наступному. Перш за все, якщо ГА моделює еволюцію на рівні геномів осіб, то алго-

ритм ЕС орієнтований на еволюцію їх фенотипів. Тому останній використовує тільки дійсне кодування осіб, на відміну від класичного ГА, який передбачає бінарне кодування.

У класичному ГА процес відбору осіб із батьківської популяції для включення їх у нову популяцію здійснюють випадковим чином пропорційно їх пристосованості. В алгоритмі ЕС відбір осіб у нову популяцію здійснюють без повторень на основі детермінованої процедури.

Алгоритм ЕС передбачає, що на кожній ітерації спочатку проводиться мутація, а потім селекція. Класичний ГА виконує ці дії у зворотному порядку. Крім того, останній використовує стаціонарні значення вільних параметрів своїх операторів (наприклад, ймовірності схрещування і мутації осіб), а алгоритм ЕС, навпаки, безперервно змінює значення цих параметрів у режимі самоадаптації.

Існують також відмінності у способах обліку обмежень на компоненти вектора варійованих параметрів. В алгоритмі ЕС осіб-нащадків, які не задовольняють хоча б одному з обмежень, що визначають область допустимих значень D , не включають у нову популяцію. Якщо ж число таких осіб перевищує деяку заздалегідь задану величину, то запускається процес адаптації параметрів алгоритму (наприклад, збільшується ймовірність мутації). У класичному ГА значення вільних параметрів генетичних операторів є стаціонарними, і зазначені обмеження враховують за допомогою штрафних функцій.

Наразі багато ГА значно відрізняються від класичного, тому відмінності між ними і ЕС можуть бути незначними.

Основою канонічного алгоритму ЕС є комбінування генетичних операторів мутації і селекції наступним чином:

1. Вибираємо з поточної популяції S особу s_i і шляхом застосування до неї оператора мутації отримуємо особу s'_i .

2. Обчислюємо пристосованості $\varphi(s_i)$, $\varphi(s'_i)$ вказаних осіб і розміщуємо в наступну популяцію $S(t+1)$ найбільш пристосовану з них.

Тут індекс i послідовно приймає значення $i=1,2,\dots,|S|$. Алгоритм селекції, що використовується на останньому кроці – це алгоритм селекції відсіканням.

Для генерації випадкової початкової популяції $S(0)$ необхідно, щоб для кожної компоненти $x_j, j \in [1:|X|]$ вектору варійованих параметрів X був заданий інтервал $[x_j^{\min}; x_j^{\max}]$ (використовуваний тільки на етапі ініціалізації). Для особи $s_i, i \in [1:|S|]$ початкові значення компонент $x_{i,j}, j \in [1:|X|]$ вектору X_i вважають рівномірно розподіленими в зазначеному інтервалі.

Як і ГА, алгоритм ЕС в якості умови зупинки може використовувати наступні події:

- досягнення заданого числа поколінь \hat{t} ;
- досягнення заданого стану популяції, коли, наприклад, пристосованості усіх осіб популяції змістились нижче деякого порогу;
- досягнення заданого рівня збіжності, коли у деякій метриці відстані між особами популяції менше певного порогового значення.

Наразі відомі так звані $(\alpha+\beta)$ - та $\alpha\beta$ -алгоритми ЕС. При цьому $\alpha=|S|$ – число осіб (батьків) поточної популяції, $\beta=|S'| > \alpha$ – число осіб-нащадків.

$(\alpha+\beta)$ -алгоритм передбачає, що в проміжну популяцію S' входять усі особи, й з їх числа проводять селекцію і включення в популяцію $S(t+1)$ найкращих осіб числом α . Цей алгоритм забезпечує відмінну від нуля ймовірність потрапляння в популяцію $S(t+1)$ деяких батьківських осіб. Слід підкреслити, що при вищенаведених позначеннях канонічний ГА відноситься до класу $(\alpha+\beta)$ -алгоритмів ЕС.

$\alpha\beta$ -алгоритм передбачає, що в проміжну популяцію S' включають тільки β осіб-нащадків, й з їх числа здійснюють відбір і включення в популяцію S' найкращих осіб числом α . Цей алгоритм виключає можливість попадання у популяцію $S(t+1)$ батьківських осіб.

Вільним параметром $(\alpha+\beta)$ - та $\alpha\beta$ -алгоритмів ЕС є співвідношення $r = \frac{\beta}{\alpha} \geq 7$. Зі збільшенням значень цього параметра зростає ймовірність того, що

кожна особа-батько виробить щонайменше одну особу-нащадка, більш пристосовану, ніж вона сама. При зростанні значень параметру r відмінності між $(\alpha+\beta)$ - і $\alpha\beta$ -алгоритмами стають менш суттєвими.

Наразі для $(\alpha+\beta)$ -алгоритму доказана його збіжність за ймовірністю, а для $\alpha\beta$ -алгоритму питання збіжності залишається відкритим.

Взагалі, $(\alpha+\beta)$ -алгоритм більш «економно», у порівнянні з $\alpha\beta$ -алгоритмом, використовує знайдені високопристосовані особи, оскільки залишає їх для конкуренції з нащадками. Ця відмінність $(\alpha+\beta)$ -алгоритму підвищує інтенсивність пошуку, але зменшує його широту, й може призвести до передчасної збіжності алгоритму до деякого локального мінімуму фітнес-функції.

Оператор мутації в канонічному алгоритмі ЕС є єдиним еволюційним оператором. Він використовує алгоритм, схожий на алгоритм гаусовської мутації, у відповідності з яким значення гену x'_j хромосоми X' визначають на основі значення гену x_j хромосоми X наступним чином:

$$x'_j = x_j + N_1(m, \sigma), \quad j \in [1:|X|] \quad (3.10)$$

де $N_1(m, \sigma)$ – функція одномірного нормального розподілу. У (3.10) середнє квадратичне відхилення σ інтерпретують як вірогідність мутації. Зазвичай використовують нульове математичне очікування $m=0$, а значення величини σ змінюють в процесі функціонування алгоритму в режимі само адаптації таким чином, що оператор мутації виявляється таким, що само адаптується. При цьому параметр σ може включатися в розширений вектор варійованих параметрів задачі, який еволюціонує за загальними правилами зміни осіб алгоритму ЕС.

У найпростішому випадку зміна середньоквадратичного відхилення σ може здійснюватись за «правилом 1/5». Згідно з цим правилом через кожне фіксоване число ітерацій перевіряють число успішних мутацій, і якщо їх відносна кількість перевищує 1/5 (популяція зосереджена навколо деякого оптимуму), то величину σ збільшують за деяким законом, інакше – зменшують.

3.6. Еволюційне програмування

Розглянемо задачу (3.1) глобальної умовної оптимізації цільової і фітнес-функції с обмеженнями типу нерівностей, коли шукається мінімум (а не максимум).

Основні відмінності алгоритму ЕП від алгоритму ЕС полягають у наступному. Перш за все, алгоритм ЕП використовує тільки $(\alpha+\beta)$ -стратегію і тільки при $\alpha=\beta$ (коли число осіб $|S|, |S'|$ поточної і проміжної популяції однакові). По-друге, останній алгоритм допускає використання як дійсного, так і будь-якого іншого кодування осіб.

Розглянемо випадок дійсного кодування. Вважаємо, що фітнес-функція $\varphi(X)$ задовольняє умові $\varphi(X^*)=0$. Рекомендований розмір популяції в алгоритмі ЕП дорівнює $|S|>200$. Алгоритм використовує наступні основні оператори й процедури.

Процедура ініціалізації призначена для генерації випадкової початкової популяції $S(0)$ й може бути реалізована за загальною для популяційних алгоритмів схемою.

Оператор реплікації реалізує створення першої частини проміжної популяції S' шляхом копіювання до неї поточної популяції S таким чином, що в популяції S' виявляється $|S'|=|S|$ осіб.

Оператор мутації застосовують до кожної з осіб $s_i \in S, i \in [1:|S|]$, й отримані особи заносять в проміжну популяцію як другу її частину. Отже, після виконання даного оператора в популяції S' виявляється $|S'|=2|S|$ осіб.

Алгоритм мутації близький до алгоритму гаусовської мутації і визначається наступним чином:

$$x'_{|S|+i,j} = x_{i,j} + N_1(0,1)\sqrt{a_j\varphi(X_i) + b_j}, \quad (3.11)$$

де $i \in [1:|S|]; j \in [1:|X|]; a_j, b_j$ – довільні параметри оператора, рекомендовані значення яких дорівнюють 1 і 0, відповідно.

Співвідношення (3.11) означає, що стандартне відхилення випадкової величини $N_1(0;1)\sqrt{a_j\phi(X_i)+b_j}$ (ширина мутації) пропорційне пристосованості батьківської особи s_i . Якщо глобальний мінімум фітнес-функції (як було прийнято вище) дорівнює нулю, то в міру наближення до точки X^* ширина мутації буде скорочуватися. Якщо фітнес-функція не задовольняє вказаній умові, то для його дотримання відповідним чином підбирають значення параметрів a_j, b_j .

Оператор селекції забезпечує формування популяції $S(t+1)$, що містить $|S|$ осіб, на основі проміжної популяції S' з $|S'| = 2|S|$ осіб. Оператор використовує турнірну селекцію і містить наступні кроки:

1. Встановлення $i=1$.

2. З популяції S' вибирається особа s'_i .

3. За допомогою генератора випадкових чисел, рівномірно розподілених в інтервалі $[1 : 2|S|]$, з популяції S' вибираються такі особи $s'_{i_1}, s'_{i_2}, \dots, s'_{i_n}$, для яких $i_j \neq i, j \in [1 : n]$. Тут n – натуральне число (довільний параметр оператора), рекомендоване значення якого лежить в інтервалі $[0,05|S| : 0,1|S|]$.

4. Порівнюється пристосованість особи s'_i з пристосованостями відібраних осіб і визначається число «перемог» v_i особи s'_i наступним чином:

$$v_i = \sum_{j=1}^n \begin{cases} 1, & \varphi(s'_i) \leq \varphi(s'_{i_j}) \\ 0, & \text{інакше} \end{cases} \quad (3.12)$$

5. Якщо популяція S' не вичерпана, то вважається, що $i=i+1$, й виконується перехід до кроку 2. В іншому випадку виконується перехід до наступного кроку 6.

6. Популяція S' представляється у вигляді списку, особи якого впорядковані у порядку убутання числа їх перемог, а при однаковій кількості перемог – в порядку зростання значень їх пристосованості.

7. В популяцію $S(t+1)$ включаються перші $|S|$ осіб із зазначеного списку.

Слід зазначити, що розглянутий механізм формування популяції $S(t+1)$ не виключає можливості включення в неї малоприспосованих осіб. З ростом значення параметру n цей механізм починає все більше наближатися до елітарного відбору.

Як *процедуру зупинки* алгоритм ЕП може використовувати ті ж самі події, що й алгоритм ЕС.

Канонічний алгоритм ЕП складається з наступних етапів.

1. Ініціалізація початкової популяції $S(0)$ і встановлення лічильника числа поколінь в значення $t=0$.

2. Формування проміжної популяції S' шляхом послідовного виконання операторів реплікації і мутації.

3. Формування наступної популяції $S(t+1)$ за допомогою оператора селекції.

4. Перевірка умови закінчення ітерації. Якщо ця умова виконана, то завершення обчислень, інакше встановлення $t=t+1$ і перехід до етапу 2.

Як і для ГА, довільні параметри алгоритму ЕП слід вибирати таким чином, щоб забезпечити бажаний баланс між інтенсивністю і широтою пошуку.

Для варіанту канонічної схеми алгоритму ЕП, в якому довільні параметри оператора мутації мають значення $a_j=1$, $b_j=0$, $j \in [1:|X|]$, відомий доказ його асимптотичної збіжності за ймовірністю, засноване на теорії марковських ланцюгів. Доказ використовує припущення, що $\varphi(X) \geq 0 \quad \forall X \in D$.

Відомо велика кількість модифікацій канонічного алгоритму ЕП. Наприклад, процедура формування проміжної популяції може бути модифікована наступним чином.

1. За розглянутою схемою виконується оператор реплікації таким чином, що в проміжній популяції S' виявляється $|S'|=|S|$ осіб.

2. За допомогою генератора натуральних випадкових чисел, рівномірно розподілених в інтервалі $[1:|S|]$, з популяції S' вибирається особа s'_i .

3. Реалізується мутація батьківської особи s'_i і отримана особа розміщується в популяцію S' , так що число осіб в ній стає рівним $|S|+1$.

4. Виконується детермінована селекція отриманої популяції S' – з неї виключається найменш пристосована особа, у результаті чого в популяції знову залишається $|S|$ осіб.

5. Кроки 2-5 повторюються декілька раз (наприклад, $|S|$ раз) або до досягнення заданої якості популяції S' (наприклад заданого рівня її різноманітності).

Наразі, існує ряд проблем, що виникають при практичному використанні алгоритму ЕП. По-перше, якщо глобальний мінімум фітнес-функції відрізняється від нуля, то алгоритм може втрачати стійкість, а якщо значення цієї функції досить великі, то пошук стає квазівипадковим. У той же час, в ситуації, коли апріорна інформація про величину глобального мінімуму відсутня (що типово для практичних задач оптимізації), скрутним є підбір значень параметрів a_j , b_j , $j \in [1:|X|]$ у формулі (3.11), які дозволяють трансформувати фітнес-функцію таким чином, щоб можна було уникнути зазначеного ефекту.

По-друге, алгоритм ЕП має у загальному випадку приблизно $2|X|$ вільних параметрів. Вибір оптимальних або хоча б розумних значень цих параметрів є серйозною проблемою. Слід підкреслити, що ГА також не вільний від цієї проблеми.

3.7. Диференціальна еволюція

Алгоритм диференціальної еволюції (ДЕ) був розроблений Прайсом (K. Price) і Сторном (R. Storn) у 1995 р. Алгоритм ДЕ використовує адаптивну мутацію осіб за правилом, подібному «правилу 1/5» (див. п. 3.5). При цьому, якщо особи популяції сильно розподілені в просторі пошуку, та алгоритм мутації проводить їх істотні варіації. Інакше, якщо особи розподілені в просторі пошуку компактно, алгоритм реалізує невеликі зміни осіб. На відміну від алгори-

тмів ЕС та ЕП тут мова йде про задачу глобальної умовної максимізації цільової та фітнес-функції з обмеженнями типу нерівностей (3.1).

Класичний адаптивний оператор мутації для кожної особи поточної популяції s створює особу нащадка s' на основі трьох інших різних батьківських осіб s_1, s_2, s_3 , випадковим чином вибраних з цієї популяції. Тут s_1 – базова особа. Хромосому X' отримують шляхом лінійної комбінації хромосом X_1, X_2, X_3 , інтерпретованих як дійсні вектори:

$$X' = X_1 + \alpha(X_2 - X_3), \quad (3.13)$$

де α – скалярний коефіцієнт впливу (вільний параметр).

На початкових стадіях алгоритму ДЕ, коли популяція S сильно розподілена (її особи знаходяться «далеко» одна від одної), модуль вектору мутації ($X_2 - X_3$), найвірогідніше, буде мати велике значення, й вектор X' виявиться таким, що сильно відрізняється від вектору X_1 . Навпаки, на заключних ітераціях пошуку, коли усі особи поточної популяції знаходяться «поблизу» один одного (і «поблизу» глобального або локального екстремуму фітнес-функції), модуль вказаного вектора з високою вірогідністю буде мати невелике значення, так що вектори X', X_1 виявляються також «близькими».

Після мутації визначають пристосованість отриманої особи s' і якщо вона вище пристосованості батьківської особи s_1 , замінюють останню особою s' .

Розглянемо ситуацію, коли множина допустимих значень вектора змінних параметрів є гіперпаралелепіпедом Π . Взагалі, алгоритм ДЕ може використовувати бінарне, цілочисельне та дійсне кодування осіб. Обмежимось розглядом дійсного алгоритму ДЕ.

Початкову популяцію $S(0) = \{s_i(0), i \in [1:|S|]\}$ в алгоритмі ДЕ отримують наступним чином:

$$x_{i,j} = x_i^- + U_1(x_i^-, x_i^+), \quad j \in [1:|X|] \quad (3.14)$$

де U_1 – функція одномірного рівномірного розподілу.

Кожна ітерація алгоритму ДЕ містить наступні основні кроки.

1. В поточній популяції $S = \{s_i(0), i \in [1:|S|]\}$ вибирається особа s_i .

2. За допомогою генератора натуральних випадкових чисел, рівномірно розподілених в інтервалі $[1:|S|]$, з популяції S вибираються особи $s_{i_1}, s_{i_2}, s_{i_3}$; $i_1 \neq i_2 \neq i_3 \neq i_1$.

3. За допомогою аналогічного генератора випадкових чисел знаходиться індекс $j_1 \in [1:|X|]$ гена в хромосомі.

4. Формується особа s'_i така, що

$$x'_{i,j} = \begin{cases} x_{i,j} + \alpha(x_{i_2,j} - x_{i_3,j}), & U_1(0;1) \leq b \text{ та } (j = j_1) \\ x_{i,j}, & U_1(0;1) > b \end{cases}, \quad (3.15)$$

де $j \in [1:|X|]$, $\alpha \in [0;2] = \text{const}$, $b \in [0;1]$ – вільний параметр алгоритму.

5. Проводиться селекція осіб за наступною схемою:

$$s'_i = \begin{cases} s'_i, & \varphi(s'_i) \geq \varphi(s_i) \\ s_i, & \text{інакше} \end{cases}. \quad (3.16)$$

Слід зазначити, що відповідно до (3.15) випадковий індекс j_1 в кожній з осіб $s_i(t)$ поточної популяції забезпечує зміну хоча б одного гену.

Формула (3.15) не гарантує, що усі компоненти отриманого вектора X'_i виявляться допустимими, тобто не вийдуть за межі гіперпаралелепіпеду Π . Якщо обмеження порушено для компоненти $x'_{i,j}$ так що $x'_{i,j} < x_i^-$ або $x'_{i,j} > x_i^+$, то обчислюється нове значення цієї компоненти:

$$x'_{i,j} = x_i^- + U_1(x_i^-, x_i^+), \quad (3.17)$$

У разі, коли множина допустимих значень вектора змінних параметрів визначає функціональні обмеження типу нерівностей $G(X) \geq 0$, ці обмеження враховуються за допомогою алгоритму штрафних функцій. При цьому фітнес-функція $\varphi(X)$ формується у вигляді:

$$\varphi(X, A) = (f(X) + b) \prod_{k=1}^{|G|} (\tilde{g}_k(X, A))^{c_i}, \quad (3.18)$$

де $A = (a_1, a_2, \dots, a_{|G|})$ – вектор ненегативних коефіцієнтів штрафу; $b, c_i \geq 1$ – вільні параметри алгоритму;

$$\tilde{g}_k(X, A) = \begin{cases} 1 + a_k g_k(X), & g_k(X) < 0, \\ 1, & \text{інакше.} \end{cases} \quad (3.19)$$

Значення параметру b обирається таким чином, щоб функція $\varphi(X, A)$ була ненегативна при усіх $X \in D$.

Відома велика кількість модифікацій канонічного алгоритму ДЕ. Наприклад, як базову особу можна використовувати найбільш пристосовану особу, а не ту, яку випадково вибрано з поточної популяції. Другий підхід полягає у використанні замість вектора мутації $(X_{i_2} - X_{i_3})$ вектора градієнта функції $\varphi(X)$ (такий спосіб побудови вектора мутації породжує гібридний метод оптимізації, заснований на використанні пришвиджуючого оператора). Ще один приклад модифікації канонічного алгоритму ДЕ полягає у використанні оператора *міграції* у випадках, коли популяція занадто однорідна.

3.8. Метод групового урахування аргументів

Метод групового урахування аргументів (МГУА) запропонований академіком О.Г. Івахненком (Інститут кібернетики НАН України) у 1968 р. МГУА – це сімейство індуктивних алгоритмів для математичного моделювання багатопараметричних даних. Метод заснований на рекурсивному селективному відборі моделей, на основі яких будуються складніші моделі. Точність моделювання на кожному наступному кроці рекурсії збільшується за рахунок ускладнення моделі.

3.8.1. Метод найменших квадратів.

Перед тим, як починати розгляд МГУА, було б корисно згадати метод найменших квадратів (МНК) – найбільш поширений метод підстроювання лінійно незалежних параметрів.

Розглянемо для прикладу МНК для 3 аргументів. Нехай функція $T=T(U, V, W)$ задана таблицею, тобто з досвіду відомі числа U_i, V_i, W_i, T_i ($i=1, \dots, n$).

Знайдемо залежність між цими даними у вигляді:

$$T(U, V, W) = aU + bV + cW, \quad (3.20)$$

де a, b, c – невідомі параметри.

Підберемо значення цих параметрів таким чином, щоб була найменшою сума квадратів відхилень експериментальних T_i і теоретичних $T_i = aU_i + bV_i + cW_i$ даних:

$$\sigma = \sum_{i=1}^n (T_i - aU_i - bV_i - cW_i)^2 \rightarrow \min. \quad (3.21)$$

Величина σ є функцією трьох змінних a, b, c . Необхідною і достатньою умовою існування мінімуму цієї функції є рівність нулю частинних похідних функції σ по всім змінним, тобто:

$$\frac{\partial \sigma}{\partial a} = 0, \quad \frac{\partial \sigma}{\partial b} = 0, \quad \frac{\partial \sigma}{\partial c} = 0 \quad (3.22)$$

Оскільки:

$$\begin{aligned} \frac{\partial \sigma}{\partial a} &= -2 \sum_{i=1}^n (T_i - aU_i - bV_i - cW_i) U_i; \\ \frac{\partial \sigma}{\partial b} &= -2 \sum_{i=1}^n (T_i - aU_i - bV_i - cW_i) V_i; \\ \frac{\partial \sigma}{\partial c} &= -2 \sum_{i=1}^n (T_i - aU_i - bV_i - cW_i) W_i, \end{aligned} \quad (3.23)$$

то система для знаходження a, b, c матиме вигляд:

$$\begin{aligned} a \sum_{i=1}^n U_i^2 + b \sum_{i=1}^n U_i V_i + c \sum_{i=1}^n U_i W_i &= \sum_{i=1}^n T_i U_i; \\ a \sum_{i=1}^n U_i V_i + b \sum_{i=1}^n V_i^2 + c \sum_{i=1}^n V_i W_i &= \sum_{i=1}^n T_i V_i; \\ a \sum_{i=1}^n U_i W_i + b \sum_{i=1}^n W_i V_i + c \sum_{i=1}^n W_i^2 &= \sum_{i=1}^n T_i W_i. \end{aligned} \quad (3.24)$$

Ця система вирішується будь-яким стандартним методом розв'язання систем лінійних рівнянь (Гауса, Жордана, Зейделя, Крамера).

Якщо ми поширимо МНК на випадок з m параметрами,

$$\sigma = \sum_{i=1}^n \left(T_i - \sum_{v=1}^m u_{iv} c_v \right)^2 \rightarrow \min \quad (3.25)$$

то шляхом міркувань, аналогічних наведеним вище, отримаємо наступну систему лінійних рівнянь:

$$\begin{cases} \overline{c_1 u_1 u_1} + \overline{c_2 u_1 u_2} + \dots + \overline{c_m u_1 u_m} = \overline{T u_1} \\ \overline{c_1 u_2 u_1} + \overline{c_2 u_2 u_2} + \dots + \overline{c_m u_2 u_m} = \overline{T u_2} \\ \dots \\ \overline{c_1 u_m u_1} + \overline{c_2 u_m u_2} + \dots + \overline{c_m u_m u_m} = \overline{T u_m} \end{cases} \quad (3.26)$$

де $\overline{T} = \{T_i\}_{i=1}^m$, $\overline{u}_v = \{u_{iv}\}_{i=1}^n$

3.8.2. Загальна схема побудови алгоритмів методу.

Запозичення у природи алгоритмів переробки інформації є однією з основних ідей кібернетики. «Гіпотеза селекції» стверджує, що алгоритм масової селекції рослин або тварин є оптимальним алгоритмом переробки інформації в складних задачах. При масовій селекції висівається певна кількість насіння. В результаті запилення утворюються складні спадкові комбінації. Селекціонери вибирають деяку частину рослин, у яких властивість, яка їх цікавить, виражена найбільше (евристичний критерій). Насіння цих рослин збирають і знову висівають для утворення нових, ще більш складних комбінацій. Через декілька поколінь селекція зупиняється, і її результат є оптимальним. Якщо надмірно продовжувати селекцію, то настане *інцухт* – виродження рослин. Існує оптимальне число поколінь і оптимальна кількість насіння, що відбираються у кожному з них.

Алгоритми МГУА відтворюють схему масової селекції (рис. 3.5). У них є генератори комбінацій, які ускладнюються із ряду в ряд, і порогові самовідбори кращих з них. Так званий «повний» опис об'єкта має вигляд

$$\varphi = f(x_1, x_2, x_3, \dots, x_m), \quad (3.27)$$

де f – деяка елементарна функція, наприклад, ступеневий поліном, замінюється декількома рядами «часткових» описів:

$$1 \text{ ряд селекції: } y_1 = f(x_1x_2), y_2 = f(x_1x_3), \dots, y_s = f(x_{m-1}x_m),$$

$$2 \text{ ряд селекції: } z_1 = f(y_1y_2), z_2 = f(y_1y_3), \dots, z_p = f(y_{s-1}y_s),$$

де $s=c^2, p=c_s^2$ і так далі.

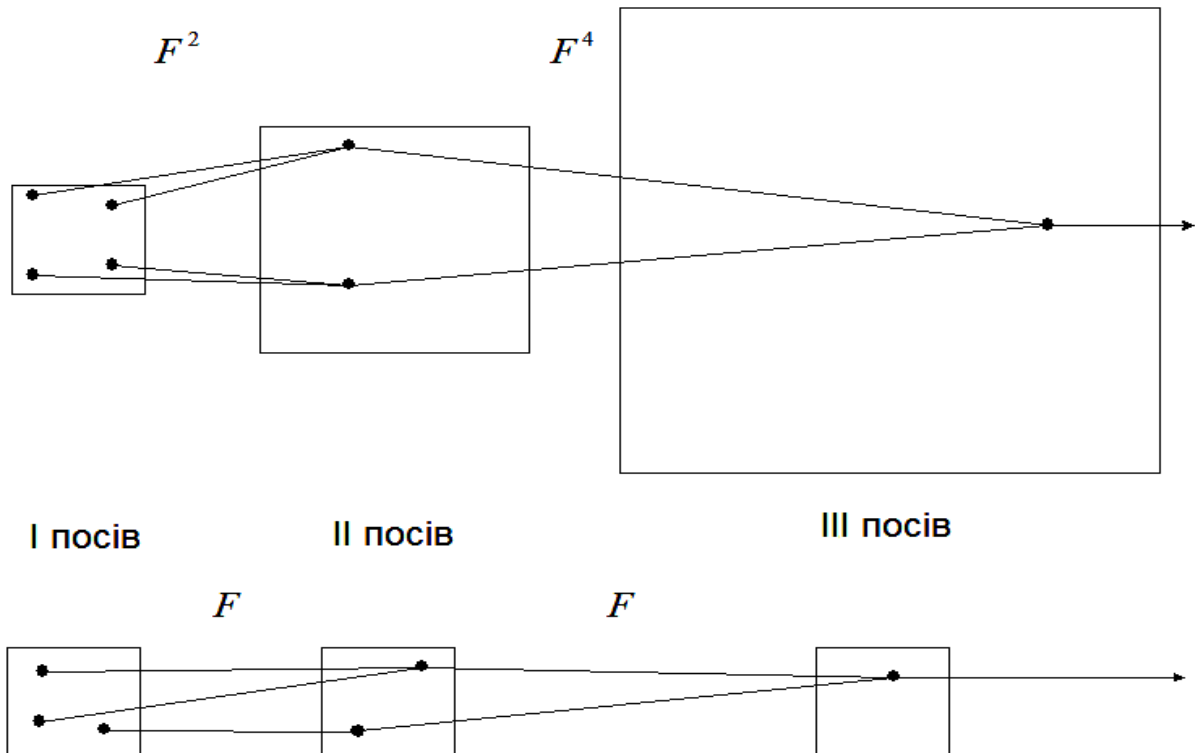


Рис. 3.5. Селекція самого чорного тюльпана при дослідному полі, яке розширюється (еквівалент повного перебору), і при постійному розмірі поля (еквівалент селекції при збереженні свободи вибору рішень $F=const$)

Вхідні аргументи і проміжні змінні сполучаються попарно, і складність комбінацій на кожному ряду обробки інформації зростає (як при масовій селекції), поки не буде отримана єдина модель оптимальної складності.

Кожний частковий опис є функцією тільки двох аргументів. Тому його коефіцієнти легко визначити за даними навчальної послідовності при малому числі вузлів інтерполяції. Виключаючи проміжні змінні (якщо це вдається),

можна отримати «аналог» повного опису. Математика не забороняє обидві ці операції. Наприклад, по десяти вузлам інтерполяції можна отримати в результаті оцінки коефіцієнтів полінома сотого ступеня і так далі.

З ряду в ряд селекції пропускається тільки деяка кількість самих регулярних змінних. Ступінь регулярності оцінюється за величиною середньоквадратичної похибки (середньої для всіх вибраних в кожному поколінні змінних або для однієї найточнішої змінної) на окремій перевірочній послідовності даних. Іноді в якості показника регулярності використовується коефіцієнт кореляції.

Ряди селекції нарощуються до тих пір, поки регулярність підвищується. Як тільки досягнуто, наприклад, мінімум похибки, селекцію, щоб уникнути інцухта, слід зупинити. Практично рекомендується зупинити селекцію навіть де-що раніше досягнення повного мінімуму, як тільки похибка починає падати занадто повільно. Це призводить до більш простих і більш достовірних рівнянь.

3.8.3. Алгоритм з коваріаціями і з квадратичними описами.

У цьому алгоритмі використовуються часткові описи, які представлені в наступних формулах та на рис. 3.6:

$$y_i = a_0 + a_1x_i + a_2x_j + a_3x_ix_j; \quad (3.28)$$

$$y_k = a_0 + a_1x_i + a_2x_j + a_3x_ix_j + a_4x_i^2 + a_5x_j^2. \quad (3.29)$$

Складність моделі збільшується від ряду до ряду селекції як за кількістю врахованих аргументів, так і за ступенем. Ступінь повного опису швидко зростає. На першому ряді – квадратичні описи, на другому – четвертого ступеня, на третьому – восьмого і так далі. У зв'язку з цим мінімум критерію селекції знаходиться швидко, але не зовсім точно. Крім того, є небезпека втрати істотного аргументу, особливо на перших рядах селекції (у разі відсутності протекції). Спеціальні теореми теорії МГУА визначають умови, при яких результат селекції не відрізняється від результату повного перебору моделей.

Для того, щоб ступінь повного рівняння підвищувалася з кожним рядом селекції на одиницю, досить розглядати усі аргументи і їх коваріації як узагальнені аргументи і користуватися складеними для них лінійними описами.

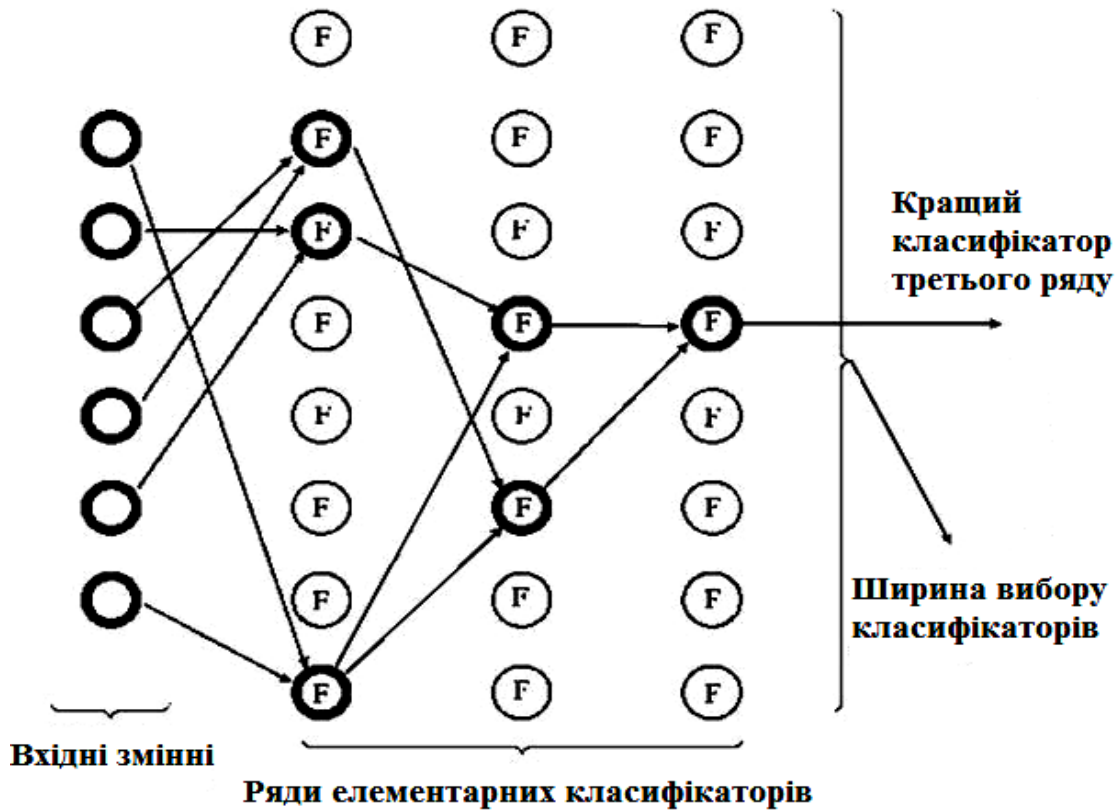


Рис. 3.6. МГУА як еквівалент масової селекції

3.8.4. Критерії селекції методу.

Критерій регулярності визначає середньоквадратичне відхилення моделі на перевірочній вибірці (N_B):

$$\Delta^2(B) = \sum_{t \in N_b} (y_t^M - y_t)^2 / \sum_{t \in N_B} y_t^2 \rightarrow \min. \quad (3.30)$$

Критерій регулярності рекомендується використовувати в якості допоміжного (регуляризуючого). Основними критеріями при цьому можуть бути критерій мінімуму зміщення або критерій балансу змінних. У разі самостійного використання критерію регулярності збільшується стійкість вибору моделі.

Критерій мінімуму зміщення (несуперечності) моделі n_{cm}^2 . Дорівнює середньоквадратичному значенню відхилень виходів моделей A і B для всієї вибірки ($t \in N$):

$$n_{cm}^2 = \sum_{t \in N_b} (y_t^A - y_t^B)^2 / \sum_{t \in N_B} y_t^2 \rightarrow \min. \quad (3.31)$$

Інтерпретація критерію така: модель, оцінка якої отримана за даними певного інтервалу спостереження або в певній точці спостереження, повинна якомога ближче збігатися із моделлю, отриманою за даними іншого інтервалу спостереження або в іншій точці спостереження. Критерій є виразом вимоги несуперечності моделі.

Критерій для малих вибірок. Усереднений критерій регулярності (УКР) дорівнює середньому значенню критеріїв регулярності, розрахованих для випадку виділення з усіх точок вимірювань по одній в перевірочну вибірку. Зі спостережень вибирають кілька точок, які по черзі утворюють перевірочну вибірку, тоді як інші реалізації служать для самоорганізації моделі. Для n реалізацій справедливий вираз:

$$УКР = \Delta^m(B) = \frac{1}{n} \sum_{i=1}^n \Delta_i(B) \rightarrow \min . \quad (3.32)$$

Критерій точності короткострокового прогнозу. При самоорганізації прогнозуючих моделей рекомендується виділення частини даних в екзаменаційну вибірку N_c для перевірки якості прогнозу. На ній визначається середньоквадратичне відхилення обчисленого за допомогою моделі y_t^M (передбаченого) значення від істинного y_t :

$$RR = \Delta^2(C) = \sum_{t \in N_c} (y_t^M - y_t)^2 / \sum_{t \in N_c} y_t^2 \rightarrow \min . \quad (3.33)$$

Він може застосовуватися у комбінації з критерієм регулярності.

Критерій балансу. Згідно із критерієм балансу, з усіх отриманих моделей обирається та, яка в певному інтервалі часу найкраще відповідає заданій закономірності.

Контрольні запитання

1. Наведіть схему канонічного генетичного алгоритму.

2. Які генетичні оператори містить генетичний алгоритм? Розкрийте їх особливості.
3. Наведіть схеми наступних типових генетичних алгоритмів: острівний генетичний алгоритм, генетичний алгоритм Genitor, генетичний алгоритм СНС.
4. Наведіть визначення основних понять теорії схем. Сформулюйте теорему схем.
5. Назвіть основні особливості алгоритму еволюційної стратегії. Наведіть схему канонічного алгоритму еволюційної стратегії.
6. Перерахуйте основні оператори алгоритму еволюційного програмування. Наведіть схему цього алгоритму.
7. Наведіть схему алгоритму диференціальної еволюції.
8. Розкрийте суть методу найменших квадратів.
9. Наведіть схему побудови алгоритмів методу групового урахування аргументів.
10. Які існують критерії селекції методу групового урахування аргументів?

Глава 4. АГЕНТСЬКІ АЛГОРИТМИ ОПТИМІЗАЦІЇ

До популяційних алгоритмів оптимізації відносять не тільки еволюційне моделювання, розглянуте в розд. 3, а й інспіровані живою природою агентські алгоритми (ройовий інтелект), серед яких отримали поширення алгоритми рою часток, мурашиної колонії та бджолиного рою, які й розглядаються в даній главі книги.

4.1. Оптимізація роєм часток

В основу *алгоритму оптимізації роєм часток* (Particle Swarm Optimization, PSO) покладено соціально-психологічну поведінкову модель натовпу. Розвиток алгоритму інспірували такі завдання, як моделювання поведінки птахів у зграї та риб у косяку. Метою було виявити базові принципи, завдяки яким, наприклад, птиці у зграї поводяться дивовижно синхронно, змінюючи як по команді напрямок свого руху, так що зграя рухається як єдине ціле. До сучасного часу концепція алгоритму рою часток розвинулась в високоефективний алгоритм оптимізації, який часто становить конкуренцію кращим модифікаціям ГА.

Існує досить багато варіантів алгоритмів рою часток. У канонічному алгоритмі, запропонованому у 1995 р. Кеннеді (J. Kennedy) і Еберхартом (R. Eberhart), при визначенні наступного положення частки враховується інформація про найкращу частку з числа її «сусідів», а також інформація про дану частку на тій ітерації алгоритму, на якій цій частці відповідало оптимальне значення фітнес-функції.

4.1.1. Канонічний алгоритм рою часток.

Розглянемо задачу багатовимірної глобальної безумовної мінімізації. У канонічному алгоритмі PSO координати частки s_i на ітерації t визначає вектор

$X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,|X|})^T$, а на ітерації $(t+1)$ – вектор $X'_i = (x'_{i,1}, x'_{i,2}, \dots, x'_{i,|X|})^T$. Початкові координати частки s_i дорівнюють $X_i(0)$; $i \in [1 : |S|]$.

Ітерації в алгоритмі PSO виконуються за схемою:

$$X'_i = X_i + V_i, \quad (4.1)$$

$$V_i = b_I V_i^- + U_{|X|}(0; b_C) \otimes (X_i^* - X_i) + U_{|X|}(0; b_S) \otimes (X_i^{**} - X_i). \quad (4.2)$$

Тут використовуються наступні позначення: $U_{|X|}$ – функція $|X|$ -мірного рівномірного розподілу; \otimes – символ прямого добутку векторів; b_I, b_C, b_S – вільні параметри алгоритму; $V_i = V_i(t) = (v_{i,1}, v_{i,2}, \dots, v_{i,|X|})^T$ – $(|X| \times 1)$ -вектор збільшення координат частки; $V_i^- = V_i(t-1)$; X_i^* – вектор координат частки s_i , що відповідає її найкращому значенню фітнес-функції за час пошуку $[0 : \hat{t}]$, тобто

$$\min_{t \in [0 : \hat{t}]} \varphi(X_i(t)) = \varphi(X_i^*(t)); \quad (4.3)$$

X_i^{**} – вектор координат сусідньої з даною часткою з найкращим за той же час значенням пристосованості, тобто

$$\min_{j \in N_i} \varphi(X_j^*) = \varphi(X_i^{**}), \quad (4.4)$$

де N_i – множина номерів часток, що є «сусідами» даної частки s_i .

Вектор V_i має сенс руху (кроку), який частка s_i здійснює при переході з поточного положення X_i в положення X'_i . Часто цей вектор називають швидкістю частки. В процесі ітерацій вектор $X_i^* = (x_{i,1}^*, \dots, x_{i,|X|}^*)$ утворює так званий *власний шлях* (private guide) частки s_i , а вектор $X_i^{**} = (x_{i,1}^{**}, \dots, x_{i,|X|}^{**})$ – *локальний шлях* (local guide) цієї частки.

Перерахунок координат часток за формулами (4.1), (4.2) може відбуватися за *синхронною схемою*, коли оновлення координат часток виконується тільки після визначення поточних приростів V_i координат усіх $|S|$ часток, або за *асинхронною схемою*, коли розрахунок нових координат відбувається до завершення зазначених обчислень.

Перший доданок у (4.2) – так звана *інерційна компонента* (позначимо її V_i^I), що виконує функції пам'яті частки про її переміщенні на попередній ітерації. Компонента попереджає сильні стрибкоподібні зміни напрямку руху частки.

Другий доданок у (4.2) – так звана *когнітивна компонента* (позначимо її V_i^C), що формалізує тенденцію частки s_i повернутися з поточного положення X_i в положення X_i^* , в якому вона мала на попередніх ітераціях найкраще значення фітнес-функції. Ця компонента є випадковим вектором, який, при рівних значеннях складових вектора $U_{|X|}(0; b_C)$, спрямований з точки X_i в точку X_i^* і довжина якого пропорційна евклідовій відстані між цими точками. Дана компонента реалізує функцію пам'яті частки про своє найбільш оптимальне положення.

Третій доданок у (4.2) – так звана *соціальна компонента* (позначимо її V_i^S), що відображає прагнення частки s_i переміститися у напрямку своєї найбільш успішної сусідньої частки s_i^{**} . Аналогічно когнітивній соціальна компонента є випадковим вектором з довжиною, пропорційною відстані між точками X_i^{**} , X_i , спрямований при рівних значеннях складових вектора $U_{|X|}(0; b_S)$ з точки X_i в точку X_i^{**} . Тобто ця компонента є реалізацію пам'яті частки про оптимальне положення її найбільш успішної сусідньої частки.

Сумарний ефект взаємодії зазначених компонентів для двовимірного вектору змінних параметрів ілюструє рис. 4.1, на якому прийнято, що всі компоненти випадкових векторів $U_{|X|}(0; b_C)$, $U_{|X|}(0; b_S)$ однакові, $X_i^*(t+1) = X_i^*(t)$, $X_i^{**}(t+1) = X_i^{**}(t)$.

Вільний параметр b_I визначає вагу «інерційних» властивостей частки (при $b_I < 1$ її рух, очевидно, сповільнюється). Рекомендоване значення параметру $b_I = 0,7298$. Відомо, що в процесі оптимізації може бути ефективним поступове зменшення коефіцієнта b_I від 0,9 до 0,4. При цьому великі значення параметра

забезпечують широкий огляд простору пошуку, а малі – точну локалізацію мінімуму фітнес-функції.

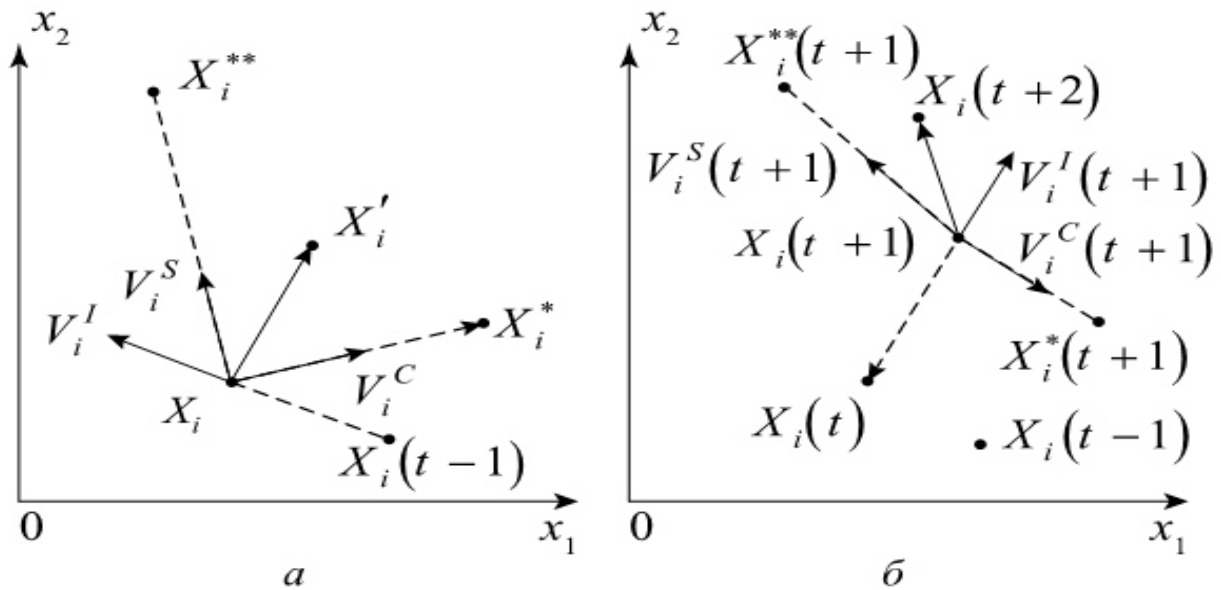


Рис. 4.1. Ілюстрація схеми алгоритму PSO, $|X|=2$:

a – покоління $(t+1)$; b – покоління $(t+2)$

Значення параметрів b_C , b_S визначають відносні ваги когнітивної та соціальної компонент. У разі нульових значень цих параметрів траєкторія частки є прямою, що йде в нескінченність. При $b_C > 0$, $b_S = 0$ частки намагаються досягти мінімуму фітнес-функції незалежно одна від одної. Успіх рішення задачі оптимізації алгоритмом PSO забезпечує координація руху часток, яку визначає значення параметру b_S . При малих значеннях параметрів b_C , b_S частки переміщуються по гладких траєкторіях. Зі збільшенням значень цих параметрів рух часток стає все більш стохастичним. Рекомендовані значення параметрів $b_C = b_S = 1,49618$.

Як і в будь-якому популяційному алгоритмі пошукової оптимізації в алгоритмі PSO зі збільшенням розміру популяції підвищується ймовірність локалізації глобального екстремуму фітнес-функції, оскільки така популяція здатна більш рівномірно покривати простір пошуку. З іншого боку при цьому підвищується обчислювальна складність кожної ітерації (хоча, можливо, зменшується

ся їх загальна кількість). Дослідження показують, що алгоритм PSO забезпечує високу ефективність пошуку навіть при використанні малочисельної популяції з 10-30 часток. Загальною рекомендацією є підбір оптимального розміру популяції для кожної задачі окремо.

Оптимальне число ітерацій \hat{t} теж є залежить від конкретної задачі. При малому числі ітерацій робота алгоритму може закінчитись передчасно, до досягнення глобального екстремуму, а при великому – можуть бути безпідставно збільшені обчислювальні витрати.

Алгоритм PSO містить наступні етапи.

1. Задаються значення вільних параметрів алгоритму і ініціалізуються популяції. Встановлюється значення лічильника поколінь $t=1$.

2. Для кожної з часток популяції s_i , $i \in [1 : S]$ знаходяться краща локальна X_i^* та краща глобальна позиція X_i^{**} за формулами (4.3) і (4.4), відповідно.

3. За формулами (4.1) і (4.2) знаходяться нові позиції усіх часток популяції X'_i .

4. Перевіряється виконання умов закінчення ітерацій. Якщо ця умова виконана – закінчення ітерацій, в іншому випадку – встановлення $t=t+1$ і перехід до етапу 2.

Ініціалізація популяції полягає в завданні початкових позицій $X_i(0)$ і швидкостей $V_i(0)$ часток. Зазвичай точки $X_i(0)$ приймають випадковими, рівномірно розподіленими в гіперпаралелепіпеді

$$\Pi = \{X \mid x_j^- \leq x_j \leq x_j^+, j \in [1 : X]\}, \quad (4.5)$$

тобто вважають $x_{i,j}(0) = U_1(x_j^-, x_j^+)$. Початкові швидкості часток також ініціалізують випадковими значеннями.

В якості умов закінчення ітерацій використовують типові для популяційних алгоритмів оптимізації умови:

- досягнення максимально допустимого числа ітерацій;
- досягнення рішення, яке задовольняє особу, яка приймає рішення;

- стагнація ітераційного процесу (відсутність поліпшень фітнес-функції протягом δ_t ітерацій, де δ_t – вільний параметр алгоритму);
- рівність нулю із заданою точністю оцінки першої похідної фітнес-функції в околиці глобально кращої частки або всіх часток популяції.

Замість формули (4.2) часто використовують її варіант

$$V_i = b_r(V_i^- + \tilde{U} \otimes (Y_i - X_i)), \quad (4.6)$$

де

$$\tilde{U} = U_{|X|}(0; b_*) + U_{|X|}(0; b_{**}), \quad (4.7)$$

$$Y_i = \frac{(U_{|X|}(0; b_*) \otimes X_i^*) + (U_{|X|}(0; b_{**}) \otimes X_i^{**})}{\tilde{U}}. \quad (4.8)$$

Вільний параметр b_r у (4.6) – обмежуючий фактор. Рекомендовані значення параметрів b_* , b_{**} у цьому випадку дорівнюють 2,05, а їх оптимальні значення визначає, загалом, клас цільової функції.

У термінах ГА канонічний алгоритм PSO можна визначити наступним чином. Алгоритм використовує тільки дійсне кодування осіб. Відбір і селекція осіб не проводяться (тобто в популяцію S' переходять всі особи поточної популяції S). Ітераційну схему (4.2) можна інтерпретувати як схему специфічного оператора кросовера, побудованого на основі лінійної комбінації векторів X_i , V_i^- , X_i^* , X_i^{**} .

4.1.2. Модифікації канонічного алгоритму рою часток.

Розглянемо найбільш характерні модифікації, які використовують стаціонарні топології сусідства часток.

Притягнуто-відштовхуючий алгоритм PSO (Attractive and Repulsive PSO, ARPSO) запропонований у роботах Вестерстрома (J. Vesterstrom).

Алгоритм PSO функціонує за принципом переміщення області пошуку кожної з часток s_i до її поточного глобально оптимального рішення X_i^{**} , хоча шуканий глобальний мінімум фітнес-функції може перебувати зовсім в іншій

частині простору пошуку. Дана обставина є причиною передчасної збіжності алгоритму PSO. Алгоритм ARPSO в значній мірі вільний від цього недоліку.

Суть алгоритму ARPSO полягає у послідовному чергуванні фази притягнення і фази відштовхування. У фазі притягнення ітерації виконуються за формулами (4.1), (4.2), а у фазі відштовхування остання формула набуває вигляду:

$$V_i = b_I V_i^- - U_{|X|}(0; b_C) \otimes (X_i^* - X_i) - U_{|X|}(0; b_S) \otimes (X_i^{**} - X_i), \quad (4.9)$$

тобто знаки когнітивної і соціальної компонент змінюються на протилежні.

В алгоритмі ARPSO також використовується *проміжна фаза* або *фаза позитивного конфлікту*, що являє собою комбінацію фаз притягнення і відштовхування і визначається наступним чином:

$$V_i = b_I V_i^- + U_{|X|}(0; b_C) \otimes (X_i^* - X_i) - U_{|X|}(0; b_S) \otimes (X_i^{**} - X_i). \quad (4.10)$$

Тобто в цій фазі частка s_i притягується до своєї найкращої позиції X_i^* і відштовхується від найкращій позиції сусідніх часток X_i^{**} .

Алгоритм рою часток з керуванням швидкістю часток. В алгоритмі PSO можлива ситуація, при якій швидкості окремих часток експоненційно зростають з ростом номера покоління. В результаті такі частки можуть залишати частину простору пошуку, яка представляє інтерес, значно зменшуючи при цьому ефективність алгоритму. Алгоритм рою часток з керуванням швидкістю покликаний виключити можливість такої ситуації шляхом постійного контролю швидкості кожної з часток і її зменшення при перевищенні деякого порогу.

Формально цей алгоритм використовує наступне правило керування швидкістю часток (див. (4.2)):

$$v_{i,j} = \begin{cases} v_{i,j}, & v_{i,j} < v_j^{\max}, \\ v_j^{\max}, & \text{інакше,} \end{cases} \quad (4.11)$$

де $i \in [1 : |S|]$; $j \in [1 : |X|]$; v_j^{\max} – максимально допустимі значення відповідних компонент вектора швидкості (вільні параметри алгоритму).

Величини v_j^{\max} мають сильний вплив на баланс між інтенсивністю і широтою пошуку. Так, великі значення цих величин збільшують широту пошуку і

зменшують його інтенсивність; а малі – можуть змусити частки залишитися поблизу знайдених ними глобальних мінімумів, а також можуть збільшити число ітерацій, необхідних для досягнення глобального екстремуму фітнес-функції.

Величини v_j^{\max} рекомендують вибирати за правилом:

$$v_j^{\max} = a(x_j^+ - x_j^-), \quad (4.12)$$

де $j \in [1 : |X|]$; x_j^- , x_j^+ – межі області пошуку по j -му виміру; $a \in (0,1)$ – коефіцієнт, що визначає частину цієї області, яка доступна за одну ітерацію.

Розглянутий механізм обмежує швидкість часток по кожному з вимірів окремо, так що є можливість керувати пошуком в областях, далеких від гіперкуба. Цей же механізм може змінювати напрямки руху часток, що ілюструє рис. 4.2, на якому $\tilde{v}_{i,2}$, \tilde{X}'_i – скориговані компоненти швидкості й положення частки, відповідно.

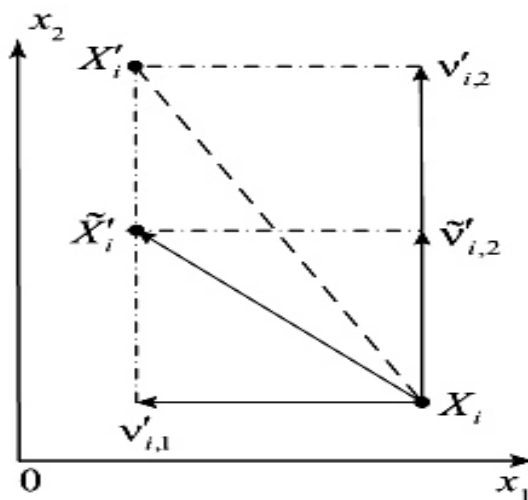


Рис. 4.2. Ілюстрація зміни напрямку руху часток в алгоритмі рою часток з керуванням їх швидкістю

Зміна напрямку руху часток в результаті обмеження їх швидкості має як позитивний (можливе розширення пошуку), так і негативний ефект (зниження його інтенсивності).

Алгоритм рою часток з повною інформацією (Fully Informed Particle Swarm, FIPS) запропонований у роботах Мендеса (R. Mendes). В його основі лежить спостереження, що поведінку особи в зграї визначає не тільки її найбільш «успішний» сусід, а й уся сукупність сусідів. Відмінність алгоритму FIPS від канонічного у тому, що в алгоритмі FIPS враховується вплив на дану частку усіх сусідніх з нею часток. Очевидно, що в цьому випадку різні топології сусідства часток повинні істотно впливати на процес оптимізації. Слід також зазначити, що у формулі (4.2) зникає необхідність в доданку, який містить вектор X_i^{**} , тобто глобальний шлях частки s_i не існує.

В ітераційній схемі алгоритму FIPS координати часток задаються згідно (4.1), де крок V_i визначається за формулою (4.6), а формули (4.7) і (4.8) змінюються на наступні:

$$Y_i = \frac{\sum_{j \in N_i} (\tilde{U}_j \otimes X_j^*)}{\sum_{j \in N_i} \tilde{U}_j}, \quad (4.13)$$

$$\tilde{U}_j = U_{|X|} \left(0; \frac{b_*}{|N_i|} \right), \quad (4.14)$$

де N_i – множина часток, що є сусідами частки s_i .

Алгоритм FIPS зазвичай має кращу збіжність, ніж канонічний алгоритм PSO. Основною проблемою алгоритму FIPS є сильний вплив на його ефективність використовуваної топології сусідства часток. Особливістю алгоритму FIPS є тенденція його збіжності до центру початкового положення часток, у зв'язку з чим рекомендують використовувати асиметричне початкове розташування.

Алгоритм FIPS з вагами (Weighted FIPS, WFIPS) використовує модифікацію канонічного алгоритму PSO, яка полягає у додаванні ваг в останній доданок ітераційної формули (4.6). При відповідному виборі ваг це дозволяє посилити вплив тих сусідніх часток, у яких кращі значення фітнес-функції. Ітераційна формула алгоритму WFIPS виглядає наступним чином:

$$V_i = b_r \left(V_i^- + \frac{\sum_{j \in N_i} (w_{i,j} U_{|X|}(0; b_*) \otimes (X_j^* - X_i))}{\sum_{j \in N_i} w_j} \right), \quad (4.15)$$

де $w_{i,j}$ – скалярний ваговий множник, який визначає «силу» впливу на дану частку сусідньої частки j , й в загальному випадку змінюється в процесі ітерацій.

Як вагу $w_{i,j}$ найчастіше беруть величину, зворотну значенню фітнес-функції в точці X_j :

$$w_{i,j} = \frac{b_w}{\varphi(X_j)}, \quad (4.16)$$

де $\varphi(X_j) \neq 0$; b_w – вільний параметр, значення якого визначається класом функції $\varphi(X)$. Ще одним варіантом обчислення ваги $w_{i,j}$ є обчислення на основі відстані між частками s_i, s_j виміряній в тій чи іншій метриці.

Ранжируваний алгоритм FIPS (Ranked FIPS, RFIPS) запропонований у роботах Джорджана (J. Jordan) з метою подолання наступного недоліку алгоритму FIPS з вагами, який може мати місце при використанні в якості ваг $w_{i,j}$ величин типу (4.16). У цьому випадку на пізніх ітераціях алгоритму, коли основні поліпшення фітнес-функції вже зроблені і всі частки знаходяться в околиці мінімуму цієї функції, значення фітнес-функції виявляються для всіх часток близькими. У результаті значення ваг $w_{i,j}$ для всіх $j \in N_i$ виявляються приблизно однаковими, так що ці ваги втрачають сенс. Як наслідок, пошук сповільнюється або навіть зупиняється.

Ідея алгоритму RFIPS полягає в заміні ваг сусідніх часток $w_{i,j}$, $j \in N_i$ їх рангами $r_{i,j}$, побудованими на основі значень фітнес-функції наступним чином.

1. Для усіх часток s_j , $j \in N_i$, які є сусідами частки s_i , обчислюються значення їх фітнес-функції $\varphi(X_{i_j}) = \varphi_{i_j}$, й частки упорядковуються у вигляді списку за зростанням величини φ_{i_j} .

2. Першій частці у цьому списку присвоюється ранг $r_{i,l}=0,5$.

3. Кожній з наступних часток списку присвоюється в два рази менший ранг, так що $r_{i,j}=0,5 r_{i,j-1}$, $j \in [2 : |N_i|]$.

Легко бачити, що з ростом величини $|N_i|$ сума усіх рангів, визначених таким чином, прагне до одиниці, тобто виконується їх нормування.

В алгоритмі RFIPS формула (4.2) має вигляд

$$V_i = b_r(V_i^- + U_{|X|}(0; b_*) \otimes (X_i^* - X_i) + \sum_{j \in N_i} r_{i,j} U_{|X|}(0; b_{**}) \otimes (X_i^{**} - X_i)). \quad (4.17)$$

Слід зазначити, що алгоритм RFIPS забезпечує також зменшення негативного впливу на ефективність пошуку топології сусідства часток.

Алгоритм FDR PSO (Fitness-Distance Ratio based PSO), запропонований у роботі Вірамаханені (К. Veeramachaneni) та заснований на так званому відношенні *значення-відстань* і відрізняється від канонічного алгоритму PSO додаванням до інерційної, когнітивної і соціальної компонент формули (4.2) ще однієї компоненти виду

$$\beta_L U_{|X|}(0; b_L) \otimes (X_i^0 - X_i). \quad (4.18)$$

Отже, в алгоритмі FDR формула (4.2) приймає наступний вигляд:

$$V_i = b_r V_i^- + \beta_C U_{|X|}(0; b_C) \otimes (X_i^* - X_i) + \beta_S U_{|X|}(0; b_S) \otimes (X_i^{**} - X_i) + \beta_L U_{|X|}(0; b_L) \otimes (X_i^0 - X_i), \quad (4.19)$$

де $\beta_C, \beta_S, \beta_L$ – ще одна група вільних параметрів алгоритму, рекомендовані значення яких дорівнюють 1,0; 1,0; 2,0 відповідно; $X_i^0 = X_i^0(t)$ – вектор координат сусідньої з даною частки s_j , $j \in N_i$, якій відповідає максимальне по k значення величини

$$d(X_i - X_k) = \frac{\varphi(X_i) - \varphi(X_k)}{\|X_i - X_k\|}. \quad (4.20)$$

Тут *значенню* відповідає чисельник, а *відстані* – знаменник; $\|\cdot\|$ – деяка векторна норма (зазвичай евклідова). Іншими словами, частку s_j визначають наступним чином:

$$\max_{k \in N_i} d(X_i, X_k) = d(X_i, X_j). \quad (4.21)$$

Слід зазначити, що величина $d(X_i, X_k)$ є оцінкою константи Ліпшиця фітнес-функції в околиці точки X_i .

Доведено, що основний вклад в швидкість збіжності алгоритму FDR має параметр (4.18), оскільки алгоритм малочутливий до значень параметрів β_C, β_S , і навіть їх обнуління не призводить до суттєвого зменшення швидкості збіжності алгоритму.

4.1.3. Топології сусідства часток.

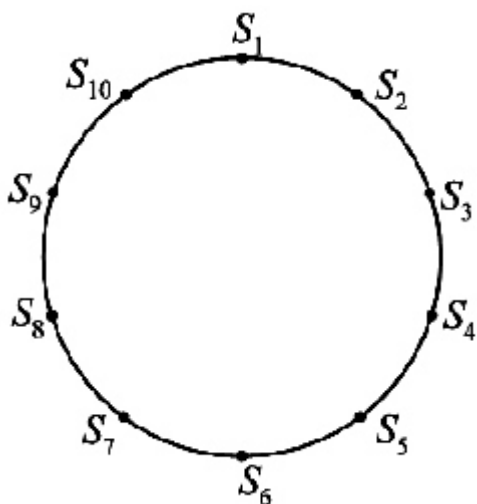
Ефективність алгоритмів PSO в значній мірі залежить від використовуваної *топології сусідства* часток (population topology, swarm topology, sociometry). Топологія сусідства визначається неорієнтованим *графом сусідства*, вершини якого відповідають часткам популяції, а ребра пов'язують безпосередніх сусідів. Сусідство часток не залежить від геометричного розташування часток в просторі пошуку, а визначається тільки їх індексами (номерама) і зберігається протягом усього пошуку (якщо не використовується та чи інша динамічна топологія сусідства).

Таке визначення сусідства часток, по-перше не вимагає великих обчислювальних витрат при визначенні сусідніх часток. У підходах, де сусідство визначають на основі, наприклад, евклідової норми різниці координат часток, на кожній ітерації потрібно $O(|S|^2)$ арифметичних операцій тільки для встановлення поточної топології сусідства. По-друге, в даному випадку сусідніми можуть вважатися частки, які містяться в далеких областях простору пошуку. Ця обставина диверсифікує пошук, попереджаючи ранню збіжність ітераційного процесу. На практиці найчастіше використовуються наступні топології сусідства часток: клік, кільце, двовимірний тор, кластерна топологія.

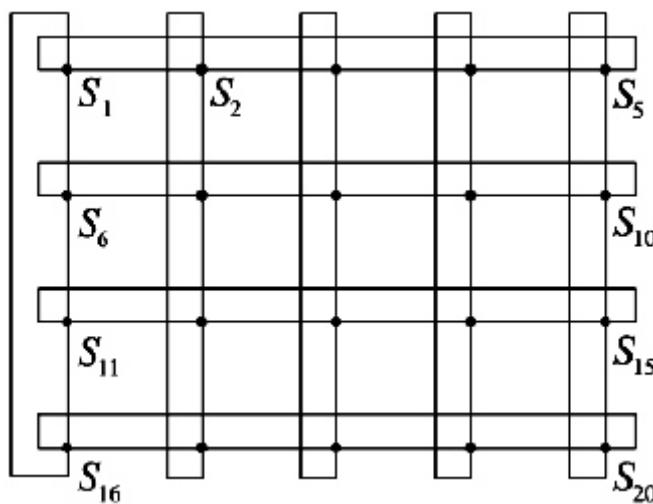
Глобально оптимальну топологію типу кліка (gbest-topology) визначає повно зв'язний граф, діаметр якого дорівнює одиниці, та у якому сусідами кожної з часток $s_i, i \in [1 : |S|]$ є інші $(|S|-1)$ частки. Діаметр неорієнтованого графа – максимальна відстань між будь-якими двома його вершинами, що дорівнює мі-

німальному числу ребер, які з'єднують ці вершини. Глобально оптимальну топологію сусідства часток називають соціальною мережею.

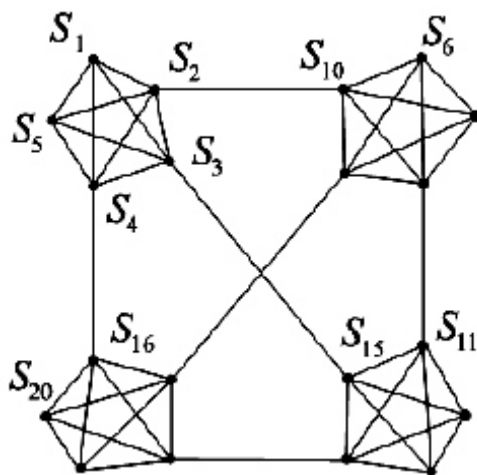
Локально оптимальній топології типу кільце (*lbest-topology*) відповідає граф сусідства, в якому сусідами кожної з часток s_i , $i \in [1:|S|]$ є дві частки. Наприклад, у графі, що наведений на рис. 4.3,а сусідами частки s_1 є частки s_2, s_{10} . Діаметр відповідного графа дорівнює $\sim |S|/2$.



а



б



в

Рис. 4.3. Топології сусідства часток: а –кільце ($|S|=10$),
б –двовимірний тор ($|S|=20$), в – кластер ($|S|=20$)

У топології двовимірний тор (топология фон Неймана) сусідами кожної з часток s_i , $i \in [1:|S|]$ є чотири частки. Наприклад, у графі, що наведений на рис. 4.3,б, який відповідає даній топології, сусідами частки s_1 є частки s_2, s_5, s_6, s_{16} . Діаметр графа типу двовимірний тор побудованого на основі $(n \times m)$ двовимірної решітки, дорівнює $\lceil n \rceil + \lceil m \rceil$, де $\lceil \cdot \rceil$ – символ найближчого цілого більшого.

У кластерній топології граф сусідства часток має в якості вершин кліки з c_{clique} вузлів в кожній. Кліки числом n_{clique} об'єднуються у повно зв'язний граф, ребра якого пов'язані з тими вузлами у кліках, ступені яких дорівнюють $(c_{clique} - 1)$. Очевидно, що діаметр такого графа дорівнює трьом. Сусідами кожної з часток s_i , $i \in [1:|S|]$ є c_{clique} часток. Наприклад, у графі на рис. 4.3,в, $n_{clique}=4$, $c_{clique}=5$, діаметр дорівнює трьом, сусідами частки s_3 є частки $s_1, s_2, s_4, s_5, s_{15}$.

Діаметр графу сусідства часток визначає швидкість поширення інформації в популяції. У популяції з топологією сусідства типу кліка краще значення фітнес-функції, досягнуте тою чи іншою часткою, відразу стає відомим усім іншим часткам популяції. З числа розглянутих топологій найнижчу швидкість поширення інформації забезпечує топологія сусідства типу кільце. Проміжні значення швидкості досягаються у популяціях, що мають топологією сусідства часток типу двовимірний тор або кластерною топологією.

Зменшення діаметра графу сусідства часток прискорює збіжність ітераційного процесу, тобто підвищує інтенсивність пошуку. З іншого боку, невеликий діаметр цього графу зменшує широту пошуку й підвищує ймовірність передчасної збіжності ітерацій до локального мінімуму фітнес-функції.

Отже, топологію сусідства часток типу кільце доцільно використовувати у разі багатовимірних багатоекстремальних цільових функцій. При оптимізації більш простих цільових функцій краще використовувати топологію типу кліка. Топології ж сусідства типу двовимірний тор та кластерна топологія з точки зору балансу інтенсивність – широта пошуку є проміжними між зазначеними то-

пологіями, а отже й універсальними, що забезпечують досить високі як інтенсивність пошуку, так і його широту.

4.1.4. Алгоритми з динамічною топологією сусідства часток.

Канонічний алгоритм PSO, а також алгоритми, що розглянуті у п. 4.1.2, засновані на статичній топології сусідства часток. Для цих алгоритмів типовою є проблема ранньої збіжності, причина якої полягає у тому, що частка, яка знайшла будь-який локальний мінімум фітнес-функції робить сильний і постійний вплив на сусідні частки, прагнучи локалізувати їх в околицях того ж локального мінімуму. В результаті зменшується ймовірність того, що ці частки можуть досліджувати інші локальні мінімуми.

Проблема ранньої збіжності особливо гостра для топологій, що забезпечують сильну зв'язаність часток (наприклад, клік) і менш гостра для слабо зв'язаних популяцій часток (кільце). Тому й виникла ідея використовувати на ранніх стадіях пошуку різні топології, тобто *динамічну топологію сусідства часток*.

Клубний алгоритм рою часток (Clubs-Based PSO, CBPSO), згідно з яким, частки популяції утворюють динамічні групи, звані клубами. Кожна частка може входити в декілька клубів. В процесі ітерацій частка може мігрувати між клубами. Сусідніми вважаються частки з одного клубу (можна вважати, що частки в клубах мають топологію сусідства типу кліка). Переміщення часток визначаються згідно формул (4.1), (4.2).

Число клубів m_i , в які входить частка s_i називають *рівнем входження* частки в клуби. У початковий момент часу $t=0$ кожна з часток популяції S випадковим чином розподіляють по $m_i=n_{cl}$ клубам, де n_{cl} – *стандартний рівень входження*. При зростанні величини n_{cl} швидкість збіжності алгоритму CBPSO зростає. Як правило, обмежуються рівнем входження від 4 до 15.

На кожній ітерації алгоритму знаходять частки s_{best} і s_{worse} з найкращим і найгіршим значеннями фітнес-функції. Частку s_{best} виключають із довільного клубу, до якого вона належить, а частку s_{worse} включають в довільний клуб, в

якому вона відсутня. Через деяке фіксоване число ітерацій для часток, які не є найкращими чи найгіршими в клубі, здійснюють випадкову зміну їх рівнів входження в клуби у напрямку стандартного рівня входження. Такий прийом покращує стабільність алгоритму і рівномірність розподілу часток по клубам.

Виключення частки s_{best} з клубу має на меті запобігти передчасній збіжності алгоритму, оскільки ця частка впливає на сусідні частки в клубі. Включення частки s_{worse} в деякий інший клуб дозволяє збільшити число сусідніх з нею часток. Оскільки частки у новому клубі, в загальному випадку, не гірше часток, що знаходяться у початковому клубі (за значеннями фітнес-функції), міграція частки s_{worse} дає їй можливість поліпшити свою пристосованість.

На параметр m_i накладається обмеження $m_i \in [m_{min} : m_{max}]$, так що кожна з часток s_i , $i \in [1 : |S|]$ в усіх випадках входить не менше ніж в m_{min} і не більше ніж в m_{max} клубів.

Алгоритм CBPSO у порівнянні з канонічним алгоритмом PSO забезпечує кращі результати з точки зору збіжності й виходу з локальних мінімумів фітнес-функції.

Алгоритм рою часток зі стереотипуванням (PSO with stereotyping) здійснює статичний розподіл часток по фіксованому числу кластерів. Критерієм розподілу є геометрична близькість часток в просторі пошуку $R^{|X|}$. Як центр кластера використовують геометричний центр координат часток, які складають цей кластер. У формулах для обчислення збільшень координат частки s_i , $i \in [1 : |S|]$, її власні координати замінюють координатами центру кластера, якому належить ця частка (так, у формулі (4.2) канонічного алгоритму PSO зазначеним чином замінюють компоненту X_i).

Алгоритм із доповненням графа сусідства часток (Dynamic Sociometry PSO, DSPSO) є найяскравішим представником алгоритмів рою часток із динамічної топологією сусідства та реалізує так звану техніку *переходу від огляду до пошуку*.

Як зазначалося вище, розріджені графи сусідства часток забезпечують високу ефективність на початкових ітераціях алгоритмів PSO, коли потрібен широкий огляд пошукового простору. Навпаки, щільні графи ефективні на завершальних ітераціях пошуку, коли необхідна швидка локалізація мінімуму фітнес-функції, в околицях якої знаходяться частки популяції.

Ідея алгоритму DSPSO полягає у наступному. На початкових ітераціях алгоритму використовується сильно розріджений граф сусідства часток. Через фіксоване число ітерацій k в цей граф додається нове ребро, наприклад, за наступним правилом.

1. За допомогою генератора натуральних випадкових чисел, рівномірно розподілених в інтервалі $[1:|S|]$ обираються два числа i_1, i_2 , таких що $i_1 \neq i_2$ і частки s_{i_1}, s_{i_2} не є сусідами в поточній топології сусідства

2. У поточний граф сусідства часток додається ребро (s_{i_1}, s_{i_2}) .

Поряд з випадковим вибором чисел може використовуватися те чи інше детерміноване правило їх вибору. В результаті розглянутого процесу при достатній кількості ітерацій отримують граф сусідства типу кліка.

4.2. Мурашина оптимізація

Мурашина оптимізація була запропонована Доріго (M. Dorigo), Ман'єццо (V. Maniezzo) і Колорні (A. Colomi) у 1992 р. як алгоритм розв'язання задачі комівояжера (Traveling Salesman Problem, TSP). Ця задача передбачає вибір найкоротшого шляху в графі, так що є пряма аналогія із задачею, яку щодня вирішують мурахи, оптимізуючи свої шляхи до їжі. Наразі відома велика кількість дискретних і безперервних алгоритмів мурашиної оптимізації, які отримали об'єднуюче найменування *алгоритми оптимізації мурашиних колоній* (Ant Colony Optimization, ACO).

Першим з алгоритмів ACO, призначеним для вирішення задачі комівояжера, став алгоритм мурашиної системи (Ant System, AS). Потім з'явилося кіль-

ка модифікацій цього алгоритму: щільнісний (ant-density), кількісний (ant-quality), циклічний (ant-cycle). Експерименти показали значну перевагу останнього алгоритму над двома іншими, які й були відкинуті, і наразі канонічним алгоритмом АСО для вирішення задачі комівояжера вважають циклічний алгоритм.

У подальшому було запропоновано ряд більш розвинених алгоритмів АСО для вирішення задачі комівояжера. Успіхи у вирішенні цієї задачі стимулювали розробку АСО-алгоритмів для вирішення інших задач дискретного програмування (задача про призначення, задача календарного планування, транспортна задача тощо). Виявилось, що алгоритми АСО показують хороші результати при вирішенні широкого класу задач, які можуть бути зведені до задачі пошуку на графах.

Першим варіантом алгоритму АСО, розробленим для вирішення задачі безперервної оптимізації став *алгоритм безперервної оптимізації колонією мурах* (Continuous Ant Colony Optimization, САСО), запропонований у 1995 році Білчевим (G. Bilchev) і Пармі (I. C. Parmee). Наразі відомо багато модифікацій алгоритму САСО, наприклад *алгоритм безперервно взаємодіючої мурашиної колонії* (Continuous Interaction Ant Colony, СІАС), *розширений алгоритм оптимізації колонії мурах для безперервних областей* (extended Ant Colony Optimization to continuous domains, АСО_R), *алгоритм безперервної системи мурашиної колонії* (Continuous Ant Colony System, САСС), *алгоритм бінарної мурашиної системи* (Binary Ant System, САС), *алгоритм прямого призначення* (Direct Application of Ant Colony Optimization, DАСО).

До недоліків АСО-алгоритмів пошукової оптимізації можна віднести ускладненість їх теоретичного аналізу і наявність великого числа вільних параметрів, оскільки, з одного боку, від значень цих параметрів у багатьох випадках сильно залежить ефективність пошуку, з іншого боку відсутні правила апріорного вибору їх оптимальних значень.

4.2.1. Біонічні передумови.

Мурахи відносяться до соціальних комах, які живуть всередині деякого колективу – колонії. Число мурах в одній колонії досягає декількох мільйонів. Поведінка мурах при транспортуванні їжі, подоланні перешкод, будівництві мурашника і інших діях часто наближається до теоретично оптимальної. «Соціальна» поведінка мурах будується на основі самоорганізації, тобто на основі множини механізмів, що забезпечують досягнення системою глобальної мети в результаті низькорівневої взаємодії її елементів. Принциповою особливістю такої взаємодії є використання елементами системи тільки локальної інформації. Будь-яке централізоване керування у цьому випадку відсутнє.

Мурахи використовують два способи передачі інформації. Прямий спосіб реалізується через *канал індивідуальних взаємодій* або *прямий канал* між парою мурах за допомогою обміну їжею, щелепних маніпуляцій, візуального та хімічного контактів.

Непрямому способу обміну інформацією між мурахами відповідає *канал стігметрії* або *слідовий канал*. Взагалі даний канал є рознесеним в часі типом взаємодії (стігметрія), коли один суб'єкт взаємодії змінює деяку частину навколишнього середовища, а інші використовують інформацію про стан цієї частини середовища пізніше, коли знаходяться в її околиці. У мурах стігметрія реалізується за допомогою *феромона* – спеціального секрету, який відкладається у вигляді сліду при переміщенні мурахи і який може сприйматися мурахами протягом деякого часу. Чим вище концентрація феромону на стежці, тим більше мурах буде по ній рухатися. Згодом феромон випаровується, що дозволяє мурахам адаптувати свою поведінку до змін зовнішнього середовища.

Суспільна поведінка мурах в колонії підпорядкована ідеї *гетерархії*, яка на відміну від ієрархії передбачає не вертикальні, а горизонтальні зв'язки між агентами системи, при цьому самі агенти рівноправні між собою. У колонії мурах ми маємо справу з щільною гетерархією, коли кожна з мурах може зв'язатися з будь-якою іншою в будь-який момент часу. При цьому інформація передається через два розглянутих вище канали зв'язку – прямий і стігметрії. Слід за-

значити, що система, побудована таким чином має *емерджентні властивості*, тобто властивості, які не притаманні складовим її агентам, що розглядаються поза системою, але притаманні системі в цілому.

Канал обміну інформацією між елементами гетерархічної системи визначає такі її характеристики: область дії каналу – розмір групи осіб популяції, в межах якої можливий обмін інформацією; обсяг і час зберігання інформації в каналі; спосіб зміни інформації в каналі – які особи або зовнішні по відношенню до системи фактори і за якими правилами можуть змінювати інформацію. Різні комбінації зазначених характеристик каналу породжують широку множину їх варіантів.

За наведеною класифікацією область дії каналу стігметрії в мурашиній колонії охоплює всю колонію, оскільки феромонна мітка може бути залишена кожною з мурах популяції і принципово доступна будь-якій іншій мурасі. Кількість феромону поступово зменшується, так що ця мітка рано чи пізно зникає. Тобто, час зберігання інформації в каналі дорівнює часу активності феромонного сліду.

Інформація по прямому каналу також може бути передана кожною мурахою популяції будь-якій іншій мурасі, але, на відміну від каналу стігметрії, у кожний даний момент часу – тільки одній мурасі. Кожною мурахою повідомлення може бути отримано від декількох інших мурах.

4.2.2. Алгоритм безперервної оптимізації колонією мурах.

Основними особливостями алгоритму CASO є комбінування глобального і локального пошуку, використання для глобального пошуку ГА, а для локального – тільки каналу стігметрії.

Розглянемо задачу багатовимірної глобальної умовної максимізації в гіперпаралелепіпеді Π . В кожній точці $X_i(\tau)$, в якій побувала мураха, $i \in [1 : |S|]$, вона залишає *феромонну точку* $\phi_i(t, \tau)$, інтенсивність якої (кількість феромону) на даній ітерації t дорівнює $\theta_i(t, \tau)$, де $\tau \in [0 : t]$ – номер ітерації, на якій зали-

шена ця точка. Кількість феромону у феромонних точках з ростом числа ітерацій зменшується (феромон випаровується) відповідно до рекурентної формули:

$$\theta_i(t, \tau) = \begin{cases} b_\theta \theta_i(t-1, \tau), & \theta_i(t-1, \tau) \geq \theta_{\min}, \\ 0, & \text{інакше,} \end{cases} \quad (4.22)$$

де b_θ – коефіцієнт стійкості феромонних точок. Формула (4.22) означає, що феромонна точка зникає, якщо кількість феромону в ній стає меншим за θ_{\min} .

Сукупність феромонних точок

$$\Phi_i(t) = \{\phi_i(t, \tau) \mid \tau \in [0; t], \theta_i(t, \tau) > \theta_{\min}\}, \quad i \in [1 : |S|] \quad (4.23)$$

утворює феромонний слід мурахи s_i . Число феромонних точок в цьому сліді позначимо $|\Phi_i(t)|$.

На поточній ітерації t активними є феромонні точки, що належать усім феромонним слідам $\Phi_i(t)$, $i \in [1 : |S|]$. Сукупність цих точок позначимо

$\Phi(t) = \{\Phi_i(t), i \in [1 : |S|]\}$. Тоді $|\Phi(t)| = \sum_{i=1}^{|S|} |\Phi_i(t)|$ – загальне число активних феромонних точок на поточній ітерації t .

Введемо ще позначення $S^g = \{s_i^g, i \in [1 : |S^g|]\}$, $S^l = \{s_i^l, i \in [(|S^g| + 1) : |S|]\}$ – множини мурах, призначених для глобального і локального пошуку, відповідно; $|S^g| + |S^l| = |S|$.

У наведених позначеннях схема алгоритму CASO має такий вигляд.

Етап ініціалізації. Створення $|S|$ регіонів пошуку. Ініціалізація в кожному з регіонів $X_i(0)$ феромонної точки. Встановлення лічильника числа ітерацій $t=1$.

Етап оптимізації.

1. Обчислення в кожному з регіонів пошуку значення фітнес-функції $\varphi(X_i) = \varphi_i$.

2. Сортування регіонів в порядку зростання величини φ_i і представлення результатів сортування у вигляді лінійного списку.

3. Виконання генетичної операції мутації (з ймовірністю мутації ξ_m) для перших $0,9|S^g|$ (найменш пристосованих) мурах зазначеного списку.

4. Аналогічно виконання генетичної операції кросовер для наступних $0,1 | S^g |$ мурах того ж списку.

5. Ініціалізація в кожному з $| S^g |$ регіонів пошуку, отриманих в результаті попередніх операцій, феромонної точки.

6. Виконання локального пошуку кращого регіону на основі каналу стігметрії для кожного з $| S^l |$ мурах:

- якщо регіон з кращим значенням фітнес-функції знайдено, ініціалізація відповідної феромонної точки та переміщення однієї з мурах множини $| S^l |$ в цей регіон;

- в іншому випадку випадковим чином проводиться спроба знаходження нового кращого регіону.

7. Моделювання випаровування феромону за формулою (4.22).

Етап завершення обчислень. Перевірка умови закінчення ітерацій. Якщо вона виконана, обчислення завершуються, інакше – повернення до етапу оптимізації.

Кожен регіон пошуку $X_i(0)$, що створений на етапі ініціалізації, є випадковою точкою в області Π , координати якої визначаються за правилом

$$x_{i,j}(0) = U_1(x_i^-; x_i^+), \quad i \in [1 : | S |], \quad j \in [1 : | X |], \quad (4.24)$$

де U_1 – функція одномірного рівномірного розподілу. Регіону $X_i(0)$ ставиться у відповідність феромонна точка $\phi_i(0,0)$ інтенсивністю $\theta_i(0,0) = \theta_{init}$, де θ_{init} – початкова кількість феромону.

Для мутації використовується гаусовий мутатор

$$x'_{i,j} = x_{i,j} + N_1(0, \sigma), \quad j \in [1 : | X |], \quad (4.25)$$

де N_1 – функція одномірного нормального розподілу (Гауса), а середньоквадратичне відхилення σ – *крок мутації*. З ростом числа ітерацій ця величина зменшується (редукується) за законом

$$\sigma(t) = \sigma_{\max} \left(1 - (U_1(0;1))^{(1-t/\bar{t})^{b_m}} \right), \quad (4.26)$$

де σ_{max} – максимальна величина кроку мутації, b_m – натуральний параметр нелінійності. Нелінійна редукція кроку мутації дозволяє на початкових ітераціях пошуку забезпечити широту пошуку, а на завершальних ітераціях – точність локалізації максимуму.

Як оператор кросовера використовується арифметичний кросовер

$$x'_{i,k} = ux_{i,k} + (1-u)x_{j,k}, \quad (4.27)$$

де $i, j \in [1 : |S^g|]$, $k \in [1 : |X|]$, $i \neq j$, $u = U_1(0;1)$. Батьківські особи s_i, s_j для схрещування обираються з числа $0,1|S^g|$ мурах множини S^g з однаковою ймовірністю.

Локальний пошук кращого регіону виконується наступним чином.

1. Вибір випадковим чином мурахи s_i з числа останніх $|S^l|$ мурах зазначеного вище лінійного списку.

2. З урахуванням *інтересу* даної мурахи обчислення координат центру ваги поточних активних феромонних точок

$$W(t) = \sum_{j,k} \bar{\omega}_{i,j,k}(t) X_{j,k}, \quad j \in [1 : |S|], \quad k \in [1 : |\Phi_j(t)|], \quad (4.28)$$

де $X_{j,k}$ – k -та точка феромонного сліду $\Phi_j(t)$. Нормований інтерес i -ої мурахи до k -ої точки феромонного сліду $\Phi_j(t)$ обчислюється таким чином

$$\bar{\omega}_{i,j,k}(t) = \frac{\omega_{i,j,k}(t)}{\sum_{l=1}^{|S|} \sum_{m=1}^{|\Phi_l(t)|} \omega_{i,l,m}(t)}. \quad (4.29)$$

3. Обчислення координат нового можливого положення мурахи s_i :

$$X'_i = X_i + V_i. \quad (4.30)$$

Тут V_i – крок у напрямку $(X_i - W_i(t))$, величина якого дорівнює $abs(b_{beg} - b_{inc} a_i)$, де b_{beg}, b_{inc} – позитивні визначені користувачем константи, що мають сенс початкової величини кроку і його інкременту, відповідно; a_i – *слабкість* або *вік* регіону X_i .

Початкове значення величина a_i отримує при створенні відповідного регіону пошуку. Якщо значення фітнес-функції в регіоні X'_i вище значення тієї ж

функції в регіоні X_i , то значення величини a_i збільшують, інакше – зменшують. Даний механізм дозволяє зменшувати крок локального пошуку з наближенням мурахи s_i до положення максимуму фітнес-функції.

Інтерес мурахи s_i до k -ої точки феромонного сліду $\Phi_j(t)$ визначається таким чином

$$\varpi_{i,j,k}(t) = \frac{\bar{\rho}(t)}{2} \theta_{j,k}(t) \exp(-\rho_{i,j,k}(t)), \quad i, j \in [1:|S|], \quad k \in [1:|\Phi_j(t)|] \quad (4.31)$$

де $\bar{\rho}(t)$ – середня поточна відстань між двома мураками популяції, $\theta_{j,k}(t)$ – кількість феромону у зазначеній феромонній точці, $\rho_{i,j,k}(t) = \|X_i(t) - X_{j,k}\|_E$ – евклідова відстань між регіоном і тією ж феромонною точкою.

Прийmemo, що на даній ітерації t активними є всі регіони, в яких побувала кожна з мурах популяції. Тоді з формули (4.28) випливає, що для мурахи s_i , $i \in [1:|S|]$ координати центру ваги всіх активних феромонних точок можна обчислити за формулою

$$W(t) = \sum_{j,k} \bar{\varpi}_{i,j,k}(t) X_{j,k}, \quad j \in [1:|S|], \quad k \in [1:t]. \quad (4.32)$$

Як умови закінчення ітерацій використовують типові для популяційних алгоритмів оптимізації умови.

Алгоритм CASO містить наступні основні вільні параметри:

θ_{init} – початкова кількість феромону (рекомендоване значення $\theta_{init}=1,0$);

θ_{min} – кількість феромону, нижче якого феромонна точка стає неактивною (рекомендоване значення $\theta_{min}=10^{-8}$);

$b_\theta \in (0;\infty)$ – коефіцієнт стійкості феромонних точок (рекомендоване значення $b_\theta=0,1$);

$\xi_m, \sigma_m, \sigma_{max}, b_m$ – ймовірність мутації, крок мутації, максимальна величина кроку мутації та параметр нелінійності мутації відповідно;

b_{beg}, b_{inc} – початкова величина кроку локального пошуку і його інкремент відповідно;

$a_i(0)$, $i \in [1 : |S|]$ – початкове значення віку регіонів.

За загальним властивості популяційних алгоритмів пошукової оптимізації зі збільшенням розміру популяції алгоритм САСО підвищує ймовірність локалізації глобального максимуму фітнес-функції (ціною підвищення обчислювальної складності кожної ітерації). Для функцій, що мають просту топологію, хороші результати може показати популяція з усього 10 мурах. Для складних багатовимірних багатоекстремальних функцій число мурах має бути близьким до 100.

Рекомендоване використання числа осіб у популяції дорівнює

$$|S| = |S|_{\max} (1 - \exp(-0,1 |X|)) + 5, \quad (4.33)$$

де $|S|_{\max} = 1000$ – максимально допустима кількість мурах у популяції.

У термінах алгоритму рою часток зменшення кількості феромону θ_{\min} і збільшення коефіцієнта стійкості феромонних точок b_θ підвищує вплив соціальної компоненти на еволюцію мурах. У цьому випадку феромонні сліди $\Phi_j(t)$, $j \in [1 : |S|]$ «подовжуються» (зростають величини $|\Phi_j(t)|$) і вплив інших мурах на кожну дану мураха збільшується. В результаті підвищується ймовірність «застрявання» мурах в локальних максимумах цільової функції.

4.2.3. Алгоритм безперервно взаємодіючої колонії мурах.

При вирішенні задач пошукової оптимізації реалізація ідеї стігметрії є не єдиним способом використання мурашиного інтелекту. Природною є ідея використання для обміну інформацією між мурахами як каналу стігметрії, та і каналу прямого обміну. Алгоритм, що використовує обидва зазначених канали – це алгоритм СІАС, який був запропонований в роботах Дрео (J. Dreco) і Сіаррі (P. Siarry).

Розглянемо задачу багатовимірної глобальної умовної максимізації в гіперпаралелепіпеді Π . Використовуємо позначення і допущення попереднього пункту. Крім того, припустимо, що кожна мураха s_i має пам'ять, в яку інші му-

рахи можуть надсилати *повідомлення*. Повідомлення $m_{i,j}$, отримане мурахою s_i від мурахи s_j , $i, j \in [1 : |S|]$, $i \neq j$ містить наступну інформацію: поточне положення мурахи X_j , значення фітнес-функції в цій точці φ_j . Таким чином, в алгоритмі СІАС мураха s_i визначає набір

$$\langle X_i, \varphi_i, m_{i,j_1}, m_{i,j_2}, \dots \rangle, j_1, j_2, \dots \in [1 : |S|], j_1 \neq j_2 \neq \dots \neq i. \quad (4.34)$$

Схема алгоритму СІАС має такі етапи.

1. Створення початкової популяції мурах $S = \{s_i, i \in [1 : |S|]\}$ і ініціалізація феромонних точок за схемою алгоритму САСО. Ініціалізація пам'яті кожної з мурах s_i шляхом розміщення в ній координат векторів $X_i(0)$ і значень фітнес-функції $\varphi_i(0)$. Установка лічильника числа ітерацій $t=1$.

2. Моделювання роботи каналу стігметрії за схемою, близькою до схеми алгоритму САСО. Відмінність полягає лише у тому, що переміщення мурахи s_i у напрямку центра тяжіння $W_i(t)$ проводиться на випадкову відстань в межах *розмірного параметра* мурахи

$$r_i = \frac{N_1(0,5; \sigma_r)}{b_r}, i \in [1 : |S|], \quad (4.35)$$

де b_r, σ_r – *дільник розмірного параметра* і середньоквадратичне відхилення відповідно.

3. Моделювання функціонування каналу прямого впливу між мурахами:

а) вибір випадковим чином одного з повідомлень із пам'яті мурахи $s_i \in S$, наприклад, повідомлення від мурахи $s_j \in S$, $i, j \in [1 : |S|]$, $j \neq i$;

б) порівняння значень фітнес-функцій φ_i, φ_j . Якщо $\varphi_j > \varphi_i$, то переміщення мурахи s_i у випадкову точку гіперсфери з центром у точці X_j і радіусом, що дорівнює величині розмірного параметра r_i цієї мурахи;

в) в іншому випадку виконання наступних дій: відправлення випадковій мураці $s_k \in S$, $k \in [1 : |S|]$, $k \neq i$ повідомлення $m_{k,i}$, що містить координати точки X_i і значення фітнес-функції φ_i ; переміщення випадковим чином мурахи s_i в межах гіпершара радіуса r_i з центром у точці X_i .

4. Моделювання випаровування феромонних точок за формулою (4.22).

5. Перевірка виконання умови закінчення ітерацій за схемою алгоритму CASO. Якщо ця умова виконана – завершення обчислень, інакше – установка $t=t+1$ і перехід до кроку 2.

За одну ітерацію кожна з мурах може отримати більше одного повідомлення. Число повідомлень, які можуть поміститись в пам'яті мурахи, обмежена натуральною величиною m_{max} . Якщо мурасі надходить нове повідомлення, а її пам'ять вже вичерпана, то це повідомлення ігнорується.

Крім вільних параметрів, які аналогічні параметрам алгоритму CASO, алгоритм CIAC містить наступні параметри:

b_r – дільник розмірного параметра;

$\sigma_r \in [0;1]$ – середньоквадратичне відхилення величини розмірного параметра при $b_r=1,0$ (рекомендоване значення дорівнює $\sigma_r=0,9$);

$m_{max} \in [0; \infty)$ – максимально допустима кількість повідомлень в пам'яті мурахи (рекомендоване значення $m_{max}=2/3|S|$).

Слід зазначити, що має місце сильна залежність основних характеристик алгоритму CIAC від значень параметрів b_r, σ_r , що визначають величину розмірного параметра $r_i, i \in [1 : |S|]$. Збільшення значень дільника розмірного параметра b_r , очевидно, зменшує величину r_i і, тим самим, широту пошуку. Підвищення значення середньоквадратичного відхилення σ_r збільшує дисперсію розмірного параметра і диверсифікує пошук. На початкових стадіях пошуку рекомендують використовувати малі значення величини b_r і великі значення величини σ_r . На завершальних стадіях величину b_r доцільно збільшувати, а σ_r – зменшувати.

На характеристики алгоритму CIAC істотний вплив може чинити також величина m_{max} . Збільшення її значень диверсифікує пошук, але знижує вплив соціальної компоненти (за яку відповідає канал стігметрії).

Встановлено, що алгоритм CIAC, який використовує тільки прямий канал, з високою ймовірністю завершується у локальному максимумі фітнес-

функції. Навпаки, канал стігметрії дозволяє цього уникнути, але істотно збільшує обчислювальні витрати. Комбінація обох каналів може забезпечити високу ймовірність локалізації глобального максимуму з прийнятними обчислювальними витратами.

Для підвищення ймовірності локалізації глобального екстремуму складних багатоекстремальних функцій може використовуватись наступна модифікація алгоритму СІАС на основі *струсу популяції*.

1. Виконання t_c ітерацій алгоритму СІАС.
2. Відбір n_w мурах з гіршими значеннями фітнес-функції і виключення їх із популяції.
3. Замість вилучених мурах створення в просторі пошуку випадковим чином n_w нових мурах.
4. Якщо умова закінчення ітерацій виконана, завершення обчислень, інакше – перехід до кроку 1.

Процедура струсу популяції має вільні параметри t_c , n_w . Крім того, виникає задача вибору правила, за яким слід призначати кількість феромону в початкових точках розташування вказаних n_w мурах.

Як величину t_c доцільно використовувати деяку монотонно спадну функцію номеру ітерації, наприклад, функцію вигляду

$$t_c = t_c^0 + \alpha t, \quad (4.36)$$

де t_c^0 – початкове значення величини t_c , $\alpha \in (0;1)$ – заданий коефіцієнт.

Для обчислення значень феромону в нових точках можна використовувати, наприклад, формулу виду

$$\theta_i(t+1) = \nu \theta_i(t) \frac{\varphi(X_i(t))}{\varphi(X_i(t+1))}, \quad (4.37)$$

де $\theta_i(t)$, $\theta_i(t+1)$ – значення феромона, які відповідають положенням мурахи s_i , до і після її переміщення; $\varphi(X_i(t))$, $\varphi(X_i(t+1))$ – аналогічні значення фітнес-функції; $\nu > 1$ – вільний параметр.

4.2.4. Безперервний ортогональний алгоритм мурашиної колонії.

Безперервний ортогональний алгоритм мурашиної колонії (Continuous Orthogonal Ant Colony, COAC) запропонований у роботах Ху (X. Hu), Чжана (J. Zhang) і Лі (Y. Li). Алгоритм істотно використовує так звані *ортогональні плани експерименту*, розроблені в рамках теорії планування експериментів.

Загальна схема алгоритму виглядає наступним чином. Після вибору мурахою регіону пошуку вона обстежує цей регіон на основі ортогонального плану експерименту, що дозволяє раціоналізувати і прискорити пошук. У кращих точках регіону пошуку (відповідних максимальним значенням фітнес-функції) мураха залишає феромонні мітки. Аналогічно алгоритму SACO, кожна мураха може сприйняти кількість феромону у цих мітках і на цій основі вибрати новий регіон пошуку. Кількість феромону у точках, що належать регіонам, в яких пошук вже закінчено, зменшується. Регіони, кількість феромону в яких нижче певного рівня, виключаються із подальшого розгляду. Тобто алгоритм використовує канал стігметрії, та не використовує прямий канал.

Варійовані параметри в теорії планування експерименту називають *факторами*, а досліджувані значення цих параметрів – *рівні факторів*. Функція, яка оптимізується (аргументами якої є фактори), є *функцією відгуку*. Експеримент, в якому кожен з факторів приймає кожен зі своїх можливих рівнів, є *повним факторним експериментом* (ПФЕ).

Для прикладу розглянемо функцію відгуку $f(X)$, яка залежить від $|X|$ чинників $(x_1, x_2, \dots, x_{|X|})$. Припустимо, що експеримент проводиться в точці $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,|X|})$ і число рівнів кожного з факторів дорівнює $(2l+1)$. Тоді ПФЕ в зазначеній точці вимагає обчислення значень функції $f(X)$ в точках $x_{i,j} \pm k\delta_{i,j}$ $j \in [1 : |X|]$, $k \in [0 : l]$, де $\delta_{i,j}$ – крок зміни j -го фактора (рис. 4.4). Кожне обчислення значення функції відгуку називають *дослідом*. Отже, ПФЕ вимагає $n = (2l+1)^{|X|}$ дослідів.

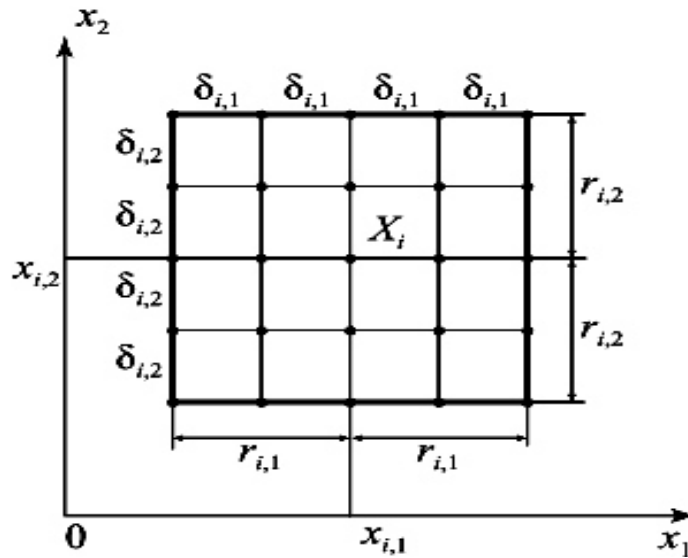


Рис. 4.4. Схема повного факторного експерименту ($|X|=2, l=2, n=25$)

За допомогою відповідного лінійного перетворення факторів X досягають того, щоб трансформований вектор \tilde{X} в точці X_i мав нульові значення всіх своїх компонент, а в інших точках ПФЕ приймав цілочисельні значення, рівні $(\pm k)$, $k \in [0:l]$.

Значення, які кожен з факторів приймає у всіх n дослідах утворюють $(n \times |X|)$ -матрицю плану експерименту P . Наприклад, для ПФЕ, в якому $|X|=2$ і $l=1$, ця матриця має вигляд, представлений в табл. 4.1.

Приклад матриці ПФЕ при $|X|=2$ і $l=1$

Таблиця 4.1

Номер дослід	Фактор		Номер дослід	Фактор	
	\tilde{x}_1	\tilde{x}_2		\tilde{x}_1	\tilde{x}_2
1	-1	-1	6	1	0
2	0	-1	7	-1	1
3	1	-1	8	0	1
4	-1	0	9	1	1
5	0	0			

Стовпець матриці P , відповідний j -му фактору позначено P_j , $j \in [1:|X|]$.

Матриця плану експерименту $P \in \text{ортогональною}$, якщо для її двох будь-яких незбіжних стовпців P_{j_1}, P_{j_2} має місце рівність $(P_{j_1}, P_{j_2}) = 0$, де $(,)$ – символ скалярного добутку векторів. Легко бачити, що матриця плану експерименту, представлена в табл. 4.1 є ортогональною.

При високій розмірності простору пошуку ПФЕ вимагає дуже великого числа дослідів (навіть при малому числі факторних рівнів). Тому в обчислювальній практиці зазвичай використовують так званий *дробовий факторний експеримент* (ДФЕ), суть якого полягає у виключенні з матриці P деяких рядків таким чином, щоб отримана матриця була ортогональною і число рядків в ній дорівнювало або перевищувало число факторів.

Поточну множину регіонів пошуку позначимо $A=(A_i, i \in [1:|A|])$, де регіон пошуку A_i визначає набір

$$\langle X_i, R_i, \Phi_i, \Theta_i, X_i^*, \theta_i^*, \varphi_i^*, \rho_i^*, n_i^* \rangle, \quad (4.38)$$

де X_i – вектор координат центру регіону; $R_i=(r_{i,1}, r_{i,2}, \dots, r_{i,|X|})$ – вектор радіусів регіону (позитивні величини); $\Phi_i=(\Phi_{i,1}, \Phi_{i,2}, \dots, \Phi_{i,|X|})$ – множина векторів активних феромонних точок Φ_i , що належать даному регіону; $\Theta_i=(\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,|X|})$ – множина відповідних кількостей феромона; $X_i^* \in \Phi_i$ – вектор координат найбільш активної феромонної точки регіону; θ_i^* – відповідна кількість феромона; φ_i^* – найкраще значення фітнес-функції активних феромонних точок регіону; ρ_i^* – ранг регіону; n_i^* – загальне число відвідувань мурахами даного регіону. Таким чином, регіону A_i відповідає гіперпаралелепіпед (рис. 4.5)

$$\Pi_i = \{(x_{i,j} - r_{i,j}) \leq x_{i,j} \leq (x_{i,j} + r_{i,j}), j \in [1:|X|]\}. \quad (4.39)$$

Початкові положення центрів регіонів пошуку $X(0)=(X_i(0), i \in [1:|A|])$ є випадковими, рівномірно розподіленими в області пошуку D . Початкові радіуси регіонів $r_{i,1}(0), r_{i,2}(0), \dots, r_{i,|X|}(0)$ є випадковими величинами, що рівномірно розподілені в інтервалі $[-r^0; r^0]$, де $r^0 > 0$ – вільний параметр алгоритму. Іншими словами $r_{i,j}(0)=U(-r^0; r^0)$.

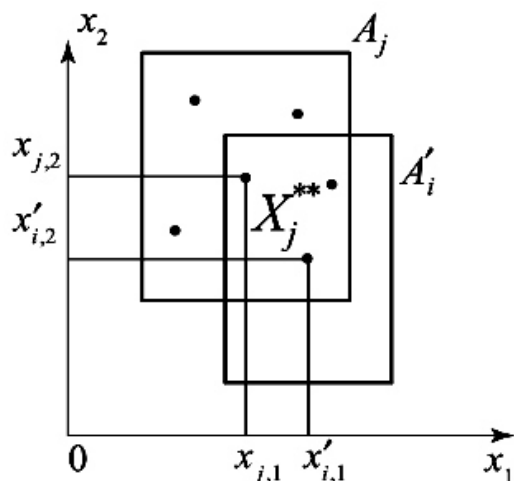


Рис. 4.5. До схеми формування нових регіонів пошуку: $|X|=2$

На поточній ітерації для мурахи s_i , $i \in [1 : S]$ новий регіон A'_i формуємо за таким правилом.

1. Генерується випадкове число $u=U_i(0;1)$. Якщо $u \leq b_A$, то обирається ще не зайнятий регіон A_j , якому належить точка X_j^{**} , що відповідає найбільш інтенсивній феромонній мітці:

$$\max_k \theta_k^* = \theta_j^{**}, k \in [1 : A]. \quad (4.40)$$

В іншому випадку обирається регіон за правилом рулетки з ймовірністю:

$$\xi_i = \frac{\theta_i^*}{\sum_k \theta_k^*}, i \in [1 : S]. \quad (4.41)$$

2. Будується новий регіон A'_i , використовуючи в якості його центру точку X_j^{**} (див. рис. 4.5). Адаптивні радіуси регіону r_{ij} , $j \in [1 : X]$ визначаються за схемою, розглянутою нижче.

Тут b_A – вільний параметр алгоритму, який має сенс ймовірності вибору мурахою s_i регіону A_i .

Після формування регіону пошуку A'_i будується $(n \times |X|)$ -матриця плану ПФЕ, відповідного даному регіону і випадковим чином вибираються $v \geq |X|$ рядків, формуючи таким чином $(v \times |X|)$ -матрицю P' плану ДФЕ. Матриця визначає

точки регіону A'_i , які підлягають обстеженню. Число рівнів факторів зазвичай приймають рівним трьом (коли $l=1$), так що $n=3^{|X|}$.

Далі відбувається послідовне переміщення мурахи у всі точки, які визначає матриця P' , і в кожній з них обчислюється значення фітнес-функції. Якщо найкраща точка регіону не змінюється протягом фіксованого числа ітерацій, то в якості центру нового регіону використовують іншу випадково обрану точку поточного регіону.

Якщо в результаті ортогонального обстеження регіону пошуку A'_i , значення фітнес-функції вдалося поліпшити, то радіуси регіону збільшують з тим, щоб урізноманітнити пошук у «вдалому» місці. Інакше радіуси регіону зменшують, щоб детальніше дослідити околиці знайденої кращої точки. Зменшення радіусів здійснюється за формулою

$$r'_{i,j} = \begin{cases} b_r r_{i,j}, & r_{i,j} \geq b_r^{\max}, \\ U_1(b_r^{\max}; 2r^0), & \text{інакше.} \end{cases} \quad (4.42)$$

а збільшення – за формулою

$$r'_{i,j} = \frac{r_{i,j}}{b_r}; \quad i \in [1 : |S|], \quad j \in [1 : |X|]. \quad (4.43)$$

Тут b_r, b_r^{\max} – вільні параметри алгоритму.

При збіжності алгоритму до глобального рішення зі збільшенням числа ітерацій частота зменшення радіуса зростає, а частота його збільшення – зменшується. В результаті в кінці кінців пошук зосереджується в одному або декількох регіонах з потенційно кращими точками.

Алгоритм СОАС крім того використовує так звану *процедуру модуляції*, яка полягає в наступному.

1. Поточні регіони A_i , $i \in [1 : |A|]$ впорядковують у вигляді списку в порядку убутання їх пристосованості. Ранг регіону ρ_i^* вважають рівним його номеру в зазначеному списку.

2. Перші $b_\rho |A|$ регіонів списку переносять у множину регіонів A' , яка формується, де $b_\rho \in [0;1]$ – *параметр елітарності*.

3. Змінюють кількість феромону в активних точках регіону A'_i , $i \in [1 : b_\rho | A |]$ за формулою

$$\theta'_i = (1 - b_\theta)\theta_i + b_\theta\theta_{init}(b_\rho | A | + 1 - \rho_i^* + n_i^*). \quad (4.44)$$

4. Регіони, які залишились в множині A' не визначені $((1 - b_\rho)|A|$ регіонів) замінюють новими випадковими регіонами за схемою їх початкової ініціалізації. Кількість феромону у точках, які опинилися в цих регіонах, вважають рівними θ_{init} .

Алгоритм СОАС на відміну від алгоритму САСО містить додаткові вільні параметри:

r^0 – величина, що визначає початкові ресурси регіонів;

$b_{A \in (0;1]}$ – поріг вибору нового регіону пошуку (рекомендоване значення $b_{A \in [0,3;0,5]}$);

$v \geq |X|$ – число рядків у використовуваному плані ДФЕ;

b_r, b_r^{max} – параметри, що визначають закон зменшення радіусів регіону (рекомендоване значення $b_r \in [0,2;0,9]$);

b_ρ – параметр елітарності (рекомендоване значення $b_\rho = 0,1$).

Залежно від складності фітнес-функції в алгоритмі СОАС рекомендують використовувати популяції з числом мурах, рівним $S \in [20:100]$, і відношення числа регіонів $|A|$ до числа мурах $|S|$ – в діапазоні від одиниці до двох.

Наразі відомі результати дослідження ефективності алгоритму СОАС, які показують, що у порівнянні із алгоритмом САСО даний алгоритм забезпечує у більшості випадків значні переваги по точності локалізації глобального максимуму і швидкості збіжності.

4.3. Оптимізація бджолиним роєм

Алгоритми бджолиної оптимізації є одними з новітніх багатоагентних алгоритмів, заснованих на моделюванні *ройового інтелекту* (swarm intelligence). Перші статті, в яких викладається алгоритм даного класу, були опубліковані у

2005 р. Основна ідея таких алгоритмів полягає у моделюванні поведінки бджіл при пошуку нектару.

Алгоритми бджолоїної оптимізації успішно застосовують при вирішенні різних задач дискретної оптимізації – задачі календарного планування, задачі комівояжера, транспортної задача тощо. Розроблено також високоефективні варіанти алгоритму, призначені для вирішення задачі безперервної оптимізації, найбільш відомими з яких є *бджолоїний алгоритм* (Bees Algorithm) і *алгоритм колонії штучних бджіл* (Artificial Bee Colony, ABC).

До недоліків бджолоїних алгоритмів слід віднести велике число вільних параметрів.

4.3.1. Біонічні передумови.

Радіус області поширення рою медоносних бджіл досягає 10 і більше кілометрів, що забезпечує високу ймовірність виявлення великої кількості джерел нектару. «Стратегію» рою можна визначити наступним чином: безперервний пошук в зазначеній області квіткових ділянок, які містять достатню для видобутку кількість нектару; відвідування квіткових ділянок, багатих на нектар та пилок, що можуть бути зібрані з меншими зусиллями великим числом бджіл у порівнянні з ділянками, що містять меншу кількість нектару і пилку.

Для опису поведінки бджіл у природі використовують такі основні поняття, як джерело нектару, робочі бджоли, бджоли-розвідники, бджоли-спостерігачі.

Джерело нектару, джерело їжі характеризується своєю *корисністю*, яку визначають такі чинники, як віддаленість від вулика, концентрація нектару, зручність його видобутку.

Робочі бджоли, зайняті фуражири – бджоли, які *пов'язані* з одним із джерел нектару, тобто добувають на ньому нектар. Зайняті фуражири володіють такою інформацією про «своє» джерело нектару: напрямок від вулика на джерело, корисність джерела.

Бджоли розвідники, незайняті фуражири, бджоли-скаути здійснюють пошук джерел нектару для їх використання.

Бджоли-спостерігачі в даний час виконують деякі роботи у вулику.

Кожна незайнята бджола може полетіти до джерела нектару, слідує за бджолою-розвідником, яка знайшла шлях до такого джерела. Бджола-розвідник виконує *вербування* незайнятих бджіл за допомогою *танцю* на спеціальному майданчику вулика – області танців. *Завербована* бджола прямує за відповідною бджолою-розвідником до області з нектаром, і стає, таким чином, зайнятим фуражиром. Вербування здійснюють ті з бджіл розвідників, які знайшли ділянки з корисністю вище деякого певного порогу.

Зайнятий фуражир після видобутку нектару повертається у вулик і залишає його там. Після цього даний фуражир може виконати одну з наступних дій: залишити «своє» джерело нектару і стати незайнятим фуражиром; продовжити заготівлю нектару з колишнього джерела, не вербуючи при цьому інших бджіл; виконати вербування. Бджола виконує одну з наведених нижче дій за деяким ймовірнісним законом.

Одночасно в межах області танців різні бджоли можуть «рекламувати» різні джерела нектару. Механізми прийняття рішень, відповідно до яких бджола вирішує слідувати за тією чи іншою бджолою-вербувальником, досліджені недостатньо добре. Логічно припустити, що ймовірність вербування тим чи іншим чином визначається корисністю відповідного джерела нектару.

Таким чином, самоорганізація бджолиного рою ґрунтується на чотирьох наступних основних механізмах.

1. Позитивний зворотний зв'язок – на основі інформації, отриманої від інших бджіл, бджола летить до одного з джерел нектару.

2. Негативний зворотний зв'язок – ґрунтуючись на інформації, отриманої від інших бджіл, дана бджола може вирішити, що «її» джерело нектару значно гірше інших знайдених джерел, і залишити це джерело.

3. Випадковість – ймовірнісний пошук бджолами-розвідниками нових джерел нектару.

4. Множинність взаємодії – інформація про джерело нектару, знайденого однією бджолою, передається багатьом іншим бджолам вулика.

4.3.2. Бджолиний алгоритм.

Бджолиний алгоритм (Bees Algorithm, В-алгоритм) запропоновано у роботах Фама (D. T. Pham) із співавторами у 2005 р. Схема В-алгоритму має наступний вигляд. Спочатку із вулика вилітає у випадковому напрямку деяке число бджіл-розвідників, які намагаються відшукати ділянки, де є нектар. Через якийсь час бджоли повертаються у вулик і повідомляють іншим бджолам, де й скільки вони знайшли нектару. Після цього на знайдені ділянки відправляються рекрутовані (робочі) бджоли, причому чим більше на даній ділянці передбачається знайти нектару, тим більше бджіл летить до цієї ділянки. Бджоли-розвідники знову відлітають шукати інші ділянки, після чого процес повторюється.

Припустимо, що мова йде про завдання максимізації фітнес-функції. Кількість нектару в деякій точці простору пошуку вважають пропорційною значенню фітнес-функції в цій точці. Одна ітерація алгоритму включає в себе наступні основні кроки.

1. У випадкові точки простору пошуку відправляють бджіл-розвідників.

2. На підставі значень фітнес-функції, обчислених в зазначених точках, виділяють певну кількість *елітних ділянок* – підобластей простору, відповідних максимальним значенням цієї функції. Аналогічно визначають деяке число *перспективних ділянок*, які відповідають значенням фітнес-функції, близьких до максимальних.

3. На кожну з елітних і перспективних ділянок посилають певне число робочих бджіл. Використовуючи відповідні значення фітнес-функції, знаходять нові елітні і перспективні ділянки. При цьому виборі враховують результати, отримані як бджолами-розвідниками, так і робочими бджолами.

Як поточне наближення до вирішення задачі приймають точку з максимальним значенням фітнес-функції. Розміри елітних і перспективних ділянок

зменшують з ростом числа ітерацій, так що на завершальних ітераціях пошук ведеться тільки в околицях максимумів цільової функції.

Формалізуємо представлену схему В-алгоритму. Розглянемо задачу глобальної умовної оптимізації в гіперпаралелепіпеді Π . Серед бджіл $S = \{s_i, i \in \{1:|S|\}\}$ виділяємо $|S^0| < |S|$ бджіл-розвідників. Покладемо для простоти запису, що ці бджоли є першими в рої S :

$$S^0 = \{s_i^0, i \in \{1:|S^0|\}\}. \quad (4.45)$$

Решта бджіл рою

$$S^w = \{s_i^w, i \in \{1:|S^w|\}\} = \{s_i, i \in [|S^0| + 1:|S|]\} \quad (4.46)$$

є робочими бджолами; $|S^w| = |S| - |S^0|$. Значення фітнес-функції $\varphi(X)$, відповідні поточному стану бджіл s_i^0, s_i^w , позначаємо $\varphi(X_i^0) = \varphi_i^0, \varphi(X_i^w) = \varphi_i^w$.

Множину елітних ділянок позначимо $A^b = \{a_j^b, j \in \{1:|A^b|\}\}$. Аналогічно, множину перспективних ділянок позначимо $A^p = \{a_k^p, k \in \{1:|A^p|\}\}$. В оригінальній роботі, в якій запропоновано В-алгоритм, сукупність елітних і перспективних ділянок називають *відібраними*.

Вважаємо, що ділянки a_j^b, a_k^p є гіперпаралелепіпедами у просторі $R^{|X|}$, грані яких паралельні координатним площинам, центри знаходяться в точках X_j^b, X_k^p і довжини сторін дорівнюють $2r_l^b, 2r_l^p$ відповідно (рис. 4.6). Тут $r_l^b, r_l^p, l \in [1:|X|]$ – радіуси ділянок за відповідними вимірами простору пошуку.

При формуванні елітних ділянок може виявитися, що евклідова відстань $\|X_{j_1}^b - X_{j_2}^b\|_E$ між деякими такими ділянками $a_{j_1}^b, a_{j_2}^b$ не перевищує деякої фіксованої вільної величини $\varepsilon_b > 0$:

$$\|X_{j_1}^b - X_{j_2}^b\|_E \leq \varepsilon_b, j_1, j_2 \in [1:|A^b|], j_1 \neq j_2. \quad (4.47)$$

У такому випадку можливі два наступних варіанти дії: ставлять у відповідність точкам $X_{j_1}^b, X_{j_2}^b$ дві різні ділянки $a_{j_1}^b, a_{j_2}^b$, що перетинаються; ставлять у

відповідність цим точкам одну ділянку a_j^b , центр якої знаходиться в точці $X_{j_1}^b$ або точці $X_{j_2}^b$, що відповідає більшому значенню фітнес-функції, тобто вважають $\varphi(X_j) = \varphi_j = \max(\varphi_{j_1}, \varphi_{j_2})$.

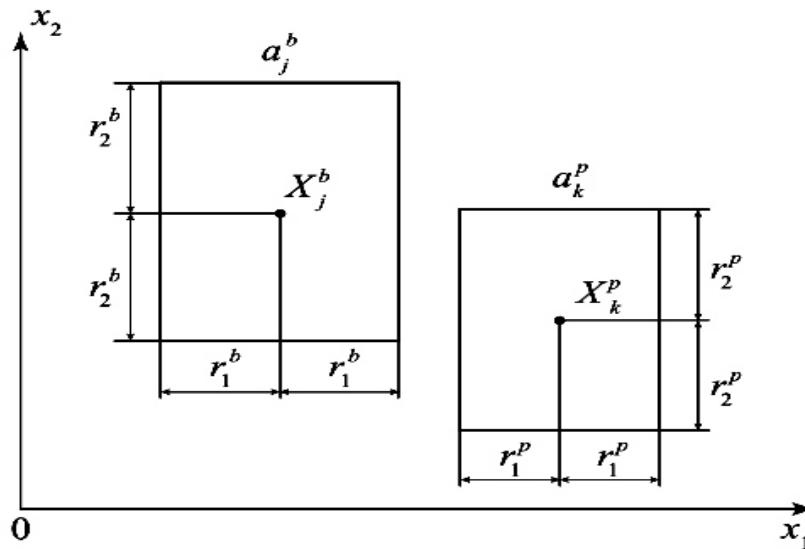


Рис. 4.6. До визначення елітних і перспективних ділянок: $|X|=2$

Аналогічним чином вирішують ситуацію, коли при формуванні ділянок має місце нерівність

$$\|X_{k_1}^p - X_{k_2}^p\|_E \leq \varepsilon_p, \quad k_1, k_2 \in [1:|A^p|], \quad k_1 \neq k_2, \quad (4.48)$$

де $\varepsilon_p > 0$ – вільний параметр.

Нарешті, можлива ситуація, коли малою виявляється евклідова відстань між деякими точками X_j^b , X_k^p . У цьому випадку також можуть бути використані різні варіанти дій, наприклад, точка X_k^p може бути просто виключена з розгляду.

У введених позначеннях схема В-алгоритму полягає в наступному.

1. Генерують випадкові точки X_i , $i \in [1:|S^0|]$, рівномірно розподілені в множині Π , і відправляють в ці точки $|S^0|$ бджіл-розвідників s_i^0 . Обчислюють в

точках X_i значення фітнес-функції $\varphi(X_i) = \varphi_i^0$, сортують величини φ_i^0 за спаданням і представляють у вигляді лінійного списку.

2. Точки X_j , що відповідають першим $|A^b|$ елементам списку, оголошують центрами X_j^b елітних ділянок a_j^b , $j \in [1:|A^b|]$. Аналогічно точки X_k , $k \in [1:|A^p|]$, відповідні наступним $|A^p|$ елементам списку, оголошують центрами перспективних ділянок a_k^p .

3. У кожному з a_j^b елітних і a_k^p перспективних ділянок посилають відповідно по n_b і n_p бджіл-фуражирів $s_i^w \in S^w$, так що

$$n_b |A^b| + n_p |A^p| = |S^w|. \quad (4.49)$$

Координати точок, в які надсилаються ці бджоли вважають випадковими величинами, рівномірно розподіленими у відповідних гіперпаралелепіпедах a_j^b , a_k^p . У всіх точках, в які послані бджоли, обчислюють значення фітнес-функції.

4. Перевіряють виконання умови закінчення ітерацій. Якщо ця умова виконана, то в якості рішення задачі приймають точку, відповідну максимальному досягнутому значенню фітнес-функції, і завершують обчислення. Інакше, пристосованості всіх точок, знайдених бджолами-розвідниками і бджолами-фуражирами, сортують по спаданню і представляють у вигляді лінійного списку. Переходять до кроку 2.

За загальним правилом, як умова закінчення ітерацій може бути використано, наприклад, досягнення заданого числа ітерацій \hat{t} або стагнація обчислювального процесу протягом δ_t ітерацій.

Крім загальних для популяційних алгоритмів, вільними є наступні основні параметри алгоритму:

$|S^0|$ – число бджіл-розвідників;

$|A^b|$ – число елітних ділянок;

$|A^p|$ – число перспективних ділянок;

n_b – число бджіл, що відправляються на елітні ділянки;

n_p – число бджіл, що відправляються на перспективні ділянки;

$r_i^b, r_i^p, i \in [1:|X|]$ – радіуси ділянок.

Збільшення всіх зазначених величин диверсифікує пошук, але уповільнює збіжність алгоритму.

Як варіант, в розглянутій схемі В-алгоритму на ділянки локального пошуку a_j^b, a_k^p можна надсилати не фіксовані числа бджіл, а їх числа, пропорційні значенням фітнес-функції в центрах цих підобластей X_j^b, X_k^p .

Ще одним варіантом В-алгоритму є *полішений бджолиний алгоритм* (Improved Bee Algorithm, IB Algorithm), основною відмінністю якого від В-алгоритму є те, що тут на кожній ітерації поточні радіуси ділянок зменшують за правилом

$$r_l^b = \rho r_l^b, r_l^p = \rho r_l^p, l \in [1:|X|], \quad (4.50)$$

де $\rho \in (0;1]$ – вільний адаптивний *параметр редукції*. Величина параметра ρ збільшується (пошук диверсифікується), якщо на даній ітерації поліпшити рішення не вдалося, і зменшується в іншому випадку.

Розглянута схема В-алгоритму враховує ту обставину, що вирішується задача умовної оптимізації тільки при розміщенні бджіл-розвідників. Якщо ці бджоли опиняються поблизу кордону множини Π , то деякі фрагменти відповідних елітних або перспективних ділянок можуть виходити за її межі. Можна запропонувати декілька варіантів вирішення даної проблеми, з яких найчастіше використовується наступні: застосування в якості фітнес-функції адитивної згортки цільової функції і функції штрафу; використання «невидимих стін», коли ділянки можуть безперешкодно виходити за межі області Π , але значення фітнес-функції бджіл, які знаходяться поза цією областю, ігноруються.

У термінах ГА можна сказати, що В-алгоритм і його модифікації (як і алгоритм мурашиної колонії) використовують речове кодування «безсмертних»

осіб. Механізм зміни координат бджіл в процесі ітерацій можна інтерпретувати як своєрідну мутацію їх координат.

4.3.3. Алгоритм колонії штучних бджіл.

Алгоритм колонії штучних бджіл (Artificial Bee Colony, ABC) запропонував Карабога (D. Karaboga) в 2005 р. для вирішення задачі безперервної безумовної оптимізації. Згодом алгоритм був модифікований у роботах того ж автора і Бастурка (B. Basturk) і поширений на задачу умовної оптимізації. ABC-алгоритм добре зарекомендував себе при вирішенні як унімодальних, так і багатоекстремальних задач оптимізації. Суть алгоритму полягає у комбінуванні алгоритмів глобального і локального пошуку з метою досягнення оптимального балансу між ними.

Розглянемо задачу глобальної умовної мінімізації в гіперкубі:

$$D = \{X \mid x^- \leq x_i \leq x^+, i \in [1:|X|]\} \subset R^{|X|}. \quad (4.51)$$

Слід зазначити, що довільний гіперпаралелепіпед Π за допомогою лінійних перетворень компонент вектору X легко трансформувати в зазначений гіперкуб.

В ABC-алгоритмі популяція містить три групи бджіл: робочі бджоли, бджоли-дослідники і бджоли-розвідники. На різних етапах алгоритму робочі бджоли перетворюються в бджіл-розвідників і навпаки, так що в кожний момент часу популяція складається з множин робочих бджіл і бджіл-дослідників, що мають рівні потужності. Кожній ділянці (джерелу нектару), яка розробляється, ставиться у відповідність одна робоча бджола. Якщо деяка ділянка перестає бути цікавою для популяції, то вона оголошують *зanedбаною* і відповідна робоча бджола стає бджолою-розвідником.

Введемо наступні позначення: s_i^w , $i \in [1:|S^w|]$ – робочі бджоли, s_j^o , $j \in [1:|S^o|]$ – бджоли-дослідники, $|S| = |S^w| + |S^o|$ – загальне число бджіл у рої. Поточні положення робочих бджіл і бджіл-дослідників визначають вектори їх координат X_i^w , X_j^o , відповідно, яким надають сенс положень джерел нектару.

Кількість нектару в його джерелі вважають рівним значенню фітнес-функції у відповідній точці. Занедбаною оголошують ділянку, яка не змінюється протягом заданого числа ітерацій.

Схема ABC-алгоритму включає в себе етап ініціалізації популяції і повторювані ітерації пошукового процесу, кожна з яких складається з наступних стадій: розробка джерел робочими бджолами; розробка джерел бджолами-дослідниками; пошук нових джерел бджолами-розвідниками; завершення ітерацій.

Ініціалізація популяції полягає в генерації в гіперкубі D початкових рішень (координат ділянок) за формулою

$$X_i^w(0) = U_{|X|}(x^-; x^+), \quad i \in [1 : |S^w|], \quad (4.52)$$

і обчисленні кількостей нектару в них

$$\varphi(X_i^w(0)) = \varphi_i^w = \begin{cases} \frac{1}{1 + f(X_i^w(0))}, & f(X_i^w(0)) \geq 0, \\ 1 + \text{abs}(f(X_i^w(0))), & f(X_i^w(0)) < 0. \end{cases} \quad (4.53)$$

Слід зазначити, що формула (4.53) зводить задачу мінімізації цільової функції $f(X)$ до задачі максимізації фітнес-функції.

Рішення $X_i^w(0)$ інтерпретують як джерело, що підлягає розробці робочою бджолою s_i^w .

Стадія розробки ділянок. На цій стадії кожна робоча бджола s_i^w , $i \in [1 : |S^w|]$ намагається відшукати в околицях своєї поточної ділянки X_i^w нову ділянку X_i' з більшою кількістю нектару за наступною схемою:

- 1) генерація точки X_i' ;
- 2) обчислення кількості нектару в цій точці і порівняння його з кількістю нектару в джерелі X_i^w ;
- 3) якщо кількість нектару в новій точці більше, то робоча бджола переміщається в цю точку.

Генерація точки X_i' виконується за наступною формулою

$$x'_{i,j} = \begin{cases} x_{i,j}^w + U_1(-1;1)(x_{i,j}^w - x_{k,j}^w), & j = l, \\ x_{i,j}^w, & j \neq l, \end{cases} \quad (4.54)$$

де $l = U_1[1:|X|]$, $k = U_1[1:|S^w|]$, $k \neq i$. Іншими словами, нове джерело знаходять на прямій, паралельній випадково обраній координатній осі $0-x_l$. Джерело на цій прямій розташовують на випадковій відстані від джерела X_i^w , пропорційному відстані по осі $0-x_l$ до випадково вибраного іншого джерела X_k^w (рис. 4.7).

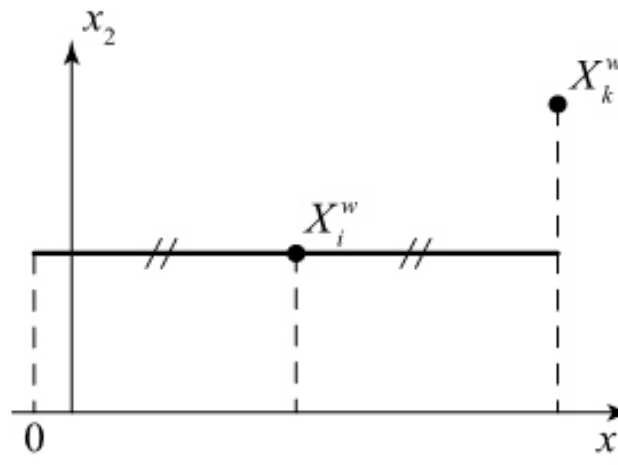


Рис. 4.7. До схеми локального пошуку: ділянка X'_i з рівною ймовірністю знаходиться в будь-якій з точок виділеного відрізка

Слід зазначити наступне – з формули (4.54) випливає, що чим менше відстань між джерелами X_i^w , X_k^w , тим менше відстань між X_i^w і новою точкою X'_i . Оскільки на завершальних ітераціях алгоритму бджоли s_i^w , $i \in [1:|S^w|]$ концентруються поблизу глобального максимуму фітнес-функції, вказана відстань, а тим самим і крок пошуку з ростом числа ітерацій зменшуються. В результаті зменшується широта огляду і підвищується точність локалізації шуканого максимуму.

Розглянута схема генерації точки X'_i може привести до ситуації, коли ця точка виходить за межі області пошуку. Відповідно до АВС-алгоритму у цьому

випадку в якості результуючої точки використовують проекцію точки X'_i на найближчу межу цієї області.

Кількість нектару φ'_i в знайдених точках X'_i , $i \in [1:|S^w|]$ обчислюють за формулою, аналогічною формулою (4.53). Якщо має місце ситуація $\varphi'_i > \varphi_i^w$, то вважають $X_i^w = X'_i$, інакше залишаємо вектор X_i^w без змін.

Стадія розробки джерел бджолами-дослідниками. На цій стадії робочі бджоли діляться інформацією про корисність своїх джерел із бджолами-дослідниками, які на основі отриманої інформації приймають рішення про те, до якого джерела їм доцільно відправитись для його подальшої розробки. Чим вище якість джерела, тим більше бджіл-дослідників до нього буде направлено. Бджоли-дослідники розробляють околиці обраного джерела аналогічно робочим бджолам. Отриману інформацію про якість джерела бджоли-дослідники передають робочим бджолам, які розробляють дане джерело. На основі цієї інформації ті приймають рішення, чи слід їм переключитися на розробку нового джерела чи ні. Після цього бджоли-дослідники повертаються назад у вулик.

Вибір джерела, до якого повинна попрямувати бджола-дослідник, здійснюють за алгоритмом рулетки. Так що ймовірність ξ_i^o того, що бджола-дослідник s_i^o вибере джерело X_i^w , визначає формула

$$\xi_i^o = \frac{\varphi_i^w}{\sum_{j \in [1:|S^w|]} \varphi_j^w}, \quad i \in [1:|S^o|]. \quad (4.55)$$

Бджола-дослідник здійснює локальний пошук в околицях обраного джерела X_i^w , $i \in [1:|S^w|]$ за розглянутою схемою стадії розробки джерел. Кожна з бджіл-дослідників, що розробляє джерело X_i^w , передає робочій бджолі s_i^w , «відповідальній» за це джерело, відповідну корисність, обчислену за формулою (4.53). На підставі інформації від всіх бджіл-дослідників, що розробляють джерело X_i^w , робоча бджола s_i^w приймає рішення про перенесення джерела у нову точку.

Стадія пошуку нових джерел бджолами-розвідниками. Якщо протягом b_t ітерацій корисність джерела X_i^w , $i \in [1:|S^w|]$ не вдається поліпшити, то його переводять в розряд занедбаних. Відповідна цьому джерелу робоча бджола стає бджолою-розвідником. Авторами ABC-алгоритму пропонується використовувати значення параметра b_t , рівне $|X| \cdot |S^w|$.

За схемою ініціалізації популяції для усіх бджіл-розвідників генерують нові джерела і знову переводять усіх цих бджіл в ранг робочих бджіл.

Стадія завершення ітерацій. На даній стадії перевіряють виконання умови завершення ітерацій, і якщо вони виконані, закінчують обчислення. В іншому випадку повертаються до стадії розробки джерел робочими бджолами. Як завжди, в якості умови закінчення ітерацій може бути використана, наприклад, умова досягнення максимального числа ітерацій або умова настання стагнації.

Загальна схема ABC-алгоритму має наступний вигляд.

1. За формулами (4.52), (4.53) ініціалізують популяцію s_i^w робочих бджіл в джерелах X_i^w , $i \in [1:|S^w|]$.

2. Обчислюють початкове наближення до рішення

$$\varphi(\tilde{X}^*) = \max_{i \in [1:|S^w|]} \varphi(X_i) = \tilde{\varphi}^* . \quad (4.56)$$

3. Для кожної робочої бджоли s_i^w за формулою (4.54) генерують точку X_i' і за формулою, аналогічній формулі (4.53), обчислюють її корисність φ_i' . Якщо $\varphi_i' > \varphi_i^w$, то джерело нектару з точки X_i^w переносять у точку X_i' .

4. За формулою (4.55) для кожного з джерел нектару X_i^w обчислюють ймовірність ξ_i^o того, що бджола-дослідник вибере його для розробки. За допомогою алгоритму рулетки призначають кожній бджолі-досліднику джерело, яке їй слід розробляти, і відправляють її до цього джерела.

5. Для кожної бджоли-дослідника за формулою (4.54) генерують точку X_i' і за формулою, аналогічній формулі (4.53) обчислюють її корисність φ_i' .

6. На підставі інформації про результати локального пошуку, виконаного бджолами-дослідниками, які розробляють кожне дане джерело нектару, приймають рішення про переміщення або не переміщення відповідної робочої бджоли в нове джерело.

7. Якщо рішення в джерелі X_i^w не покращується протягом b_t ітерацій, то дане джерело виключають з процесу розробки і відповідну йому робочу бджолу оголошують бджолою-розвідником.

8. Для кожної бджоли-розвідника за формулою (4.52) генерують нове джерело нектару і оголошують її робочою бджолою, яка розробляє це джерело.

9. У кожному поточному джерелі нектару X_i^w , $i \in [1:|S^w|]$ обчислюють значення фітнес-функції $\varphi(X_i)$ і визначають максимальне з цих значень $\varphi(X_{i_j}^w) = \max_{i \in [1:|S^w|]} \varphi(X_i^w)$. Якщо $\tilde{\varphi}^* < \varphi(X_{i_j}^w)$, то вважають $\tilde{\varphi}^* = \varphi(X_{i_j}^w)$.

10. Перевірка умови закінчення ітерацій. Якщо вона виконана, то приймають $X^* = \tilde{X}^*$ і завершують обчислення. Інакше переходять до кроку 3.

Крім загальних для всіх популяційних алгоритмів, вільним параметром алгоритму є параметр b_t – максимальне число невдалих ітерацій.

Наразі відомо значне число модифікацій АВС-алгоритму, що мають на меті підвищення його ефективності. Так, в роботі А.А. Олійника і С.А. Суботіна (2008 р.) запропонована модифікація, яка полягає у наступному.

Якщо загальне число бджіл у даному джерелі X_i досягло свого порогового значення n_{max} , то одна з них (нехай це буде робоча бджола s_i^w) змінює свої координати за правилом, близьким до правила (4.54)

$$x'_{i,j} = \begin{cases} x_{i,j}^w + u_{sign}^{\pm 1} b_x, & j = l, \\ x_{i,j}^w, & j \neq l, \end{cases} \quad i \in [1:|S^w|], \quad j \in [1:|X|], \quad (4.57)$$

де $l = U_1[1:|X|]$ – випадкова натуральна величина, яка визначає координатну вісь, уздовж якої переміщається бджола s_j^w ; $u_{sign}^{\pm 1}$ – випадкова дискретна вели-

чина, яка з однаковою ймовірністю приймає значення ± 1 і визначає напрямок зсуву бджоли s_i^w ; b_x – вільний параметр, крок зсуву.

Корисність φ_i^w , $i \in [1:|S^w|]$ джерела нектару X_i^w визначає формула

$$\varphi_i^w = \begin{cases} 1, & \tilde{\varphi}_i^w > 1, \\ \tilde{\varphi}_i^w, & \tilde{\varphi}_i^w \in [\delta_\varphi; 1], \\ 0, & \tilde{\varphi}_i^w < \delta_\varphi, \end{cases} \quad \tilde{\varphi}_i^w = 1 - \bar{\varphi}_i^w + U_1(-b_n; b_n) \quad (4.58)$$

Тут $\delta_\varphi > 0$ – поріг корисності джерела (рекомендоване значення дорівнює 0,1); $b_n > 0$ – рівень випадкових збурень (рекомендоване значення дорівнює 0,1); $\bar{\varphi}_i^w$ – нормована корисність джерела нектару X_i^w , яка розрахована на основі його мінімальної і максимальної корисностей протягом усіх ітерацій $[1:t]$.

На поточній ітерації ймовірність ξ_i^w виконання бджолою s_i^w процедури вербування інших бджіл оцінюють у відповідності з виразом

$$\xi_i^w = \frac{a_i}{b_\xi}, \quad i \in [1:|S^w|], \quad (4.59)$$

де $a_i = \max(\varphi_i^w - b_a \bar{\varphi}^w, 0)$ – привабливість джерела X_i^w ; $b_\xi > 0$ – параметр, керуючий ймовірністю ξ_i^w ; $\bar{\varphi}^w$ – середня поточна корисність усіх $|S^w|$ ділянок; $b_a > 0$ – вільний коефіцієнт, який регулює вплив величини $\bar{\varphi}^w$.

Ймовірність того, що незайнятий фуражир стане розвідником, визначає формула

$$\xi^o = \exp\left(-\frac{a^2}{2b_o}\right), \quad (4.60)$$

де $a = \sum_{i \in [1:|S^w|]} a_i$ – поточна привабливість джерел; $b_o > 0$ – вільний параметр.

Авторами даної модифікації ABC-алгоритму виконано порівняння її ефективності з канонічним алгоритмом рою часток. Дослідження показало, що при оптимізації складних багатовимірних багатоекстремальних функцій дана модифікація забезпечує більш високу точність локалізації глобального екстремуму.

Контрольні запитання

1. Наведіть схему канонічного алгоритму рою часток. Визначте поняття *власний шлях* частки і її *локальний шлях*.
2. Сформулюйте основні відмінні ознаки наступних модифікацій алгоритму рою часток: притягнуто-відштовхуючий алгоритм, алгоритм рою часток з керуванням швидкістю часток, алгоритм рою часток з повною інформацією (FIPS), алгоритм FIPS з вагами, ранжируваний алгоритм FIPS, FDR алгоритм рою часток.
3. Дайте визначення поняття *топологія сусідства часток*. Проілюструйте топології сусідства типу кліка, кільце, двовимірний тор, кластерна топологія.
4. Визначте поняття *динамічна топологія сусідства часток*. Назвіть основні особливості наступних алгоритмів рою часток, що використовують динамічні топології сусідства: клубний алгоритм рою часток, алгоритм рою часток зі стереотипуванням, алгоритм з доповненням графа сусідства часток.
5. Які біонічні передумови мурашиних алгоритмів оптимізації?
6. Наведіть схему алгоритму безперервної оптимізації колонією мурах CASO. Деталізуйте зміст основних етапів цього алгоритму.
7. Назвіть основні відмінні ознаки алгоритму безперервно взаємодіючої колонії мурах CIAC.
8. Наведіть схему безперервного ортогонального алгоритму мурашиної колонії COAC.
9. Наведіть схему бджолиного алгоритму (Bees Algorithm).
10. Наведіть схему алгоритму колонії штучних бджіл ABC. Деталізуйте основні оператори цього алгоритму.

ЧАСТИНА II.

МОДЕЛЮВАННЯ НЕЛІНІЙНИХ ДИНАМІЧНИХ ПРОЦЕСІВ

Глава 5. НЕЛІНІЙНА ДИНАМІКА ТА ХАОС

5.1. Динамічні системи

5.1.1. Визначення динамічної системи, класифікація.

Під динамічною системою (ДС) розуміють будь-який об'єкт або процес, для якого однозначно визначено поняття стану як сукупності деяких величин або функцій в даний момент часу, і заданий закон, який описує зміну (еволюцію) початкового стану з плином часу. Цей закон дозволяє за початковим станом прогнозувати майбутній стан ДС, і його називають законом еволюції.

Розглянемо ДС, що моделюються кінцевим числом звичайних диференціальних рівнянь (ЗДР). Опис стану задамо величинами x_1, x_2, \dots, x_N в деякий момент часу $t=t_0$. Тоді закон еволюції ДС має вигляд

$$\frac{dx_i}{dt} = \dot{x}_i = f_i(x_1, x_2, \dots, x_N), \quad i = 1, 2, \dots, N. \quad (5.1)$$

Якщо розглядати величини x_1, x_2, \dots, x_N як координати точки x в N -вимірному просторі, то виходить наочне геометричне уявлення стану ДС у вигляді цієї точки. Останню називають *зображуючою*, а частіше – *фазовою точкою*, а простір станів – *фазовим простором* ДС. Зміні стану системи у часі відповідає рух фазової точки вздовж деякої лінії, яка називається *фазовою траєкторією*.

Необхідно уточнити взаємозв'язок понять числа ступенів свободи і розмірності фазового простору ДС. Під числом *ступенів свободи* розуміється найменше число незалежних координат, необхідних для однозначного визначення стану системи. Під координатами спочатку розумілись саме просторові змінні, що характеризують взаємне розташування тіл і об'єктів. У

той же час для однозначного рішення відповідних рівнянь руху необхідно, окрім координат, задати відповідні початкові значення імпульсів або швидкостей. У зв'язку з цим система з n ступенями свободи характеризується фазовим простором розмірністю в два рази більшою ($N=2n$).

ДС формально визначена, якщо задані:

1) множина станів X , яке утворює повний метричний простір (фазовий простір);

2) множина моментів часу Θ ;

3) оператор еволюції $T_{t_0}^T$ – деяке відображення $T_{t_0}^T : X \rightarrow X$, яке кожному

стану $x_0 \in X$ в початковий момент часу $t_0 \in \Theta$ однозначно ставить у відповідність деякий стан $x_t \in X$ в будь-який інший момент часу $t = t_0 + \tau \in \Theta$. Таким чином, можна записати:

$$\mathbf{x}_t = T_{t_0}^T \mathbf{x}_0, t = t_0 + \tau, \quad (5.2)$$

де τ – інтервал (зсув) часу.

Оператор еволюції є безперервним в X і має такі властивості:

$$T_{t_0}^T \mathbf{x}_0 = \mathbf{x}_0; \quad (5.3)$$

$$T_{t_0}^{T+s} \mathbf{x}_0 = T_{t_0+s}^T \circ T_{t_0}^S \mathbf{x}_0 = T_{t_0+\tau}^S \circ T_{t_0}^T \mathbf{x}_0, \quad (5.4)$$

де \circ – означає суперпозицію операторів.

Виходячи з характеру множин X , Θ і властивостей оператора еволюції можна дати найбільш загальну класифікацію ДС.

Якщо $\Theta=R^1$ (час приймає безперервну множину значень), то оператор еволюції безперервний по τ і відповідна ДС є системою з безперервним часом або потоком.

Якщо множина Θ є рахунковою, то ДС є системою з дискретним часом або каскадом.

Можна виділити клас лінійних ДС, для яких оператор еволюції є лінійним, тобто задовольняє правилу суперпозиції:

$$T_{t_0}^\tau (\mathbf{x} + \mathbf{y}) = T_{t_0}^\tau \mathbf{x} + T_{t_0}^\tau \mathbf{y}. \quad (5.5)$$

Якщо оператор нелінійний (не задовольняє правилу (5.5)), то ДС є *нелінійною*.

Якщо оператор еволюції $T_{t_0}^\tau$ визначений для всіх значень зсуву в часі τ , як для $\tau \geq 0$, так і для $\tau < 0$, то він є оборотним, тобто існує зворотний до нього оператор $T_{t_0+\tau}^{-\tau}$, який дозволяє, знаючи стан системи в момент $t=t_0+\tau$, знайти стан системи у попередній момент t_0 . ДС також називається *оборотною в часі*. Якщо оператор еволюції визначено тільки для $\tau \geq 0$, то він незворотний і попередній стан системи однозначно визначити не можна. Система у цьому випадку є *незворотною у часі*.

Якщо оператор еволюції $T_{t_0}^\tau$ не залежить від моменту часу t_0 , а визначається тільки початковим станом \mathbf{x}_0 і інтервалом τ , то відповідна ДС є *автономною*, в іншому випадку система є *неавтономною*. В позначенні оператора еволюції автономної системи не потрібно вказувати початковий момент часу, тобто $T_{t_0}^\tau = T^\tau$, а властивість (5.4) набуває вигляду

$$T^{\tau+s} \mathbf{x}_0 = T^\tau \circ T^s \mathbf{x}_0 = T^s \circ T^\tau \mathbf{x}_0. \quad (5.6)$$

З фізичної точки зору автономність системи означає, що на систему не діють ніякі зовнішні сили і параметри системи постійні у часі.

Якщо оператор еволюції зберігає фазовий обсяг, то ДС є *консервативною*. Повна енергія консервативної системи залишається постійною. Якщо оператор еволюції стискає фазовий обсяг, то система є *дисипативною*. В такій системі відбувається розсіювання (дисипація) енергії.

Способи завдання оператора еволюції можуть бути різними: у вигляді інтегрального перетворення, у вигляді матриці або таблиці, у вигляді графіка або функції тощо.

Важливу групу ДС представляють системи, в яких можливі коливання. Коливальні системи з точки зору їх математичних моделей поділяють на певні класи. Розрізняють *лінійні* і *нелінійні* коливальні системи, *зосереджені* і *розподілені*, *консервативні* і *дисипативні*, *автономні* і *неавтономні*. Особливий клас представляють так звані *автоколивальні системи*.

Коливальна система називається *лінійною* або *нелінійною* залежно від того, лінійна або нелінійна система диференціальних рівнянь, яка її описує.

Коливальні системи, що моделюються кінцевим числом звичайних диференціальних рівнянь, називають *зосередженими* або *точковими* системами. Вони описуються за допомогою кінцеве вимірною фазового простору і характеризуються кінцевим числом ступенів свободи. Одна й та ж система в різних умовах може розглядатися як зосереджена або як розподілена. Математичні моделі *розподілених* систем – це диференціальні рівняння в часткових похідних, інтегральні рівняння або звичайні рівняння з аргументом, який запізнюється. Число ступенів свободи розподіленої системи нескінченно, і потрібно нескінченне число даних для визначення її стану.

За енергетичною ознакою ДС діляться на консервативні і неконсервативні. *Консервативні* системи характеризуються незмінним у часі запасом енергії. У механіці їх називають *гамільтоновими*.

ДС із змінним в часі запасом енергії називаються відповідно *неконсервативними*. Системи, в яких енергія зменшується в часі через тертя або розсіювання, є *дисипативними*. Відповідно до цього системи, енергія яких у часі наростає, називаються системами з негативним тертям або негативною дисипацією. Такі системи можна розглядати як дисипативні при зміні напрямку відліку часу на протилежний. Принциповою особливістю дисипативних систем є залежність елемента фазового обсягу від часу. У системах з поглинанням енергії фазовий обсяг в часі зменшується, в системах з негативним тертям – збільшується.

Коливальні системи називаються *автономними*, якщо вони не схильні до дії зовнішніх сил, змінних у часі. Рівняння автономних систем явної залежності від часу не містять.

Більшість реальних коливальних систем у фізиці, радіофізиці, біології, хімії та інших галузях знань неконсервативні. Серед них виділяється особливий клас так званих *автоколивальних* систем, які принципово неконсервативні і нелінійні. Автоколивальною називають ДС, яка перетворює енергію джерела в енергію незатухаючих коливань, причому основні характеристики коливань (амплітуда, частота, форма коливань тощо) визначаються параметрами системи і в певних межах не залежать від вибору початкового стану.

5.1.2. Регулярні і хаотичні атрактори.

Всі траєкторії у фазовому просторі дисипативної системи можна розділити на траєкторії, що відповідають перехідним процесам (процесам релаксації системи до деяких сталих режимів), і траєкторії, що належать інваріантним граничним множинам. Траєкторії другого типу відповідають усталеним режимам функціонування системи.

Дамо загальне визначення граничних множин ДС. Точка $p \in R^N$ називається ω -граничною точкою траєкторії $x(t)$, $t \geq t_0$, якщо існує послідовність $t_k \rightarrow \infty$ при $k \rightarrow \infty$, така, що послідовність станів $x(t_k)$ сходиться до точки p . Аналогічно, точка $q \in R^N$ називається α -граничною точкою траєкторії $x(t)$, $t \leq t_0$, якщо існує послідовність $t_k \rightarrow -\infty$ при $k \rightarrow \infty$, така, що послідовність станів $x(t_k)$ сходиться до точки q . Множина усіх ω -граничних точок траєкторії $x(t)$ називається ω -граничною множиною даної траєкторії. Позначимо її $\omega(x(t))$. Множина всіх α -граничних точок траєкторії $x(t)$ є α -граничною множиною даної траєкторії. Позначимо її $\alpha(x(t))$.

Граничні множини будь-якої фазової траєкторії самі складаються з фазових траєкторій. Крім того, граничні множини ω і α інваріантні щодо оператора еволюції. Це означає, що оператор еволюції відображає будь-яку точку граничної множини в точку цієї ж множини.

Розглянувши, куди прагнуть різні траєкторії ДС в прямому і зворотному часі, можна виділити усі інваріантні граничні множини в фазовому просторі. Типові граничні множини траєкторій на фазовій площині – це *стани рівноваги*, *періодичні рухи* і *особливі траєкторії* типу сепаратрисних контурів, двійково асимптотичні до сідлових станів рівноваги. Зазначені граничні множини повністю вичерпують можливі ситуації на фазовій площині. Їм відповідають три різних типи рішень рівнянь. Сепаратрисні контури і петлі – це особливі криві, які в дисипативних системах не є *структурно стійкими (грубими)*. Вони існують тільки при певних значеннях параметрів і зникають при як завгодно малому збуренні оператора еволюції. Навпаки, точки рівноваги і граничні цикли при виконанні деяких умов будуть структурно стійкими граничними множинами і можуть існувати в деякій області простору параметрів.

Розглянемо структурно стійкі граничні множини в R^N . Якщо деяка множина Q є граничною множиною у фазовому просторі ДС, це означає одну з трьох можливостей:

1. Q є ω -граничною множиною для всіх траєкторій з деякої області фазового простору U , які не належать до Q . У цьому випадку множина Q є *притягаючою граничною множиною* або *атрактором* ДС;

2. Q є α -граничною множиною для всіх траєкторій з U , які не належать до Q . Тоді Q називають *відштовхуючою граничною множиною* або *репелером*;

3. В U існують траєкторії, які не належать Q , для яких Q є ω -граничною множиною, і траєкторії, для яких Q є α -граничною множиною. У цьому випадку множина Q є *сідлом* (наприклад, сідлова точка рівноваги, сідловий граничний цикл, сідловий тор).

Єдиного загальноновживаного визначення атрактора не існує. Найбільш часто, особливо в математичній літературі, використовується визначення атрактора як *максимального атрактора поглинаючої області*. Пояснимо це визначення.

Нехай ДС задана оператором еволюції $T^\tau: R^N \rightarrow R^N$ і нехай B є поглинаючою областю в R^N , тобто для B виконується умова: $T^\tau \bar{B} \subset B, \tau \gg 0$. Максимальним атрактором A_{\max} в поглинаючій області B є множина

$$A_{\max} = \bigcap_{\tau > 0} T^\tau B. \quad (5.7)$$

Деяку інваріантну множину A назвемо *атрактором* ДС, якщо існує поглинаюча область, для якої A є максимальним атрактором. *Басейном* (областю) тяжіння атрактора A називається множина U , така, що всі траєкторії з U прагнуть до A при $t \rightarrow \infty$.

Важливим результатом досліджень останніх років стало виявлення принципово нових типів рухів в ДС. Таким рухам у фазовому просторі розмірністю $N \geq 3$ відповідають складним чином влаштовані притягуючі множини, траєкторії зображуючих точок яких не належать до жодного з описаних вище типів атракторів. Фазові траєкторії представляються тут у вигляді нескінченної кривої, яка ніде не перетинається. При $t \rightarrow \infty$ траєкторія не покидає замкнутої області і не притягується до відомих типів атракторів. Саме із існуванням таких траєкторій пов'язують можливість хаотичної поведінки детермінованих ДС з розмірністю фазового простору $N \geq 3$. Вперше подібні властивості ДС у 1963 р виявив Е. Лоренц при чисельному дослідженні динаміки тривимірної моделі теплової конвекції. Через 8 років в теоретичній роботі Д. Рюеля і Ф. Такенса притягаюча область у фазовому просторі ДС, що характеризується режимом сталих неперіодичних коливань, була названа *дивним атрактором*. Цей термін був відразу сприйнятий дослідниками і затвердився для позначення математичного образу режиму нерегулярних коливань детермінованих ДС.

Атрактори у вигляді станів рівноваги, граничних циклів або l -мірних торів називають *простими* або *регулярними*, підкреслюючи тим самим, що рух на них відповідає сформованим уявленням про стійку по Ляпунову детерміновану поведінку ДС. З *дивним* атрактором зв'язується реалізація

нерегулярного (в сенсі відсутності періодичності) коливального режиму, який багато у чому схожий із нашими уявленнями про стаціонарні випадкові процеси.

Однак термін «випадковий» має цілком певний сенс. Випадковий рух непередбачуваний або передбачуваний з певною ймовірністю. Іншими словами, траєкторії випадкового руху не можна багаторазово і однозначно відтворити ні в чисельному, ні в фізичному експериментах. Прикладом служить класичний рух броунівської частинки. У разі дивного атрактора є суворе передбачуваність в сенсі детермінованості закону еволюції. Рішення рівнянь (як і для регулярних атракторів) підпорядковується теоремі єдності й однозначно відтворюється при фіксованих початкових умовах. Тому для позначення складних «шумоподібних» автоколивань, математичним образом яких служить дивний атрактор, використовуються терміни типу *динамічна стохастичність*, *детермінований хаос* і подібні. Важливо відрізнити ці процеси від стохастичних у класичному сенсі, які при описі вимагають врахування флуктуацій у початкових динамічних рівняннях або безпосередньо підпорядковуються рівнянням для щільності розподілу ймовірностей статистичної теорії.

Прикладом системи з хаотичним атрактором є рівняння генератора з інерційною нелінійністю (генератора Аніщенко-Астахова). Ця система є узагальненням рівнянь Ван дер Поля на випадок тривимірного простору:

$$\begin{aligned} \dot{x} &= mx + y - xz, & \dot{y} &= -x, & \dot{z} &= gz + gI(x)x^2, \\ I(x) &= \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases} \end{aligned} \quad (5.8)$$

Результати чисельного рішення рівняння (5.8) для значень параметрів $m=1.5$, $g=0.2$ наведені на рис. 5.1, який ілюструє хаотичний атрактор.

5.1.3. Системи з дискретним часом. Відображення послідування.

Системи з дискретним часом (каскади) відіграють важливу роль в нелінійній динаміці.

Їх називають також *точковими відображеннями* або *відображеннями послідування*, які можуть бути записані у вигляді

$$x(n+1) = P(x(n)), \quad (5.9)$$

де $x \in R^N$ – вектор стану, n – номер ітерації (дискретний час), $P(x)$ – деяка вектор-функція, звана *функцією послідування*. Функція послідування в явному вигляді задає оператор еволюції на одному кроці за часом (за одну ітерацію): $T_n^1 x = P(x(n))$. Знаючи вектор стану x в деякий момент часу n , ми можемо знайти стан системи у будь-який інший момент часу $t=n+k$, $k>0$:

$$x(n+k) = T_n^k x = P^{(k)}(x(n)) = P(P(\dots P(x(n)) \dots)). \quad (5.10)$$

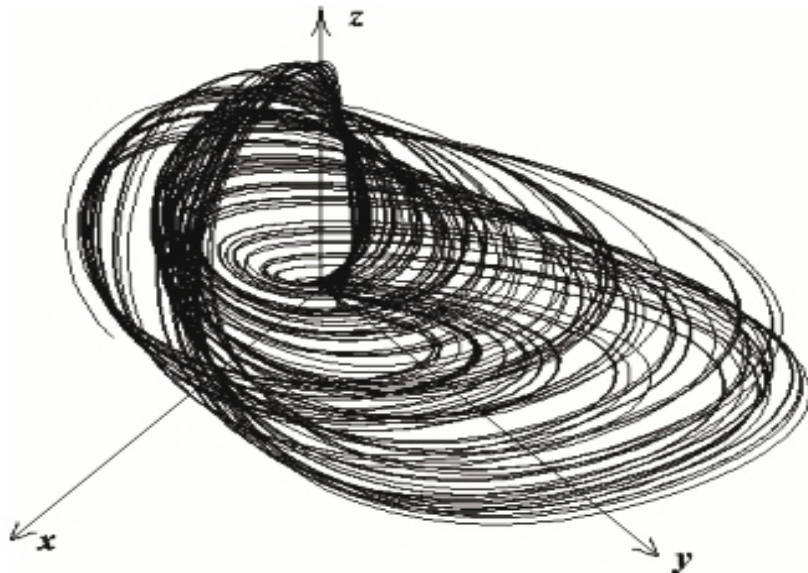


Рис. 5.1. Дивний атрактор в моделі генератора Аніщенко-Астахова

Також як потоки, відображення послідування можуть бути дисипативними (стискаючими) і консервативними (зберігаючими обсяг). У фазовому просторі дисипативних відображень можна виділити атрактори й інші граничні множини. Фазова траєкторія відображення (5.9) складається з послідовності точок фазового простору: $x_1, x_2, \dots, x_i, \dots$, де $x_i = x(i)$. Фазова траєкторія, що належить граничній множині системи (5.9), може складатися з

однієї точки x_0 , званої *нерухомою точкою відображення*. Для неї справедливо $P(x_0) = x_0$.

Якщо траєкторія замкнута і складається з k точок x_i , $i = 1, 2, \dots, k$, таких, що для будь-якого i $P^k(x_i) = x_i$, то множина точок $x_i \in$ *циклом відображення періоду k* . Точки циклу називають також *нерухомими точками кратності k* . Квазіперіодичні та хаотичні траєкторії відображення послідування є незамкнутими послідовностями точок, які ніколи не повертаються строго у свої попередні положення.

Оборотні відображення послідування можуть бути безпосередньо пов'язані з поточковими системами, що задаються ЗДР. Для того, щоб від потокової системи перейти до відображення з дискретним часом, потрібно ввести січну поверхню S (у багатовимірному випадку – гіперповерхню), так, щоб всі фазові траєкторії перетиналися з нею строго трансверсально. Якщо розглядати точки перетину траєкторій з поверхнею S при русі в одному напрямку, то потік породжує в S відображення послідування, зване також відображенням Пуанкаре (рис. 5.2). Відображення Пуанкаре обов'язково оборотно однозначно (але не взаємно однозначно) пов'язане з початковим потоком. Порядковий номер перетину заданої траєкторії із січною поверхнею є дискретним часом n .

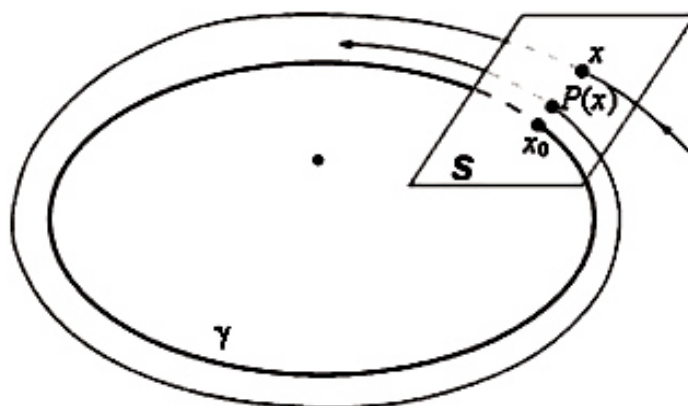


Рис. 5.2. Побудова відображення Пуанкаре на січній поверхні S

Розмірність усіх граничних множин у фазовому просторі системи при переході до відображення Пуанкаре знижується на одиницю, що робить фазові портрети відображення більш наочними. Граничним циклом потокової системи відповідають нерухомі точки або цикли відображення Пуанкаре, що складаються з k точок (рис. 5.2). Число k визначається тим, скільки раз траєкторія перетнула поверхню S в обраному напрямку. Квазіперіодичним і хаотичним траєкторіям потокової системи відповідають квазіперіодичні та хаотичні незамкнуті траєкторії відображення. Квазіперіодичні траєкторії, породжувані ергодичними траєкторіями на двовимірному торі, заповнюють інваріантну замкнуту криву, яка є образом двовимірного тора в відображенні (рис. 5.3,*a*). Хаотичні траєкторії відображення Пуанкаре належать множинам, які мають складну геометричну структуру. Якщо розмірність хаотичного атрактора невелика, то його геометрична структура в січній S більш наочна, ніж «клубок ниток», що спостерігається у фазовому просторі потокової системи (рис. 5.3,*б*).

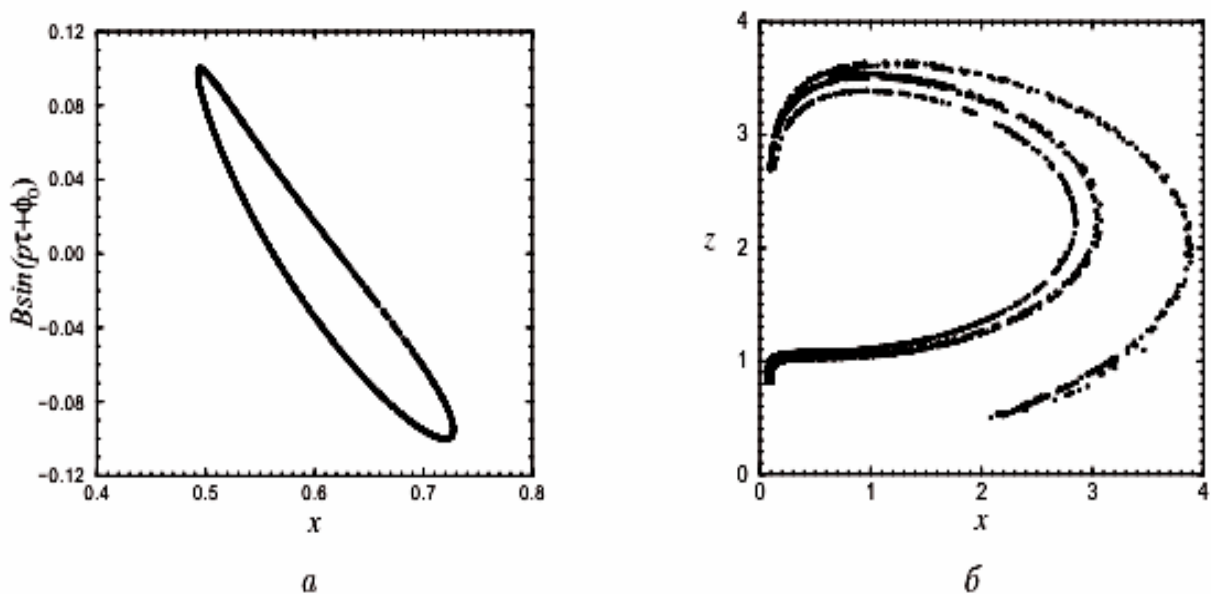


Рис. 5.3. Перетини Пуанкаре складних граничних множин: *a*– інваріантна замкнута крива; *б*– множина зі складною геометричною структурою

Тут інваріантна замкнута крива (див. рис. 5.3,*a*) в перерізі двовимірного тора в (5.8) отримана для параметрів $a=1$, $\varepsilon=0.1$, $b=0.3$, $B=0.1$, $p=1.35$, $\varphi_0=0$. А множина зі складною геометричною структурою (див. рис. 5.3,*б*) в перерізі хаотичного атрактора в системі (5.8) отримана при $m=1.5$, $g=0.2$. Як січна поверхня S в обох випадках використовувалась площина $y=0$.

Якщо ДС має розмірність фазового простору $N=3$, то відображення Пуанкаре буде двовимірним. У цьому випадку воно може бути зведене до відображення площини, що робить динаміку системи особливо наочною. Якщо у фазовому просторі потокової системи є сильне стиснення вздовж деяких напрямків, то може виявитись, що точки в перетині S лягають практично на криву. Насправді це не одновимірна крива, а фрактальна множина зі складною поперечною структурою, яка не помітна через сильне стиснення.

Отримання двовимірних і одновимірних відображень послідування в явній формі (у вигляді системи (5.9)) для конкретних потокових ДС виявляється в більшості випадків складним завданням або взагалі неможливо. Однак існує безліч модельних одновимірних і двовимірних точкових відображень, які, не є безпосередньо пов'язані з конкретними потоковими системами, проте широко застосовуються в нелінійній динаміці для дослідження й опису тих чи інших явищ фундаментального характеру:

1. *Відображення розтягування:*

$$x(n+1) = \alpha x(n), \text{ mod } 1.$$

Відображення задано на інтервалі $[0; 1]$, *mod* 1 означає, що береться дрібна частина числа; $\alpha > 0$ – параметр відображення.

2. *Логістичне відображення:*

$$x(n+1) = \alpha x(n)(1 - x(n)).$$

Відображення задано на інтервалі $[0; 1]$, $\alpha \in [0; 4]$ – параметр. Логістичне відображення може також бути приведене до вигляду

$$x(n+1) = a - x^2(n) \text{ або } x(n+1) = 1 - \varepsilon x^2(n).$$

3. *Відображення окружності:*

$$x(n+1) = \alpha x(n)(1 - x(n)).$$

Відображення задано на інтервалі $[0; 1]$, $\Omega \in [0; 1]$ і $K \geq 0$ – параметри.

4. *Відображення Ено*:

$$x(n+1) = 1 - ax^2(n) + y(n),$$

$$y(n+1) = by(n).$$

a і b – параметри відображення.

5. *Відображення Лозі*:

$$x(n+1) = 1 - a|x(n)| + y(n),$$

$$y(n+1) = by(n).$$

Відображення площині Ено і Лозі є оборотними.

5.1.4. Стійкість динамічних систем.

Існує досить багато різних визначень стійкості, з яких найбільш часто використовуються наступні: стійкість по Пуассону, стійкість по Ляпунову і асимптотична стійкість.

Нехай ДС задана системою ЗДР (5.1) і нехай нас цікавить стійкість траєкторії $x^0(t)$.

Стійкість по Пуассону означає, що через деякий час фазова траєкторія повертається в як завгодно малу околицю початкової точки $x_0^0 = x^0(t_0)$. Причому, якщо система оборотна, то повернення відбувається як в прямому, так і в зворотному часі. Таким чином, кожна точка стійкої по Пуассону траєкторії одночасно є її α - і ω -граничною точкою. Інтервал часу, після якого траєкторія повертається в околиця точки x_0^0 заданого радіуса ε , називається *періодом повернення Пуанкаре*. Періоди повернення можуть відповідати періоду або квазіперіоду регулярного руху або представляти випадкову послідовність в режимі динамічного хаосу (рис. 5.4).

Стійкість по Пуассону є важливою, але слабкою властивістю стійкості. Ми нічого не можемо сказати про поведінку сусідніх траєкторій, початково близьких до $x^0(t)$. У практичних задачах нас найчастіше цікавить інша

властивість стійкості, пов'язана з малим збуренням заданої траєкторії. Залежно від поведінки малого збурення в часі розрізняють стійкість по Ляпунову і асимптотичну стійкість.

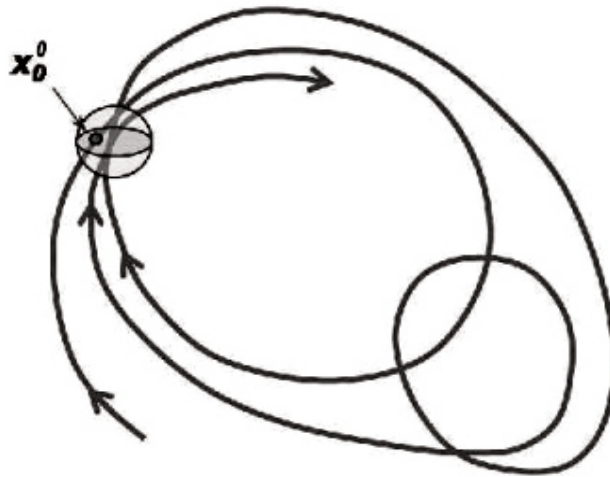


Рис. 5.4. Стійка по Пуассону незамкнута траєкторія

Траєкторія $\mathbf{x}^0(t)$ називається *стійкою по Ляпунову*, якщо для будь-якого довільно малого $\varepsilon > 0$ існує таке $\delta(\varepsilon) > 0$, що для будь-якої траєкторії $\mathbf{x}(t)$, для якої $\|\mathbf{x}(t_0) - \mathbf{x}^0(t_0)\| < \delta$, при всіх $t > t_0$ виконується нерівність $\|\mathbf{x}(t) - \mathbf{x}^0(t)\| < \varepsilon$. Символ $\|\dots\|$ позначає норму в R^N . Таким чином, мале початкове збурення стійких по Ляпунову фазових траєкторій не збільшується з плином часу (рис. 5.5,а). Якщо мале збурення δ з часом зменшується, тобто $\|\mathbf{x}(t) - \mathbf{x}^0(t)\| \rightarrow 0$ при $t \rightarrow \infty$, то траєкторія має сильнішу стійкість, а саме *асимптотичну стійкість* (рис. 5.5,б). Будь-яка асимптотично стійка фазова траєкторія стійка по Ляпунову. Протилежне твердження у загальному випадку не вірно.

Властивості стійкості фазових траєкторій, що належать граничним множинам (наприклад, аттракторам), мають особливу важливість при дослідженні ДС. Зміна характеру стійкості тієї або іншої граничної множини у багатьох випадках призводить до зміни режиму функціонування системи.

Розглянемо автономну ДС, що задається векторним диференціальним рівнянням виду

$$\dot{x} = F(x), \quad x \in R^N. \quad (5.11)$$

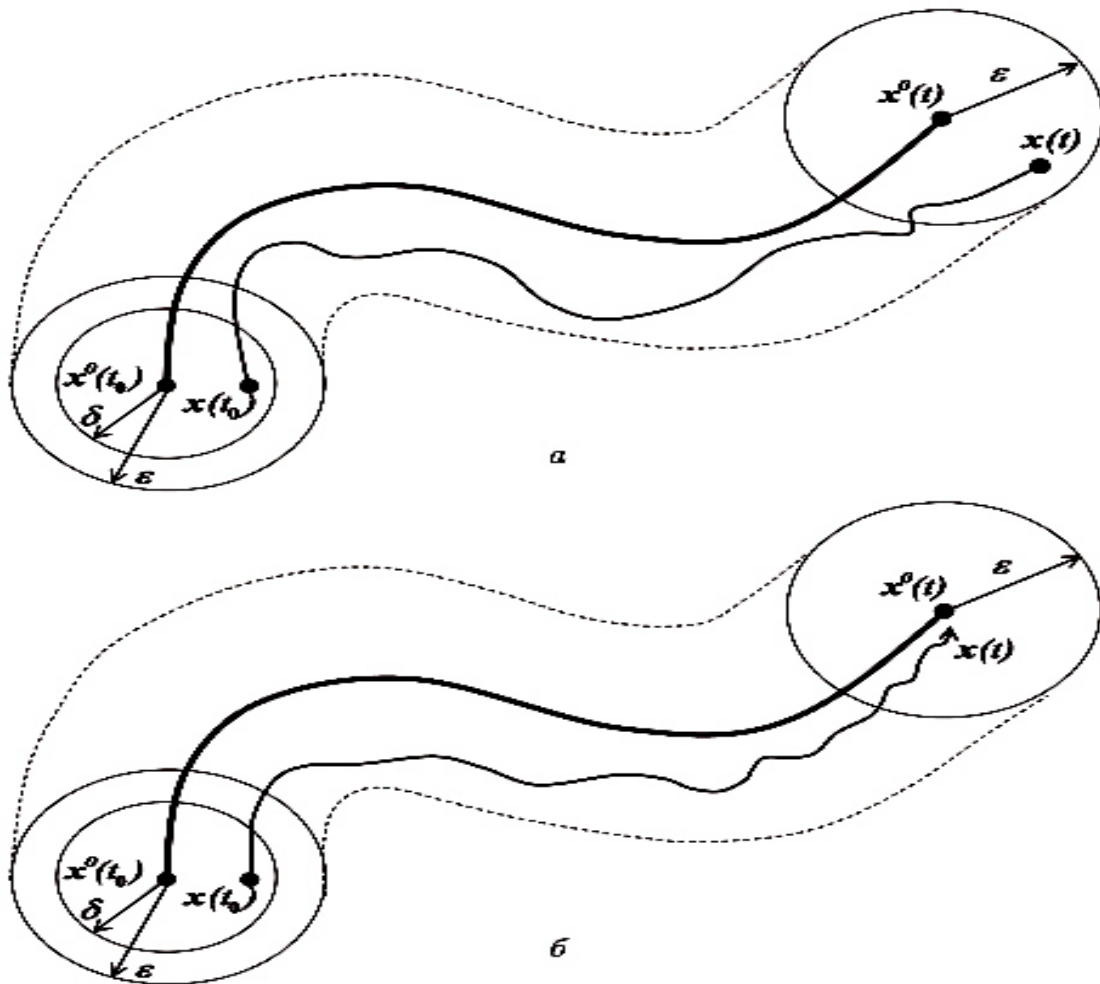


Рис. 5.5. Ілюстрації до визначення стійкості:

а – по Ляпунову; *б* – асимптотичної стійкості траєкторії $x^0(t)$

Проаналізуємо стійкість часткового рішення $x^0(t)$. Рівняння (5.11) може мати в якості рішення не тільки точки рівноваги, але також періодичні, квазіперіодичні та хаотичні траєкторії.

Введемо вектор збурення $y = x(t) - x^0(t)$, припускаючи, що його довжина $\|y\|$ – мала. Для y можна записати рівняння

$$\dot{y} = F(x^0 + y) - F(x^0). \quad (5.12)$$

Розкладаючи $F(x^0 + y)$ в ряд в околиці x^0 і беручи до уваги той факт, що збурення по нормі є малим, приходимо до наступного лінеаризованого рівняння відносно y :

$$\dot{y} = \widehat{A}(t)y, \quad (5.13)$$

де $\widehat{A}(t)$ – матриця з елементами

$$a_{j,k}(t) = \left. \frac{\partial f_j}{\partial x_k} \right|_{x(t)=x^0(t)}, j,k=1,2,\dots,N, \quad (5.14)$$

яка звана *матрицею лінеаризації* системи в околиці рішення $x^0(t)$, а f_j – компоненти вектор-функції F .

Оскільки елементи матриці \widehat{A} залежать від точки на досліджуваній траєкторії, то в загальному випадку вони змінюються в часі. Матриця характеризується N власними значеннями $s_i(t)$, також мінливими в часі. Власні значення є коренями характеристичного рівняння

$$\text{Det}[\widehat{A}(t) - s\widehat{E}] = 0, \quad (5.15)$$

де \widehat{E} – одинична матриця. N власним значенням (включно кратні) відповідають N лінійно незалежних власних вектора $e_i(t)$, що змінюють свої напрямки при русі вздовж траєкторії $x^0(t)$ і задовольняють рівнянню

$$\widehat{A}(t)e_i(t) = s_i(t)e_i(t), i=1,2,\dots,N. \quad (5.16)$$

Розглянемо матрицю лінеаризації в точці досліджуваної траєкторії в момент часу t . Початкове збурення у напрямку i -го власного вектора, задане в момент t , змінюється за малий інтервал часу τ наступним чином:

$$y^i(t + \tau) = y^i(t) \exp[\tau s_i(t)]. \quad (5.17)$$

Збільшення або зменшення норми збурення $\|y^i(t + \tau)\|$ визначається знаком дійсної частини $s_i(t)$. При русі вздовж траєкторії $x^0(t)$ показник

експоненти приймає різні значення. Отже, не виключено, що коли збурення $\mathbf{y}(t+\tau) = \sum_{i=1}^N \mathbf{y}^i(t+\tau)$ експоненційно зростає із зростанням τ в одних точках досліджуваної траєкторії, то зменшується в інших.

Розглянемо еволюцію компоненти малого збурення $\mathbf{y}^i(t)$, спрямованої уздовж i -го власного вектора матриці. Стійкість траєкторії вздовж власного вектора $\mathbf{e}_i(t)$ визначається *характеристичним показником Ляпунова*

$$\lambda_i = \overline{\lim}_{t \rightarrow \infty} \frac{1}{t - t_0} \ln \left\| \frac{\mathbf{y}^i(t)}{\mathbf{y}^i(t_0)} \right\|, \quad (5.18)$$

де ліміт обмежений зверху, t_0 – початковий момент часу.

Таким чином, стійкість траєкторії в R^N визначається набором з N показників Ляпунова. Розташовані в порядку спадання $-\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$, вони утворюють так званий *спектр характеристичних показників Ляпунова* (спектр ЛХП) фазової траєкторії $\mathbf{x}^0(t)$.

З'ясуємо, як показники Ляпунова пов'язані з власними значеннями матриці лінеаризації $s_i(t)$. Розглянемо (5.17) в початковий момент часу $t=t_0$, припускаючи, що інтервал τ малий. Перейдемо в точку $\mathbf{x}(t_1)$, де $t_1=t_0+\tau$, і в якості початкового збурення візьмемо

$$\mathbf{y}^i(t_1) = \mathbf{y}^i(t_0) \exp [s_i(t_0)\tau]. \quad (5.19)$$

Оскільки τ мале, будемо вважати, що напрямок власних векторів \mathbf{e}_i майже не змінюється за час τ , і можна вважати, що вектор $\mathbf{y}^i(t_1)$ спрямований уздовж i -го власного вектора $\mathbf{e}_i(t_0)$. Вважаємо, що початкове збурення $\mathbf{y}^i(t_0)$ настільки мале, що воно залишається малим і в наступні моменти часу. Пересуваючись по кривій $\mathbf{x}^0(t)$ з малим кроком τ , отримуємо наближений вираз, що описує еволюцію малого збурення у напрямку i -го власного вектора:

$$\mathbf{y}^i(t) \approx \mathbf{y}^i(t_0) \exp \left[\sum_k s_i(t_k)\tau \right]. \quad (5.20)$$

Переходячи до границі $\|\mathbf{y}^i(t_0)\| \rightarrow 0$ і $\tau \rightarrow 0$, отримуємо сувору рівність

$$y^i(t) = y^i(t_0) \exp \left[\int_{t_0}^t s_i(t') dt' \right]. \quad (5.21)$$

В результаті підстановки (5.21) в (5.18) приходимо до рівності

$$\lambda_i = \overline{\lim}_{t \rightarrow \infty} \frac{1}{t - t_0} \int_{t_0}^t \operatorname{Re} s_i(t') dt'. \quad (5.22)$$

Таким чином, i -й показник Ляпунова λ_i можна розуміти як усереднену уздовж досліджуваної траєкторії дійсну частину власного значення s_i матриці лінеаризації $\hat{A}(t)$. Він показує, що відбувається з відповідною компонентою початкового збурення у середньому уздовж траєкторії.

Якщо часткове рішення $x^0(t)$ системи (5.11) є станом рівноваги, тобто $F(x^0)=0$, то матриця лінеаризації $\hat{A}(t)$ розраховується тільки в одній точці фазового простору і, отже, є матрицею з постійними елементами $a_{i,j}$. Власні вектори і власні значення матриці $\hat{A}(t)$ постійні у часі, а показники Ляпунова рівні дійсним частинам власних значень: $\lambda_i = \operatorname{Re} s_i$. Сигнатура спектра ЛХП показує, чи є стан рівноваги стійким чи ні. Для аналізу поведінки фазових траєкторій в локальній околиці стану рівноваги необхідно знати також і уявні частини власних значень матриці лінеаризації. На фазовій площині (випадок $N=2$) положення рівноваги характеризується двома власними значеннями матриці \hat{A} : s_1 і s_2 . Можливі такі випадки:

- 1) s_1 і s_2 є дійсними негативними числами. У цьому випадку стан рівноваги є стійким вузлом;
- 2) s_1 і s_2 – дійсні позитивні числа. Стан рівноваги є нестійким вузлом;
- 3) s_1 і s_2 – дійсні числа, але з різними знаками. Стан рівноваги у цьому випадку – сідло;
- 4) s_1 і s_2 – комплексно-зв'язані числа з $\operatorname{Re} s_{1,2} < 0$. Стан рівноваги – стійкий фокус;

5) s_1 і s_2 – комплексно-зв'язані з $\text{Re } s_{1,2} > 0$. Стан рівноваги – нестійкий фокус;

6) s_1 і s_2 – чисто уявні числа: $s_{1,2} = \pm j\omega$. Стан рівноваги у цьому випадку є центром.

На рис. 5.6 показана діаграма станів рівноваги, існуючих на фазовій площині при різних значеннях детермінанта і сліду матриці \hat{A} (відповідно, $\text{Det } \hat{A} = s_1 s_2$ і $\text{Sp } \hat{A} = s_1 + s_2$).

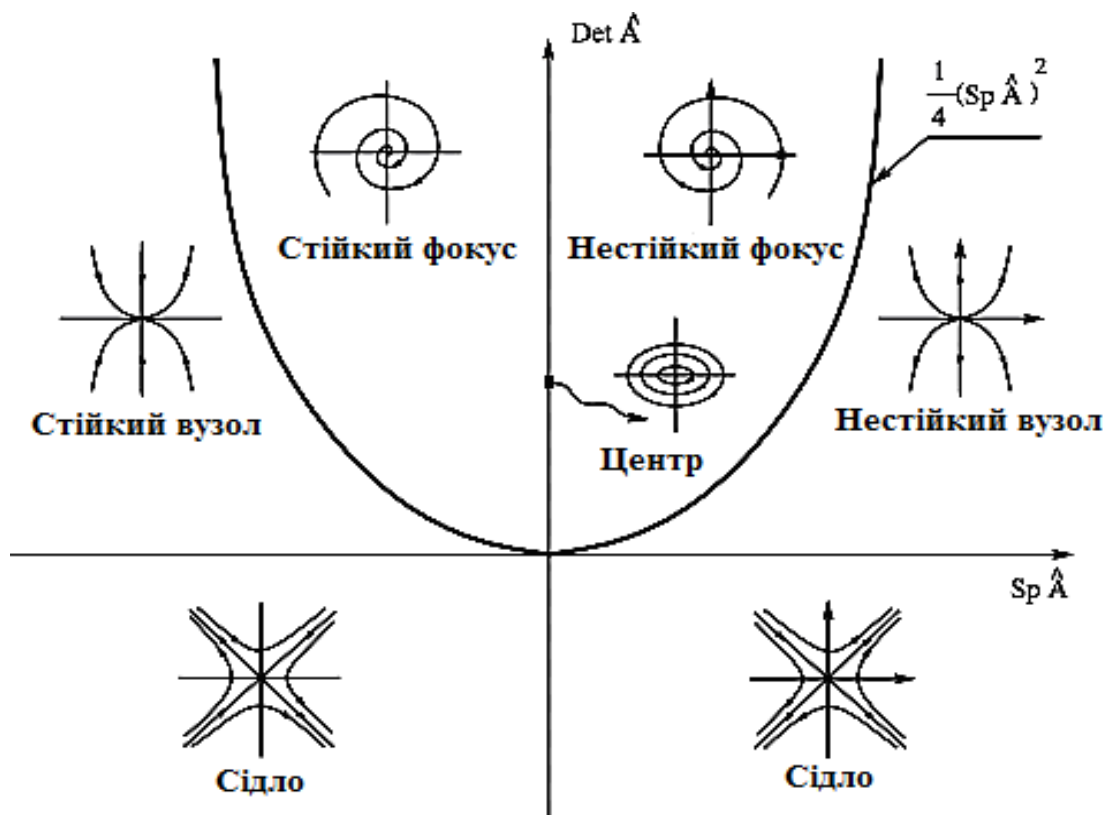


Рис. 5.6. Діаграма станів рівноваги на площині (фазові портрети показані в перетворених координатах)

Крім вищезазначених станів рівноваги, в просторі з розмірністю $N \geq 3$ можливі і інші типи станів рівноваги, наприклад, нестійкий по Ляпунову стан рівноваги, званий сідло-фокусом. На рис. 5.7 показані два варіанти стану

рівноваги сідло-фокусного типу, які реалізуються в R^3 . Вони розрізняються розмірностями їх стійких і нестійких різноманіть.

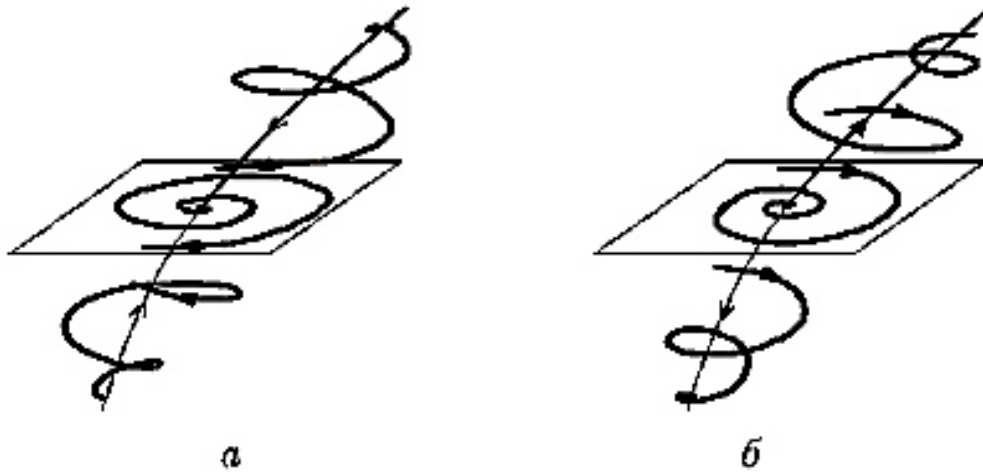


Рис. 5.7. Сідло-фокуси в тривимірному фазовому просторі:
 $a - s_1 -$ дійсно і негативно, $s_{2,3} -$ комплексно-зв'язані $\text{Re } s_{2,3} > 0$;
 $b - s_1 -$ дійсно і позитивно, $s_{2,3} -$ комплексно-зв'язані $\text{Re } s_{2,3} < 0$

Знаючи показники Ляпунова, неважко визначити, до якого типу граничних множин належить досліджуваний стан рівноваги. Положення рівноваги є атрактором, якщо воно асимптотично стійке за всіма напрямками, тобто його спектр ЛХП складається тільки з негативних показників (стійкий вузол або фокус). Якщо стан рівноваги нестійкий у всіх напрямках, то він є репелером (нестійкий вузол або фокус). Якщо спектр ЛХП включає як позитивні, так і негативні показники, то стан рівноваги належить до сідлового типу (просте сідло або сідло-фокус). Крім того, число показників $\lambda_i \geq 0$ і $\lambda_j \leq 0$ визначає розмірність нестійкого і стійкого різноманіть.

Розглянемо тут ще стійкість періодичних рішень. Будь-яке періодичне рішення $\mathbf{x}^0(t)$ системи (5.11) задовольняє умові

$$\mathbf{x}^0(t) \equiv \mathbf{x}^0(t + T), \quad (5.23)$$

де T – період рішення. Матриця лінеаризації $\hat{A}(t)$, що обчислюється в точках траєкторії, відповідної періодичному рішенню $\mathbf{x}^0(t)$, також є періодичною:

$$\widehat{A}(t) = \widehat{A}(t + T). \quad (5.24)$$

У цьому випадку рівняння для збурень (5.13) являє собою лінійне рівняння з періодичними коефіцієнтами. Стійкість періодичного рішення можна оцінити, визначивши, як мале збурення $y(t_0)$ змінюється за період T . Його еволюція може бути представлена наступним чином:

$$y(t_0 + T) = \widehat{M}_T y(t_0), \quad (5.25)$$

де \widehat{M}_T – постійна матриця, звана *матрицею монодромії*. Власні значення матриці монодромії, тобто коріння характеристичного рівняння

$$\text{Det}[\widehat{M}_T - \mu \widehat{E}] = 0, \quad (5.26)$$

називаються *мультиплікаторами* періодичного рішення $x^0(t)$. Мультиплікатори визначають стійкість періодичного рішення. Дійсно, дія оператора монодромії (5.25) за період T зводиться до наступного: компоненти розкладання первісного збурення за власними векторами матриці $\widehat{A}(t_0)$ помножуються на відповідні мультиплікатори μ_i . Таким чином, для того щоб періодичне рішення $x^0(t)$ було стійке по Ляпунову, необхідно і достатньо, щоб його мультиплікатори задовольняли вимозі $|\mu_i| \leq 1$, $i=1,2,\dots,N$. В усякому разі, один з мультиплікаторів завжди дорівнює $+1$. Як власні значення матриці монодромії, мультиплікатори задовольняють умовам:

$$\sum_{i=1}^N \mu_i = \text{Sp} \widehat{M}_T, \quad \prod_{i=1}^N \mu_i = \text{Det} \widehat{M}_T. \quad (5.27)$$

Мультиплікатори пов'язані із показниками Ляпунова для періодичного рішення наступним чином:

$$\lambda_i = \frac{1}{T} \ln |\mu_i|. \quad (5.28)$$

Нульовий показник в спектрі ЛХП граничного циклу відповідає мультиплікатору, рівному 1. Граничний цикл є атрактором, якщо всі інші показники негативні, а відповідні мультиплікатори по абсолютній величині менше 1. Якщо спектр ЛХП включає показники різного знака, то граничний

цикл є сідловим. Розмірність нестійкого різноманіття сідлового циклу дорівнює числу невід'ємних показників в спектрі ЛХП, а розмірність його стійкого різноманіття дорівнює числу показників, для яких $\lambda_i \leq 0$. Якщо все $\lambda_i > 0$, то граничний цикл є абсолютно нестійким (репелером).

5.2. Біфуркації динамічних систем

Багато фізичних задач при їх математичному описі приводять до диференціальних рівнянь, які залежать від одного або декількох параметрів.

Розглянемо як приклад рівняння коливань звичайного маятника або (що з математичної точки зору повністю ідентично) паралельного *RLC*-контур:

$$\ddot{x} + \alpha \dot{x} + \omega_0^2 x = 0, \quad \omega_0 = 1. \quad (5.29)$$

Рівняння (5.29) містить два параметри: α – параметр загасання, що характеризує тертя, і ω_0 – параметр, що визначає частоту коливань. Якщо втрати енергії відсутні, параметр загасання $\alpha=0$, то рішенням рівняння (5.29) будуть гармонійні незгасаючі коливання. При малому терті $0 < \alpha < 1$ рух системи буде коливальним з амплітудою, яка зменшується в часі за експоненціальним законом. Нарешті, при досить великому терті ($\alpha > 1$) рух маятника стає аперіодичним, загасаючим у часі. Уже в цьому простому прикладі виділяються два особливих значення параметра $\alpha=0$ і $\alpha=1$, відхилення від яких якісно змінюють властивості системи.

Керуючи параметрами ДС (наприклад, коефіцієнтом α в диференціальному рівнянні (5.29)), ми можемо спостерігати якісну зміну фазового портрета. Таке явище називають *біфуркацією* (від слова роздвоєння) ДС, а значення параметрів, при яких має місце біфуркація, називаються *біфуркаційними*. Під якісною зміною фазового портрета ми розуміємо зміну його структури, що складається в появі або зникненні граничних множин, а також у зміні характеру стійкості траєкторій, що належать граничним множинам. Наприклад, якщо у фазовому просторі системи існує стійка точка

рівноваги, а при зміні одного або декількох керуючих параметрів вона стає нестійкою, то докорінно змінюється поведінка всіх фазових траєкторій системи з області тяжіння цієї точки рівноваги. Тепер вони будуть притягатися до іншого атрактора, тобто в системі відбувається біфуркація.

В математиці і фізиці існує поняття *грубості* або *структурної стійкості*. Суть цього поняття у тому, що при малій зміні параметра груба система хоч і змінює в деталях режим функціонування, але не принципово. Можна сказати, що якісної перебудови фазового портрета не відбувається. З цієї точки зору для грубих систем перехід через точку біфуркації означає зміну одного структурно стійкого режиму на інший. При цьому в точці біфуркації система не є грубою: мала зміна параметра в ту чи іншу сторону призводить до різких змін стану.

Аналіз біфуркацій ДС при зміні параметрів системи дозволяє побудувати *біфуркаційну діаграму* системи. Біфуркаційна діаграма являє собою набір точок, ліній і поверхонь в просторі параметрів, які відповідають різним біфуркаціям граничних множин системи. Якщо при деяких значеннях параметрів існують кілька граничних множин, то біфуркаційна діаграма є «багатолистою». Співіснування великого (навіть нескінченного) числа граничних множин типово для систем зі складною динамікою. У багатовимірному просторі параметрів системи біфуркації характеризуються деякою кількістю умов, що накладаються на параметри системи. Число таких умов визначає корозмірність біфуркації. Наприклад, корозмірність 1 означає, що є тільки одна біфуркаційна умова.

5.2.1. «М'які» і «жорсткі» біфуркації. Катастрофи.

Незважаючи на багаторічну історію існування і розвитку класичної теорії стійкості і біфуркацій, настав момент, коли до цієї теорії була привернута загальна увага. Причиною цього послужили популярно викладені версії робіт французького математика Рене Тома про так звану теорію катастроф. Теорія катастроф на початку 1970-х рр. стала модною, зрозумілою для неспеціалістів

та універсальністю своїх претензій стала нагадувати псевдонаукові теорії минулих часів.

Суть справи полягає у тому, що мова йде про ті ж біфуркації, але при цьому обирається один з типів – так звані жорсткі біфуркації або кризи. Для пояснення розглянемо два простих приклади. У першому випадку (рис. 5.8,*а,б*) у результаті біфуркації початковий стаціонарний стан втрачає стійкість і народжуються два нових стійких стаціонарних стани. При цьому два стаціонарні стани, які знову з'явилися (рис. 5.8,*в*) розташовані в безпосередній близькості від вихідного стану, що втратило стійкість (позначено зірочкою). Біфуркації такого типу називають м'якими, маючи на увазі те, що режим функціонування системи, який знов народився, як би з'являється із режиму, який втратив стійкість, і співіснує поруч із ним.

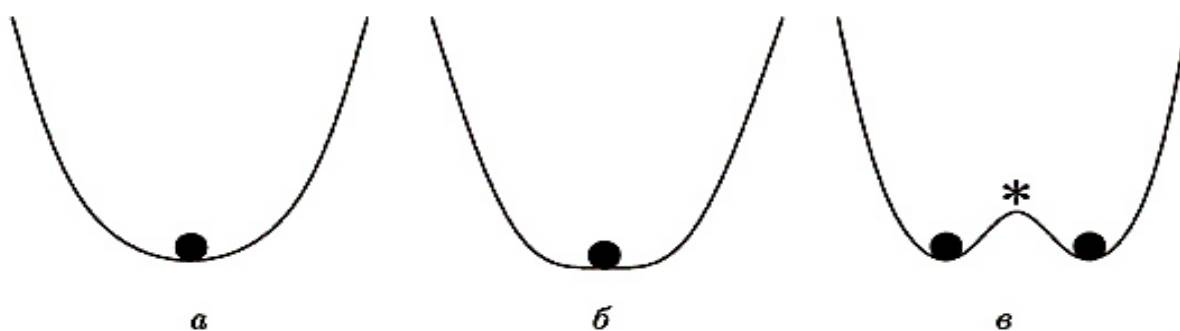


Рис. 5.8. Приклад м'якою біфуркації. Стаціонарний стан (*а*) втрачає стійкість (*б*) і поблизу нього з'являються два нових стійких стаціонарних стани (*в*)

Інший приклад біфуркації якісно представлений на рис. 5.9. При значенні параметра $\alpha < \alpha^*$ (рис. 5.9,*а*) кулька знаходиться у стійкому стаціонарному стані. При цьому існує ще один, нестійкий стан (помічено зірочкою).

У точці біфуркації $\alpha = \alpha^*$ стійкий і нестійкий стани зливаються в один (рис. 5.9,*б*). Далі вони зникають і система обирає новий режим (наприклад, як це показано на рис. 5.9,*в*), який істотно відрізняється від попереднього і не знаходиться в безпосередній близькості від початкового режиму. Такий тип

біфуркацій називають жорстким і саме жорсткі біфуркації були предметом аналізу в теорії катастроф.

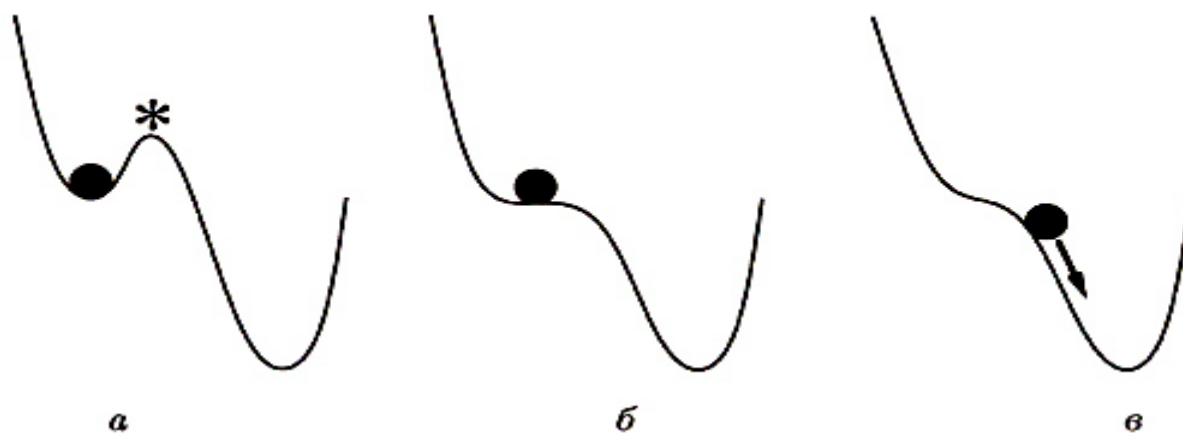


Рис. 5.9. Жорстка втрата стійкості стаціонарним станом, катастрофа

5.2.2. Біфуркація «триразова рівновага».

Розглянута вище біфуркація «дворазова рівновага» відноситься до так званих біфуркацій корозмірності 1. Біфуркація «триразова рівновага», яку ми збираємося тепер проаналізувати, відноситься до біфуркації корозмірності 2 і вимагає керування двома параметрами. Ця біфуркація складається в злитті трьох станів рівноваги: двох вузлів Q_1 , Q_2 і сідла Q_0 , розташованого між ними. В результаті залишається один стійкий вузол в точці Q_0 (рис. 5.10).

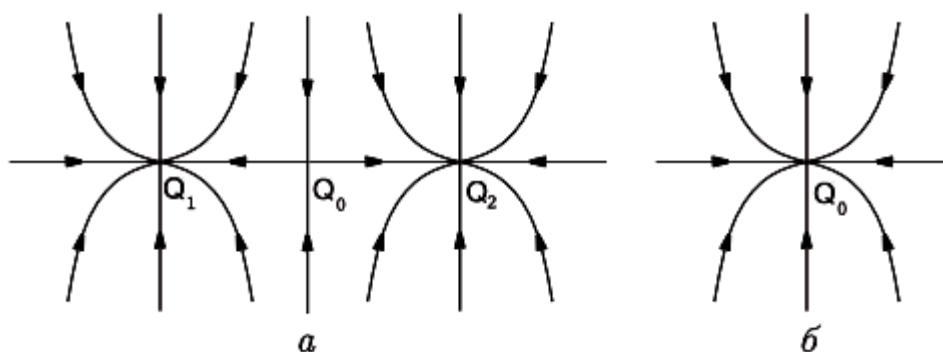


Рис. 5.10. Ілюстрація біфуркації «триразова рівновага»:

a – два стійких вузла і сідло до біфуркації;

б – один стійкий вузол після біфуркації

Модельна система для такої біфуркації може бути записана у вигляді

$$\dot{x} = \alpha_1 + \alpha_2 x + x^3. \quad (5.30)$$

Аналіз станів рівноваги показує, що при $\alpha_2 > 0$ незалежно від значення $\alpha_1 > 0$ у системи є єдиний стан рівноваги Q_0 з власним значенням $s_{Q_0} < 0$, тобто воно є асимптотично стійким. При $\alpha_2 < 0$ існує область значень параметра α_1 (заштрихована область на біфуркаційній діаграмі, зображеній на рис. 5.11,а), в якій система має три стани рівноваги, Q_0 , Q_1 і Q_2 .

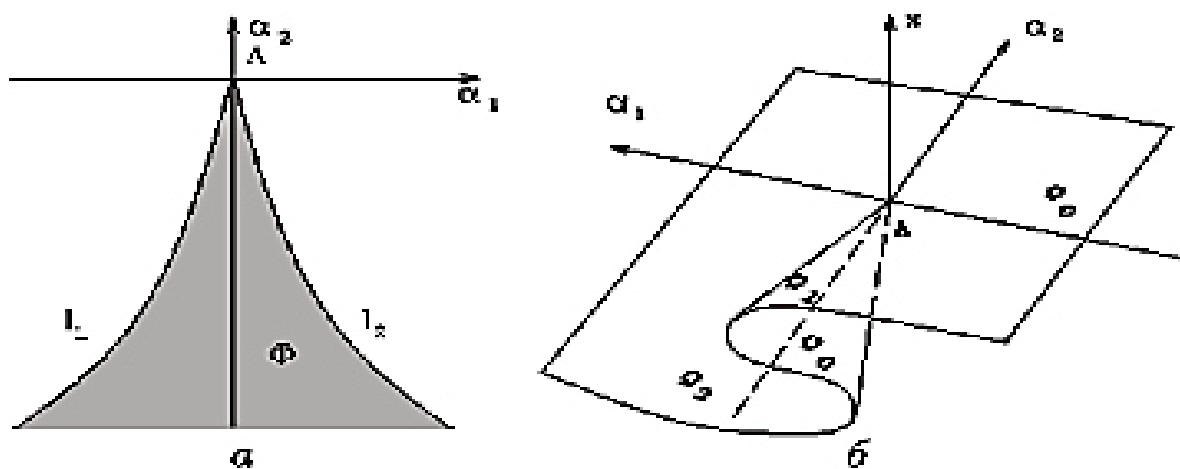


Рис. 5.11. Ілюстрація біфуркації триразової рівноваги:

а – біфуркаційна діаграма; б – фазо-параметрична діаграма

Один з них, Q_0 , є нестійким з $s_{Q_0} > 0$, а два інших, Q_1 і Q_2 , стійкі з $s_{Q_{1,2}} \leq 0$. Область бістабільності на біфуркаційній діаграмі (див. рис. 5.11) обмежена лініями l_1 і l_2 , які відповідають сідло-вузловим біфуркаціям вузлів $Q_{1,2}$ з сідлом Q_0 . Лінії l_1 і l_2 сходяться до точки A ($\alpha_1 = \alpha_2 = 0$), званою *точкою зборки*, або *каспом*. В цій точці одночасно виконуються дві біфуркаційних умови: $s_{Q_1}(\alpha_1, \alpha_2) = 0$ і $s_{Q_2}(\alpha_1, \alpha_2) = 0$. Тому біфуркація триразової рівноваги має корозмірність 2. У фазо-параметричному просторі системи (5.30) має місце структура, звана зборкою (див. рис. 5.11,б). В області збірки верхній і нижній листи біфуркаційної діаграми відповідають стійким станам рівноваги, а центральний – нестійкому.

5.2.3. Біфуркація Андронова–Хопфа.

В ДС з розмірністю $N \geq 2$ можлива ситуація, коли пара комплексно-сполучених власних значень точки рівноваги типу «стійкий фокус» перетинає уявну вісь. Це означає, що виконана біфуркаційна умова $\text{Re}s_{1,2}=0$. Нехай при цьому $\text{Im}s_{1,2} \neq 0$. Цей випадок відповідає біфуркації Андронова-Хопфа, інакше званої біфуркацією народження (зникнення) граничного циклу. Така біфуркація була вперше досліджена А.А. Андроновим для випадку $N=2$ і далі узагальнена Е. Хопфом на системи з довільним числом розмірності N . Існують два різних види біфуркацій Андронова-Хопфа: *суперкритична* біфуркація і *субкритична*, або *жорстка* біфуркація. Суперкритична біфуркація є м'якою, а субкритична біфуркація відповідає кризі атрактора. Біфуркація Андронова-Хопфа визначається єдиною біфуркаційною умовою і тому має корозмірність 1.

Суперкритична біфуркація Андронова-Хопфа показана на рис. 5.12,а-в і полягає у наступному. При $\alpha < \alpha^*$ існує стійкий фокус F , який в точці біфуркації $\alpha = \alpha^*$ перетворюється в центр і має пару чисто уявних власних значень $s_{1,2} = \pm j\omega_0$. При $\alpha > \alpha^*$ фокус F стає нестійким ($\text{Re}s_{1,2} > 0$), і поблизу нього народжується стійкий граничний цикл C_0 .

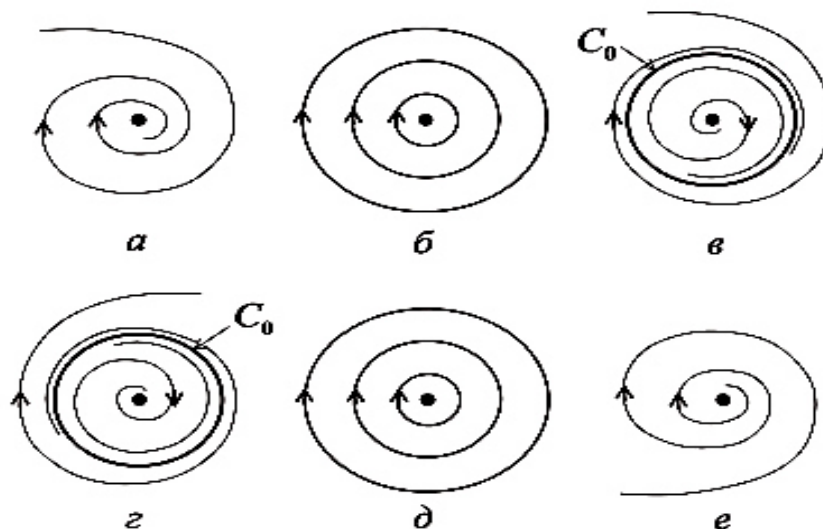


Рис. 5.12. Біфуркації Андронова–Хопфа:
а-в – суперкритична; г-е – субкритична

Субкритична біфуркація Андронова-Хопфа відбувається, коли при $\alpha=\alpha^*$ нестійкий (у загальному випадку для $N>2$ сідловий) граничний цикл C_0 «стягується» в точку фокусу F , який був стійким при $\alpha<\alpha^*$. У результаті цикл зникає, а фокус стає нестійким (рис. 5.12, $z-e$).

Модельна система для біфуркації Андронова-Хопфа має вигляд:

$$\dot{a} = (\alpha + j\omega_0)a + L_1 a |a|^2, \quad \omega_0 \neq 0, L_1 \neq 0, \quad (5.31)$$

де a – миттєва комплексна амплітуда. Величина L_1 називається *першою ляпуновською величиною* стану рівноваги. Якщо $L_1<0$, біфуркація є суперкритичною. Якщо $L_1>0$, то біфуркація – субкритична (характер біфуркації в особливому (виродженому) випадку $L_1=0$ потребує додаткового аналізу з урахуванням вищих ступенів a). Для дійсної миттєвої амплітуди і миттєвої фази коливань з (5.31) отримуємо:

$$\dot{A} = \alpha A + L_1 A^3, \quad \dot{\Phi} = \omega_0, \quad (5.32)$$

де $A=|a|$ і $\Phi=\text{Arg}(a)$. З рівняння для стаціонарної амплітуди $\alpha A + L_1 A^3=0$ отримуємо значення, відповідні фокусу ($A_F=0$) і граничному циклу ($A_0 = \sqrt{-\alpha / L_1}$). Граничний цикл існує за умови, що $\alpha/L_1<0$. Величина ω_0 визначає його період $T=2\pi/\omega_0$. Аналіз лінеаризованого рівняння для збурення амплітуди дозволяє знайти власні значення для рішень $A=A_F$ і $A=A_0$: $s_{F,0}=\alpha+3L_1A_{F,0}^2$. Звідси видно, що для $L_1<0$ цикл існує і стійкий при $\alpha>0$, а фокус стійкий при $\alpha<0$ і нестійкий при $\alpha>0$. У випадку $L_1>0$ при $\alpha<0$ існують нестійкий цикл і стійкий фокус, тоді як при $\alpha>0$ – тільки нестійкий фокус.

5.2.4. Біфуркації граничних циклів.

Розглянемо локальні біфуркації корозмірності 1 невивродженого граничного циклу. Відкинемо одиничний мультиплікатор і розташуємо мультиплікатори, які залишилися, в порядку убутання абсолютних значень. В цьому випадку біфуркації граничного циклу пов'язані з одним дійсним або двома комплексно-сполученими старшими мультиплікаторами $\mu_{1,2}$. Оскільки біфуркація корозмірності 1 передбачає тільки одну біфуркаційну умову, що

відповідна рівності $|\mu_1|=1$, то можливі лише три різних типи біфуркацій: $\mu_1(\alpha^*)=+1$, $\mu_1(\alpha^*)=-1$, і $\mu_{1,2}(\alpha^*)=\exp(\pm j\varphi)$, де α^* – біфуркаційне значення параметра. Для аналізу біфуркацій граничного циклу доцільно використовувати перетин Пуанкаре. Нерухомі точки в відображенні послідування характеризуються тими ж самими мультиплікаторами, що й вихідний граничний цикл, а перехід до перетину робить аналіз більш зручним.

Сідло-вузлова біфуркація. При досягненні параметром α біфуркаційного значення $\alpha=\alpha^*$ мультиплікатор μ_1 стійкого циклу стає рівним $+1$. Рис. 5.13 ілюструє цю біфуркацію для випадку тривимірного фазового простору ($N=3$). При $\alpha<\alpha^*$ існують два граничних циклу: стійкий цикл C_1 і сідловий цикл C_2 (див. рис. 5.13,а).

Їм відповідають стійка Q_1 і нестійка Q_2 нерухомі точки в перетині Пуанкаре. Умова $\mu_1=1$ визначає біфуркацію, подібну сідло-вузловій біфуркації станів рівноваги, розглянутій вище. У біфуркаційній точці $\alpha=\alpha^*$ відбувається злиття циклів C_1 і C_2 , в результаті чого виникає негруба замкнута траєкторія C типу сідло-вузол (див. рис. 5.13,б), яка зникає при $\alpha>\alpha^*$. Зміна параметра α в зворотному напрямку призводить до народження пари циклів C_1 і C_2 із згущення фазових траєкторій.

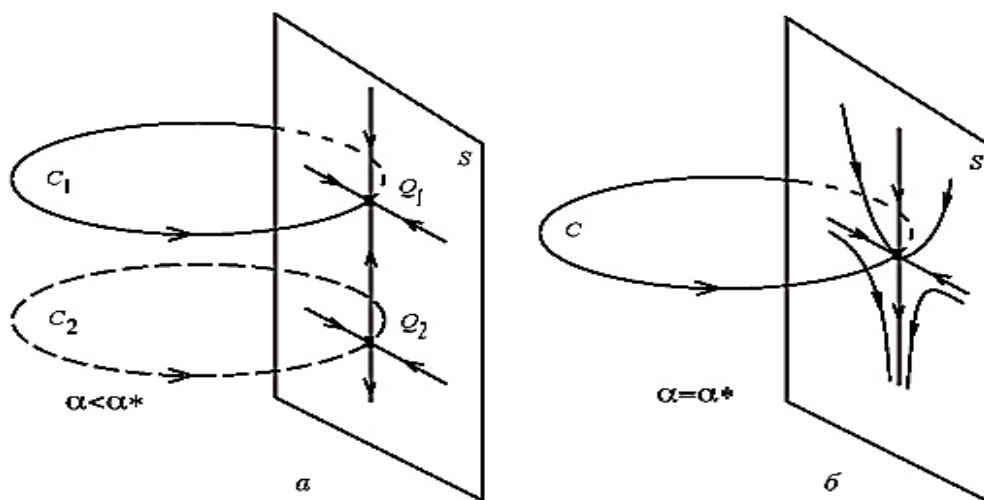


Рис. 5.13. Сідло-вузлова біфуркація граничних циклів:

а – до біфуркації; б – після біфуркації

Біфуркація подвоєння періоду. В біфуркаційній точці $\alpha=\alpha^*$ мультиплікатор $\mu_1(\alpha^*)$ стає рівним -1 , причому $d\mu/d\alpha|_{\alpha^*} \neq 0$. Біфуркація, що визначається такою умовою, називається *біфуркацією подвоєння періоду*. Ця біфуркація може бути суперкритичною (внутрішньою) або субкритичною (кризою). Суперкритична біфуркація подвоєння відбувається наступним чином: нехай при $\alpha < \alpha^*$ існує стійкий граничний цикл C_0 з періодом T_0 . При $\alpha > \alpha^*$ цикл C_0 стає сідловим, і в його околиці народжується стійкий граничний цикл C з періодом T , близьким до подвоєного T_0 ($T \approx 2T_0$). Фазові траєкторії C_0 і C , а також їх перетин Пуанкаре поблизу точки біфуркації зображені на рис. 5.14,а.

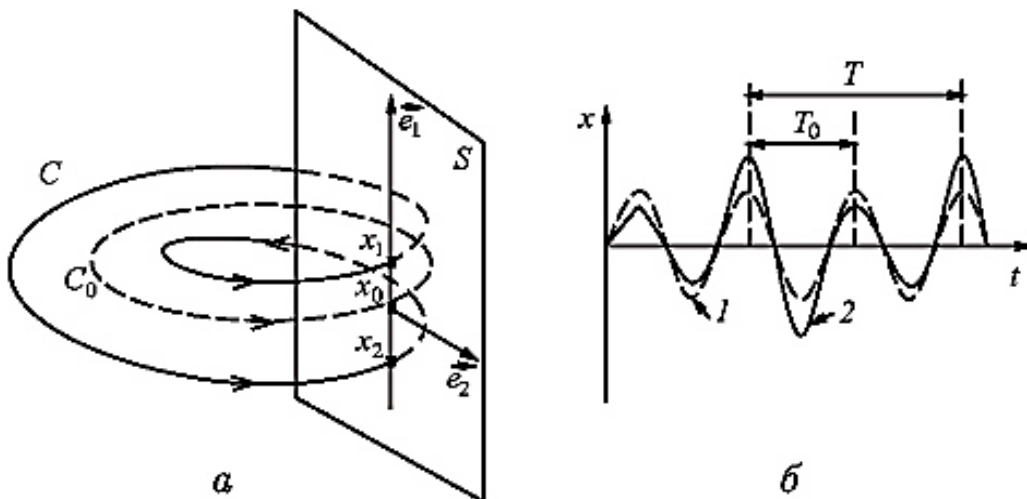


Рис. 5.14. Суперкритична біфуркація подвоєння періоду:

а – цикли C_0 і C і їх перетин Пуанкаре;

б – форма коливань до (крива 1) і після (крива 2) біфуркації

Рис. 5.14,б показує, як змінюється форма коливань однієї з динамічних змінних при проходженні точки біфуркації.

Коли відбувається субкритична біфуркація подвоєння періоду, стійкий цикл C_0 і сідловий цикл C з подвоєним періодом, існуючий при $\alpha < \alpha^*$, зливаються в точці біфуркації, після чого в фазовому просторі залишається тільки цикл C_0 , що став сідловим.

Біфуркація народження (зникнення) двовимірного тора (біфуркація Неймарка). Ця біфуркація (вона також називається біфуркацією Неймарка-Сакера) відбувається, коли пара комплексно-сполучених мультиплікаторів граничного циклу виходить на одиничну окружність. В біфуркаційній точці $\alpha=\alpha^*$ має місце наступне співвідношення: $\mu_{1,2}(\alpha^*)=\exp(\pm j\varphi)$, де $\varphi \in [0, 2\pi]$ і $\varphi(\alpha^*) \neq 0, \pi/2, \pi/3$ (виключені так звані *сильні резонанси*). Дана біфуркація також може бути суперкритичною (внутрішньою) і субкритичною (кризою). В залежності від характеру біфуркації можуть виникати різні ситуації. У разі суперкритичної біфуркації зі стійкого граничного циклу C_0 народжується стійкий двовимірний тор T^2 . Цикл C_0 в результаті біфуркації втрачає стійкість і стає сідловим. Субкритична біфуркація має місце, коли нестійкий (сідловий) тор T^2 «стягується» до стійкого циклу C_0 , який в цей момент втрачає стійкість. Народження тора з граничного циклу зображено на рис. 5.15,а.

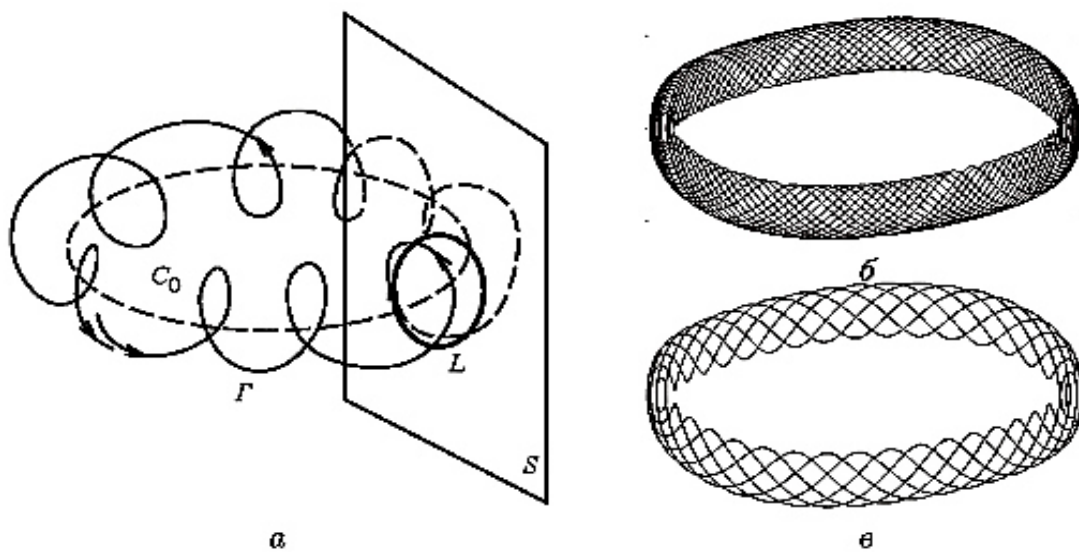


Рис. 5.15. Біфуркація народження тора з граничного циклу C_0 :

а – траєкторія Γ на торі в околиці нестійкого циклу C_0 ;

б – ергодичний тор; в – резонанс на торі

Поблизу точки біфуркації $\alpha=\alpha^*$ вектор малого збурення циклу C_0 обертається вздовж траєкторії C_0 . У той же час величина збурення залишається

незмінною, оскільки виконується умова $|\mu_{1,2}(\alpha^*)|=1$. Таким чином, зображаючи точку в перетині Пуанкаре рухається вздовж замкнутої кривої L , званої *інваріантною замкнутою кривою*. Величина $\theta(\alpha)=\varphi/2\pi$ називається числом обертання на торі T^2 (або на відповідній інваріантній замкнутій кривій). Якщо число обертання $\theta(\alpha^*)$ приймає ірраціональне значення, будь-яка траєкторія C на торі незамкнута, і тор, який виник, є *ергодичним* (рис. 5.15,б).

Якщо $\theta(\alpha^*)=p/q$, де p і q – будь-які позитивні цілі числа, то говорять, що на торі має місце резонанс порядку p/q . Траєкторія замикається, утворюючи граничний цикл, що лежить на поверхні тора (рис. 5.15,в).

Біфуркації порушення симетрії. Біфуркації граничних циклів, що задаються умовами $\mu_1(\alpha^*)=\pm 1$ або $\mu_{1,2}(\alpha^*)=\exp(\pm j\varphi)$, можуть призвести до ситуації, коли граничний цикл втрачає свою симетрію. Такі біфуркації типові, наприклад, для систем, що складаються з двох або більше ідентичних підсистем. Властивість симетрії пов'язана з існуванням в фазовому просторі системи деякого інваріантного різноманіття U . Як приклад розглянемо випадок двох пов'язаних ідентичних підсистем:

$$\dot{x} = F(x, \alpha) + \gamma g(y, x); \quad (5.33)$$

$$\dot{y} = F(y, \alpha) + \gamma g(x, y); \quad (5.34)$$

де $x, y \in \mathbf{R}^N$ – вектори стану підсистем, α – керуючий параметр, γ – параметр зв'язку. Функція g задає функцію зв'язку між підсистемами, причому $g(x, x)=0$. У цьому випадку підпростір $x=y$ є інваріантним симетричним різноманіттям.

Нехай стійкий граничний цикл розташований в U , тобто є симетричним. Якщо який-небудь мультиплікатор симетричного циклу, що відповідає власному вектору, що не лежить в U , приймає біфуркаційне значення, то має місце біфуркація порушення симетрії. В результаті виникає несиметричний атрактор, який не розташований в U . Кажуть, що в результаті біфуркації атрактор втрачає симетрію. Біфуркації порушення симетрії, визначені умовами $\mu_1(\alpha^*)=-1$ або $\mu_{1,2}(\alpha^*)=\exp(\pm j\varphi)$ дуже схожі на аналогічні біфуркації в системі без симетрії. Біфуркація, визначена умовою $\mu_1(\alpha^*)=+1$, є особливим випадком,

коли симетричний цикл $C_0 \in U$ продовжує існувати, але стає сідловим. Це призводить до народження двох стійких циклів з тим же самим періодом. Вони не лежать в U , але є взаємно симетричними. Така біфуркація відома під назвою *біфуркації вил* (рис. 5.16).

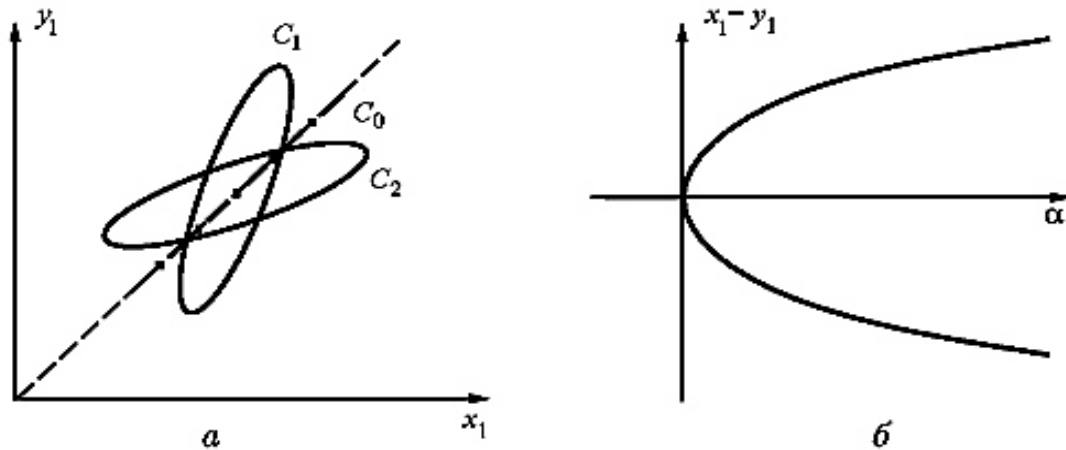


Рис. 5.16. Біфуркація вил в системі із симетрією:
 а – проекція граничних циклів після біфуркації;
 б – якісна ілюстрація фазо-параметричної діаграми

Ми розглянули локальні біфуркації станів рівноваги і граничних циклів. Різні біфуркації відбуваються також і з більш складними множинами (торами, хаотичними атрactorами). Однак їх вивчення часто засноване на експериментальних результатах. Теорія біфуркацій квазіперіодичних і хаотичних атрactorів ще не завершена і знаходиться в стані розвитку.

5.3. Детермінований хаос в динамічних системах

Вихід з фазової площини в простір більшої розмірності призводить до принципових якісних змін. Значно збільшується число можливих біфуркацій станів рівноваги і граничних циклів, багато з яких поки ще не досліджені. Стають можливими такі сідлові множини, як стан рівноваги сідло-фокусного типу, сідловий граничний цикл, а в просторі з розмірністю $N \geq 4$ – цикл сідло-

фокусного типу, сідловий тор тощо. Поява багатовимірних стійких і нестійких різноманіть сідлових множин і нових типів двояко асимптотичних траєкторій, таких як сепаратрисні петлі сідло-фокусів і гомоклінічні криві Пуанкаре, у багатьох випадках забезпечує складну структуру фазового портрета ДС. Реалізовані режими поведінки виявляються набагато складніші і різноманітніші. Крім періодичних коливань стають можливими квазіперіодичні і хаотичні коливання. Виникають нові типи атракторів – двовимірні і багатовимірні тори, відповідні квазіперіодичним режимам, дивні хаотичні атрактори, що є образами динамічного хаосу. При деяких умовах можливі особливі типи поведінки ДС і особливі «екзотичні» атрактори: дивні хаотичні і хаотичні недивні.

Найбільш важливим і цікавим типом поведінки систем в R^N при $N > 3$ є детермінований (динамічний) хаос.

Хаотичні процеси в детермінованих нелінійних дисипативних системах – одна з фундаментальних проблем сучасного природознавства, та є предметом пильної уваги дослідників. Переконаливо доведено, що в таких системах причина генерування складних коливальних процесів, які можуть не відрізнятися за фізичними характеристиками від істинно випадкових, криється не в великій кількості ступенів свободи і не в наявності флуктуацій, як раніше вважалося, а в експоненційній нестійкості режимів, що породжує вразливу залежність від точності завдання початкового стану системи. Можливість подібних явищ прекрасно розумів і передбачав А. Пуанкаре. У нестійких системах «...зовсім мізерна причина, що вислизає від нас по своїй малості, викликає значну дію, яку ми не можемо передбачити. ... Передбачення стає неможливим, ми маємо перед собою явище випадкове». Так писав він ще у 1908 р. в книзі «Наука і метод». Розвиток ідей А. Пуанкаре наразі призвів до створення фундаменту хаотичної динаміки детермінованих систем. Як виявилось, необхідною умовою виникнення хаосу в диференціальних системах є розмірність фазового простору $N > 3$ і збудження незатухаючих хаотичних

пульсацій стає принципово можливим в генераторах всього з півтора ступенями свободи.

5.3.1. Детермінованість і хаос.

Коли говорять про детермінованість, мають на увазі однозначний взаємозв'язок причини і слідства (див. п. 5.1). У застосуванні до еволюційних законів це означає, що якщо задано деякий початковий стан системи при $t=t_0$, то воно однозначно визначає стан системи в будь-який момент часу $t>t_0$. Наприклад, якщо тіло рухається рівно прискорено, то його швидкість визначається детермінованим законом:

$$v(t) = v(t_0) + at. \quad (5.35)$$

При завданні початкової швидкості $v(t_0)$ ми однозначно визначаємо значення швидкості $v(t)$ в будь-якій момент часу $t>t_0$.

У загальному випадку залежність майбутнього стану $x(t)$ від початкового $x(t_0)$ можна записати у вигляді: $x(t)=F[x(t_0)]$, де F – детермінований закон (або оператор), який здійснює строго однозначне перетворення початкового стану $x(t_0)$ в майбутній стан $x(t)$ для будь-якого $t>t_0$. Цей закон може являти собою функцію, диференціальне або інтегральне рівняння, просто деяке правило, задане таблицею або графіком тощо. Важливо головне: закон F *однозначно* трансформує початковий стан (причину) в майбутній стан (наслідок).

Тепер внесемо ясність в поняття «хаос». Давайте проведемо уявний експеримент з броунівською часткою. Помістимо частку в початковий момент $t=t_0$ в розчин рідини і за допомогою мікроскопа почнемо фіксувати її положення в часі, відзначаючи координати частки через рівні інтервали Δt . Неважко переконатися, що під дією випадкових поштовхів з боку оточуючих молекул частка буде здійснювати нерегулярні блукання, які характеризуються заплутаною траєкторією. Повторимо експеримент декілька раз поспіль, здійснюючи в межах можливостей відтворення початкових умов. Які будуть результати? Їх, головним чином, два. Перший – кожен раз траєкторія руху частки буде складною, неперіодичною. Другий – будь-яка спроба однозначного

повторення експерименту призведе до негативного результату. Кожен раз при повторенні експерименту з однаковими (в межах наших можливостей) початковими умовами ми будемо отримувати різні траєкторії руху частки, які навіть близько не нагадують одна одну!

Класичне явище руху броунівської частки дає нам чіткі фізичні уявлення про хаос як про непередбачуваний, випадковий процес. Таким чином, якщо ми говоримо про хаос, ми маємо на увазі, що зміна в часі стану системи є випадковим (його не можна однозначно передбачити) і таким, що не відтворюється (процес не можна повторити).

5.3.2. Стійкість і нелінійність.

Нам знадобиться розглянути поняття *стійкості* (*нестійкості*) руху системи. Почнемо з найпростішого, розглянувши стан спокою або рівноваги системи. Помістимо маленьку кульку в нижню точку всередині порожньої сфери. Злегка штовхнемо її і поспостерігаємо за рухом. Після скоєння кількох загасаючих коливань кулька знову займе положення на дні сфери. У цьому випадку положення рівноваги стійке: малі збурення початкового стану загасають у часі. Якщо ми помістимо кульку на вершину сфери (ззовні), то реакція на мале збурення буде іншою: при будь-якому як завгодно малому відхиленні кульки від стану рівноваги він скочується з вершини. Це положення рівноваги нестійке: малі збурення початкового стану наростають у часі.

Фізичний сенс поняття «*стійкість*» (*нестійкість*), розглянутий нами стосовно стану рівноваги, застосовується й до будь-якого іншого режиму. Режим функціонування ДС називають стійким, якщо малі збурення в околиці цього режиму загасають у часі, прагнучи до нуля. Якщо цього не відбувається і малі відхилення від режиму функціонування системи наростають у часі, такий режим буде нестійким (див. п. 5.1.4).

Тепер обговоримо іншу важливу властивість складних систем – *нелінійність*. Нехай ми маємо справу з нестійким режимом. Злегка порушивши режим малим впливом, ми спочатку будемо фіксувати наростання збурення. Чи

буде воно нескінченним? У реальному житті – ні. Відхилення буде наростати до тих пір, поки не вступить в дію якийсь механізм нелінійного обмеження процесу наростання збурення.

З фізичної точки зору наростання амплітуди не може відбуватися нескінченно. На першому етапі, коли відхилення від вихідного стану мале, воно може наростати. Далі, в силу обмеженості енергетичних ресурсів системи, це наростання має припинитися або змінитися зменшенням амплітуди відхилення. Будь-який новий режим повинен мати кінцеву амплітуду і управляють цими процесами нелінійні закони.

Ми говоримо про нелінійність в тому випадку, коли властивості системи безпосередньо залежать від її стану. Наведемо приклад. Нехай залежність амплітуди відхилення $f(x)$ від вихідного стану x визначається наступним співвідношенням:

$$f(x) = kx - bx^3, \quad (5.36)$$

де k і b – постійні позитивні коефіцієнти. Якщо $x \ll 1$, то $bx^3 \ll kx$ і

$$f(x) \cong kx. \quad (5.37)$$

У разі (5.37) $f(x)$ лінійно зростає з ростом x . Якщо ж x стає порівняним з одиницею, то членом bx^3 нехтувати вже не можна. У разі (5.36) зростання відхилення $f(x)$ за рахунок члена kx почне відчувати нелінійне обмеження в силу віднімання величини bx^3 . При деяких значеннях x величина відхилення (5.36) знову буде близька до нуля і все почнеться спочатку: відхилення почне наростати, досягне максимуму і потім, відчуваючи обмеження, знову зменшиться. Система буде як би автоматично себе регулювати, оскільки її властивості залежать від її поточного стану.

5.3.3. Нестійкість і нелінійне обмеження.

Розглянемо нестійку детерміновану систему з урахуванням дії механізму нелінійного обмеження наростань збурень. Для простоти розглянемо стан рівноваги, якому відповідає точка в просторі фазових координат системи. Виведемо систему з рівноваги малим відхиленням. Це збурення почне

наростати в силу нестійкості. Далі наростання збурення почне сповільнюватися (вступить в силу механізм нелінійного обмеження). Що можна очікувати в цій ситуації? По-перше: в силу нелінійного обмеження відхилення зменшиться строго до нуля. Система повернеться в початковий стан рівноваги. Теоретично це можливо, однак дуже мало ймовірно, тому що початковий стан рівноваги нестійкий. Більш вірогідною є інша ситуація: система повернеться в малу околицю вихідного стану (підійде дуже близько до стану нестійкої рівноваги) і знову (в силу нестійкості) почне від нього віддалятися. Цей процес буде продовжуватися до нескінченності у часі. Але реалізація такого процесу вимагає деяких спеціальних умов.

Припустимо, що ми маємо справу з двовимірною диференціальною ДС. Простір її станів – фазова площина з координатами x і y . Якщо мале збурення стану рівноваги в такій системі буде наростати, а в результаті нелінійного обмеження далі зменшуватися, то можливі два варіанти: поява нових стійких станів рівноваги поблизу нестійкого, або вихід траєкторії на новий режим, який відповідає періодичним коливанням.

Другий варіант ілюструє рис. 5.17. При малих амплітудах збурення (див. рис. 5.17,*а*) траєкторія по спіралі віддаляється від точки рівноваги O . При великих відхиленнях (див. рис. 5.17,*б*) траєкторія повертається. В результаті замість стану рівноваги, що втратив стійкість, з'являється новий режим – періодичні автоколивання, яким відповідає граничний цикл Γ на фазовій площині.

Нестійкість стану рівноваги в двовимірній системі при наявності механізму нелінійного обмеження наростання збурень породжує новий режим – режим стійких періодичних коливань. Якщо ми уявимо собі іншу ситуацію, коли відхилення від стану рівноваги спочатку наростає, а потім в силу нелінійності знову прагне до нуля, ми прийдемо до протиріччя: фазова траєкторія зобов'язана буде само перетинатися (рис. 5.18).

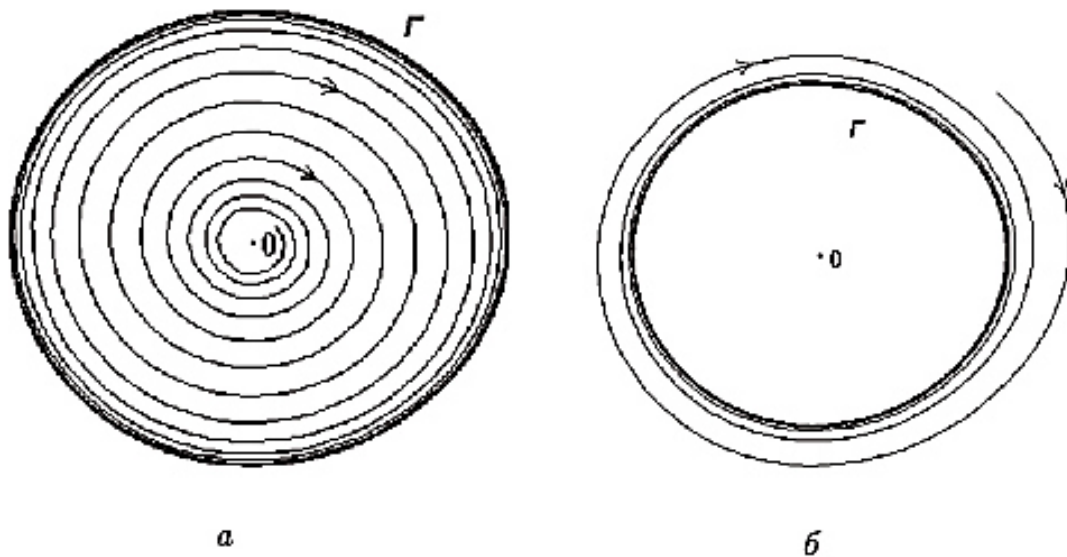


Рис. 5.17. Народження стійкого граничного циклу Γ в околиці нестійкої рівноваги O . Поведінка траєкторій: a – при малих відхиленнях від рівноваги; b – при великих

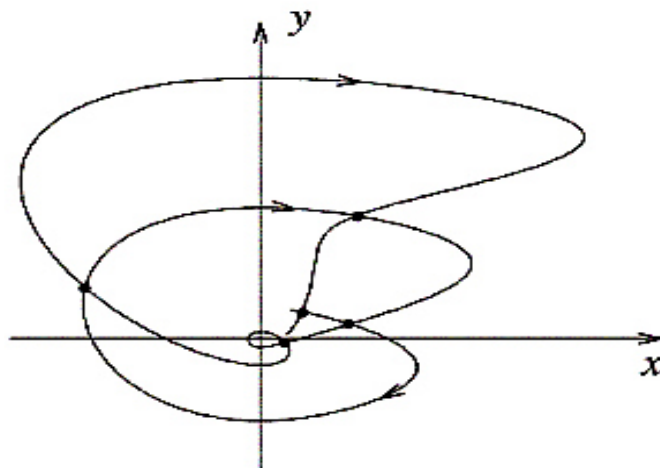


Рис. 5.18. Поведінка динамічної системи, яку неможливо реалізувати на площині в силу перетину фазових траєкторій (виходить шляхом проєкції тривимірної траєкторії на площину)

Але це буде означати, що існують початкові умови, що призводять в процесі еволюції до різних станів. Це неможливо в силу поняття детермінізму, яке в даному прикладі проявляється в змісті теореми єдності рішення: при заданих початкових умовах рішення існує і воно єдине, іншого не дано.

5.3.4. Детермінований хаос.

Картина принципово зміниться, якщо ми розглянемо ДС, стан якої характеризується трьома незалежними змінними (фазовими координатами). Іншими словами, давайте повторимо наші міркування, здійснивши вихід з площини в тривимірний фазовий простір. Ніщо не забороняє нам реалізувати ситуацію (див. рис. 5.18) в просторі трьох вимірів. Траєкторія розкручується в тривимірному просторі, віддаляючись від точки O по спіралі. Досягнувши деяких значень і відчуваючи дію механізму нелінійного обмеження, траєкторія знову повернеться в околицю початкового стану. Далі, з огляду на нестійкість, процес буде повторюватися.

Можливі два варіанти. Перший: траєкторія, зробивши декілька оборотів в тривимірному просторі, через деякий час замкнеться. Це буде означати наявність в системі складного, але періодичного процесу коливань. Другий варіант: траєкторія, хоча й буде повертатися в околицю нуля координат, але не буде замикатися, демонструючи якийсь аперіодичний процес, що триває нескінченно. Другий випадок і відповідає режиму детермінованого хаосу. Дійсно, працює основний принцип детермінізму: майбутнє однозначно визначено початковим станом. Однак, процес еволюції системи складний, неперіодичний. Чисто зовні він нічим не відрізняється від випадкового.

При більш детальному аналізі розкривається одна важлива відмінність цього процесу від випадкового: цей процес такий, що відтворюється. Дійсно, повторивши ще раз початковий стан, в силу детермінованості ми знову однозначно відтворимо ту ж саму траєкторію незалежно від ступеня її складності. Значить цей неперіодичний процес не є хаотичним в сенсі визначення хаосу, даного нами в п. 5.3.1. Так, це складний, схожий на випадковий, але тим не менш детермінований процес. Важливо тут те, що він характеризується нестабільністю і ця обставина дозволяє нам зрозуміти ще одну принципово важливу властивість систем з детермінованим хаосом – *перемішування*.

5.3.5. Перемішування.

У дисипативних системах, розмірність фазового простору яких $N \geq 3$, теоретично можливий режим складних неперіодичних пульсацій. Цей тип руху детермінований і характеризується нестійкістю.

Спочатку поговоримо про стійкі режими руху в детермінованих дисипативних ДС. Розглянемо в якості початкового стану не крапку x^0 з певними координатами в просторі станів, а малу сферу радіуса $\varepsilon > 0$, яка оточує цю точку. Будь-яка точка всередині сфери характеризує мале відхилення від x^0 . Сфера включає сукупність можливих відхилень від початкового стану, що не перевищують по модулю ε . Тепер застосуємо оператор еволюції і простежимо за трансформацією цієї сфери. В силу стійкості обраного нами режиму будь-яке мале відхилення в часі має затухати. Це означає, що під дією детермінованого закону еволюції кулька радіусу ε в часі буде зменшуватися і при $t \rightarrow \infty$ її радіус зменшиться до нуля (рис. 5.19).

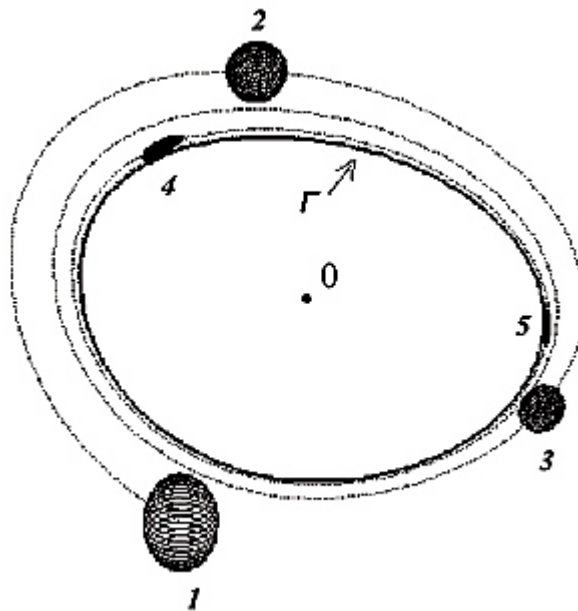


Рис. 5.19. Стиснення первісної області невизначеності 1 в часі в разі, коли цикл Γ є стійким граничним режимом

Вихідний фазовий обсяг в дисипативних системах в часі зменшується. Це означає, що малі збурення в результаті будуть затухати і система знову повернеться в початковий стійкий режим. А якщо вихідний режим нестійкий, то фазовий обсяг може збільшуватись до нескінченності, якщо нестійка система лінійна. Але якщо система нелінійна і дисипативна, то процес еволюції початкового малого фазового обсягу буде вельми нетривіальним.

Нестійкість режиму веде до зростання збурень. Це одна обставина. Друга – дисипативні системи незалежно від виду стійкості викликають зменшення елемента фазового обсягу в часі до нуля, що пов'язано із втратами енергії. Як поєднати ці два фактори? Існує єдине вирішення цієї дилеми: елемент фазового обсягу за деякими напрямками повинен розтягуватися, а за іншими – стискатися. Причому ступінь стискання в середньому повинна обов'язково превалювати над ступенем розширення, щоб у результаті фазовий обсяг в часі зменшувався. У нелінійних дисипативних системах це виявляється можливим. В силу наявності механізму нелінійного обмеження фазова траєкторія складного режиму коливань зосереджена в обмеженій області фазового простору. У цьому кожна мала околиця початкового стану еволюціонує і в підсумку переміщується по всій області, що зайнята траєкторією.

Проведемо уявний експеримент. У склянку з водою помістимо маленьку чайнку і розмішаємо воду чайною ложкою, викликавши нестійкість. Чайнка буде при цьому рухатися по складній спіралеподібній траєкторії, яка обумовлена рухом води в стакані. При цьому в будь-який заданий момент часу ми теоретично можемо зафіксувати її координати $x(t)$ в обсязі води. Тепер замість чайнки помістимо в стакан з водою дуже маленьку крапельку чорнила і знову розмішаємо воду чайною ложкою. Що при цьому відбудеться? Чорнила практично рівномірно розбіжаться по всьому обсягу води. Частинки чорнила, спочатку зосереджені в маленькому обсязі крапельки, через час перемішування можна буде виявити в будь-якій частині обсягу води у склянці. У житті цей процес ми звикли називати перемішуванням. У математиці це поняття також існує і, з точки зору фізичної інтерпретації, виявляється дуже близьким за

змістом. Дійсно, потік води в склянці, створений рухом чайної ложки, можна інтерпретувати як дію детермінованого еволюційного оператора динамічної системи. Чайнка при цьому буде рухатися по складній, але детермінованій (хоча й дуже складній) траєкторії. А крапелька чорнила, яку можна інтерпретувати як деякий маленький обсяг у фазовому просторі навколо чайнки, під дією оператора еволюції переміщається по всьому обсягу води.

5.3.6. Ймовірнісні властивості детермінованих систем.

У нестійких режимах в детермінованих нелінійних системах з перемішуванням ми можемо передбачити майбутній стан однозначно тільки у разі суворого завдання початкових умов. Однак, якщо врахувати скільки-завгодно малу, але кінцеву похибку (тобто розглянути крапельку чорнила замість чайнки), то детерміноване передбачення стає неможливим. Мала область первісної невизначеності розвивається за рахунок перемішування на кінцеву область у фазовому просторі. Тепер ми маємо справу з процесом, який асоціюється зі справжньою випадковістю, зі справжнім хаосом.

Основною властивістю ДС, що демонструють режим детермінованого хаосу, є чутлива залежність режиму функціонування до як завгодно малих змін початкових умов. Саме ця обставина веде до втрати детермінованої передбачуваності і необхідності вводити ймовірнісні характеристики для опису динаміки таких систем. У цьому сенсі стає зрозумілим термін «детермінований хаос», який характеризує народження випадкової, непередбачуваної поведінки системи, що керується детермінованими законами.

Невизначеність в завданні початкового стану – ситуація цілком реальна з точки зору фізики. Дійсно, в силу кінцевої точності реєстрації стану будь-якими приладами, він визначається з кінцевою (нехай як завгодно малою) похибкою. Це означає, що ми повинні аналізувати еволюцію у часі не початкової точки, а початкової області навколо цієї точки. В силу перемішування ми зіткнемося з процесом, детально описаним вище.

5.3.7. Детермінований хаос – властивість матеріального світу.

Шляхом простих міркувань ми прийшли до висновку про можливість режиму детермінованого хаосу в нелінійних системах з дисипацією енергії. У сучасній науці цей ефект строго обґрунтований теоретично і достовірно підтверджений експериментально. Після відкриття детермінованого хаосу, ясного розуміння властивостей ефекту і розробки методів його діагностики, хаос було виявлено практично в усіх областях сучасного природознавства: у фізиці, радіотехніці, хімії, біології, механіці, економіці та інших.

Хоча теоретично переважна кількість реальних матеріальних систем і процесів нелінійні, існує широкий клас процесів, досить коректно описуваних в лінійному або квазілінійному наближенні. Лінійна теорія ДС і процесів розроблена досить повно і дозволяє дати їх вичерпний опис, який добре узгоджується з експериментом. Але детермінований хаос – явище, притаманне виключно нелінійним системам. А щодо нелінійної теорії справи йдуть набагато гірше. Поки не існує, наприклад, загальної теорії розв'язання нелінійних диференціальних рівнянь. Аналіз динаміки нелінійних систем і наразі вимагає творчого підходу, індивідуального у кожному конкретному випадку.

Саме відсутність строгих теоретичних результатів стосовно нелінійних систем стримувало відкриття і розуміння цього універсального явища. Експериментатори давно стикалися із проявом хаосу. Однак обмеженість теоретичних знань, обумовлена впливом лінійної і квазілінійної структури наукового мислення, призводила до похибок в трактуванні спостережуваних результатів. Робився висновок про те, що шумоподібні коливання обумовлені або дією флуктуацій, або великою кількістю ступенів свободи системи, або несправністю вимірювальної апаратури.

Зараз ситуація змінилася. Наше життя все більше настійно вимагає кількісного обліку таких чинників, як надвисока щільність, надвисока температура, тиск, надвисокі швидкості, щільності населення тощо. А, як відомо, облік цих факторів вимагає принципово нелінійного підходу до опису

еволюційних процесів. Ці процеси моделюються і аналізуються за допомогою комп'ютерів, для яких нелінійність моделі не є перешкодою для її детального аналізу. І з'ясувалося, що в таких системах хаотичний режим функціонування швидше правило, ніж виняток.

5.3.8. Дивні хаотичні атрактори.

Математичним чином усталеного режиму функціонування дисипативної ДС служить *атрактор* – гранична множина у фазовому просторі, яка притягує, і до якої прагнуть всі вихідні траєкторії. Якщо сталий режим є стійкий стан рівноваги – атрактор системи буде просто нерухомою точкою, якщо це стійкий періодичний рух – атрактором буде замкнута крива, звана граничним циклом. Раніше вважалось, що атрактор є образом виключно стійкого режиму функціонування системи. Наразі ми розуміємо, що режим детермінованого хаосу теж атрактор в сенсі визначення граничної траєкторії в обмеженій області фазового простору. Однак такий атрактор має дві істотні відмінності: траєкторії на такому атракторі неперіодичні (вони не замикаються) і режим функціонування нестійкий (малі відхилення від режиму наростають). Саме ці відмінності й привели до необхідності ввести в розгляд новий термін. З легкої руки французького дослідника Ф. Такенса такі атрактори стали називати *дивними атракторами*.

Основним критерієм хаотичного атрактора є нестійкість траєкторії. Причому нестійкість повинна бути експоненційною. Це означає, що мале збурення режиму $D(0)$ має за часом збільшуватися по експоненті:

$$D(t) = D(0)\exp(\lambda t), \quad \lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{D(t)}{D(0)}, \quad (5.38)$$

де λ – старший показник Ляпунова.

Виявилось, що позитивність величини λ говорить не тільки про експоненційну нестійкість режиму коливань, але й доводить наявність в системі перемішування. Якщо встановлено, що досліджуваний режим має $\lambda > 0$, то наслідком будуть: неперіодичність в залежності від часу будь-якої з координат

стану, суцільний спектр потужності (в спектрі коливань присутні всі частоти з деякого інтервалу) і щільна у часі автокореляційна функція. До недавнього часу з такою поведінкою зазначених характеристик однозначно пов'язували уявлення про випадковий процес. Наразі відомо, що подібні властивості може мати процес, породжуваний детермінованими законами. Ця обставина і є підставою називати такі процеси детермінованим хаосом.

5.3.9. Дивні нехаотичні і хаотичні недивні атрактори.

Хаотичні атрактори, описані вище, об'єднують дві принципові властивості: складна геометрична структура (як наслідок, дробна метрична розмірність – фрактал) і експоненціальна нестійкість індивідуальних траєкторій. Саме ці властивості використовуються експериментаторами в якості критеріїв при діагностиці режимів детермінованого хаосу.

Однак режими складної динаміки вищеописаними типами хаотичних атракторів не вичерпуються. З'ясувалось, що хаотична поведінка в сенсі наявності перемішування і геометрична «дивина» атрактора можуть не відповідати одна одній. Дивні в геометричному розумінні атрактори можуть не бути хаотичними з причини відсутності експоненційної нестійкості фазових траєкторій. З іншого боку, є приклади, дисипативних систем, що перемішуються, атрактори яких не є в строгому сенсі дивними, тобто характеризуються фрактальною структурою і дробовою метричною розмірністю.

Іншими словами, існують конкретні приклади дисипативних ДС, атрактори яких характеризуються наступними властивостями:

- 1) при регулярній геометричній структурі з точки зору цілочисельної метричної розмірності індивідуальні фазові траєкторії в середньому експоненційно нестійкі;
- 2) при складній геометричній структурі траєкторії асимптотично стійкі, перемішування відсутнє.

Перший тип називають *хаотичним недивним атрактором*, другий – *дивним нехаотичним атрактором*.

У підсумку, всі атрактори ДС можна розділити на два типи: регулярні, до яких відносяться різноманіття, які диференціюються– точки рівноваги, граничні цикли, двовимірні й багатовимірні тори, і нерегулярні, до яких відносяться всі атрактори, що володіють властивістю «дивацтва» або «хаотичності» (або, як це найчастіше буває, відразу обома цими властивостями).

Таким чином, у результаті простого якісного розгляду особливостей нелінійних дисипативних ДС ми прийшли до ряду важливих висновків.

1. В диференціальних системах з розмірністю фазового простору $N \geq 3$ теоретично можливі сталі режими коливань, які не є ні періодичними, ні квазіперіодичними. Вони являють собою динамічний хаос.

2. Принциповою особливістю хаотичних коливань є їх нестійкість, що призводить до чутливої залежності динаміки системи від малих збурень.

3. Нестійкість нелінійної системи в сукупності з обмеженістю енергії коливань може викликати перемішування.

4. Наявність перемішування призводить до необхідності введення статистичного опису динаміки детермінованих систем з хаотичними атракторами як найбільш зручного.

Перераховані результати переконують нас у тому, що режими функціонування детермінованих нелінійних систем з дивними хаотичними атракторами дійсно мають ряд специфічних властивостей, сукупність яких включається в поняття «детермінований хаос».

5.4. Сценарії розвитку хаосу. Типи атракторів

Нелінійні властивості ДС по-різному проявляються при зміні її керуючих параметрів. Але, зазвичай, з ростом впливу нелінійності відбувається ускладнення динамічного режиму. Прості атрактори у фазовому просторі

дисипативної системи замінюються більш складними. При певних умовах нелінійність призводить до виникнення динамічного хаосу. Рух у просторі параметрів уздовж відповідного напрямку дозволяє спостерігати послідовність біфуркацій, у результаті якої формується хаотичний атрактор. Такі типові біфуркаційні послідовності об'єднуються поняттям *біфуркаційних механізмів* або *сценаріїв розвитку хаосу*.

Перший із подібних сценаріїв був запропонований Л.Д. Ландау у 1944 р. і незалежно від нього Е. Хопфом у зв'язку зі спробами пояснити виникнення турбулентної поведінки рідини при збільшенні числа Рейнольдса. Відповідний біфуркаційний механізм, який отримав назву *сценарію Ландау-Хопфа*, передбачає послідовність біфуркацій (типу біфуркації Андронова-Хопфа), кожна з яких породжує нову некратну частоту: $\omega_1 \rightarrow \omega_1, \omega_2 \rightarrow \dots \rightarrow \omega_1, \omega_2, \dots, \omega_k$. В результаті виникає багаточастотний квазіперіодичний режим, відповідний багатовимірному тору у фазовому просторі ДС. Якщо число біфуркацій k досить велике, то спектр процесу з урахуванням флуктуацій, які завжди присутні в реальних системах, стає досить широкосмуговим, як і спектр хаотичних коливань. Однак багаточастотні квазіперіодичні коливання, навіть у присутності шуму, можуть залишатися стійкими по Ляпунову. Перемішування в такій системі буде пов'язане тільки з шумом, а не з детермінованим оператором еволюції. Таким чином, сценарій Ландау-Хопфа не передбачає обов'язкового переходу до хаотичної динаміки і, строго кажучи, не є сценарієм розвитку хаосу. Крім того, даний сценарій не може пояснити виникнення коливань із суцільним спектром в маломірних системах.

Ідея розвитку турбулентності через квазіперіодичні коливання на початку 1970-х рр. була перероблена з нових позицій Д. Рюелем, Ф. Такенсом і С. Ньюхаусом. Вони зв'язали турбулентну поведінку з динамічним хаосом і вперше ввели поняття *дивного атрактора* як математичного образу хаосу в детермінованій системі. При цьому було показано, що дивний атрактор може виникати в системах навіть з невеликою розмірністю фазового простору ($N > 3$).

Наразі відкриті і досліджені три типових біфуркаційних сценарії розвитку хаосу в дисипативних системах, що реалізовані вже у тривимірному фазовому просторі. Причому кожен з них має властивості універсальності, тобто деякими загальними закономірностями, що не залежать від конкретного виду оператора еволюції.

5.4.1. Перехід до хаосу через послідовність біфуркацій подвоєння періоду.

Перехід до хаосу через *послідовність біфуркацій подвоєння періоду* спостерігається в чисельних і натурних експериментах у величезній кількості найрізноманітніших ДС, починаючи з найпростіших відображень послідування і закінчуючи розподіленими середовищами. Цей перехід допускає однопараметричний аналіз (оскільки біфуркація подвоєння має корозмірність 1) і полягає в наступному.

Нехай ДС при деякому значенні керуючого параметра $\alpha = \alpha_0$ має стійкий граничний цикл C з періодом $T(\alpha)$. Нехай при збільшенні параметра до значення $\alpha = \alpha_1$ відбувається суперкритична біфуркація подвоєння періоду, що призводить до народження стійкого граничного циклу $2C$ з періодом $2T(\alpha)$. Далі спостерігається нескінченна послідовність біфуркацій подвоєння періодів циклів $2^k C$ в точках $\alpha = \alpha_k$, $k=1, 2, 3, \dots$. В спектрі виникають субгармонійні частоти $\omega_0 = 2\pi/T_0$, тому послідовність біфуркацій подвоєння іноді називають *субгармонійним каскадом*. Біфуркаційні точки α_k сходяться в границі $k \rightarrow \infty$ до деякого критичного значення $\alpha = \alpha_{cr}$, при якому період стає нескінченним, а спектр суцільним. При $\alpha > \alpha_{cr}$ виникають аперіодичні коливання, нестійкі по Ляпунову. Цим коливань відповідає хаотичний атрактор в фазовому просторі системи.

Як приклад на рис. 5.20 представлені зміни, що відбуваються в процесі переходу до хаосу через послідовність подвоєнь періоду в генераторі з інерційною нелінійністю Аніщенко-Астахова, описуваному рівняннями:

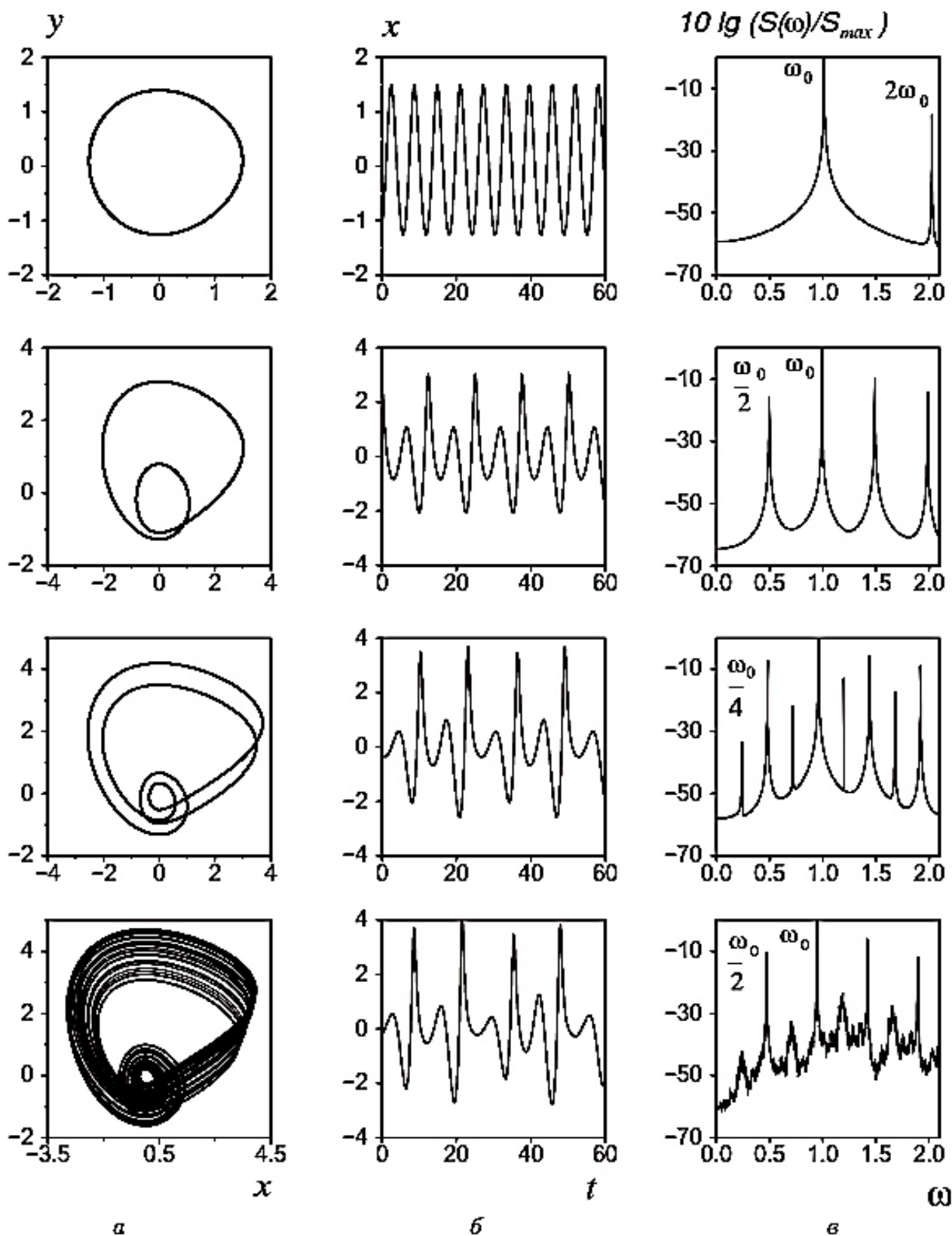


Рис. 5.20. Послідовність біфуркацій подвоєння періоду:

a – проєкції фазових траєкторій, *б* – форма коливань, *в* – спектри потужності для циклів з періодами $2T_0^k$, $k=1, 2, 3$, і дивного атратора

$$\begin{aligned}
\dot{x} &= mx + y - xz, \\
\dot{y} &= -x, \\
\dot{z} &= -gz + g\Phi(x),
\end{aligned}
\tag{5.39}$$

де $\Phi(x)=x^2$ при $x \geq 0$ і $\Phi(x)=0$ при $x < 0$.

Експериментально було встановлено, що у всіх без винятку тривимірних потокових системах хаотичний атрактор, який формується в результаті послідовності подвоєнь періоду, має фрактальну розмірність $2 < d < 3$, а його перетин за формою нагадує підкову Смейла. У цьому випадку найпростішою моделлю, яка виникає в січній поверхні відображення може служити відображення Ено (Henon).

Нелінійне відображення Ено описує рух часток в умовах тертя й імпульсних впливів (зіткнень), а також – ротатор зі збудженням:

$$x_1[n+1] = 1 - ax_1^2[n] - bx_2[n]; \quad x_2[n+1] = x_1[n],
\tag{5.40}$$

де a, b – параметри; x_1, x_2 – координати; n – такт часу.

Параметр a характеризує просторовий розподіл взаємодії (зіткнення) частинок, а параметр b – загасання в системі. Він обернено пропорційний коефіцієнту тертя (реальними є значення $b > 0$).

Поводження системи (режими її роботи) наочно зображуються у вигляді карти динамічних режимів на площині параметрів системи (a, b) , яка для відображення Ено наведена на рис. 5.21. Нижня частина 1 карти (див. рис. 5.21) відповідає встановленню стаціонарного стану рівноваги. З ростом параметра a і зменшенням b реалізуються режими періодичних коливань періоду 2, а потім спостерігається послідовність біфуркацій 6 подвоєння періоду 4, 8, ... і, нарешті, виникає режим хаосу 3. Біла область 5 у верхній частині відповідає нестійкому стану.

Теорія переходу до хаосу через послідовність біфуркацій подвоєння періоду на базі одновимірних відображень отримала назву *сценарію*

Фейгенбаума. Найпростішою моделлю для дослідження сценарію Фейгенбаума служить логістичне відображення

$$x_{n+1} = f(x_n) = r - x_n^2, \quad (5.41)$$

де r – параметр відображення, $r \geq 0$.

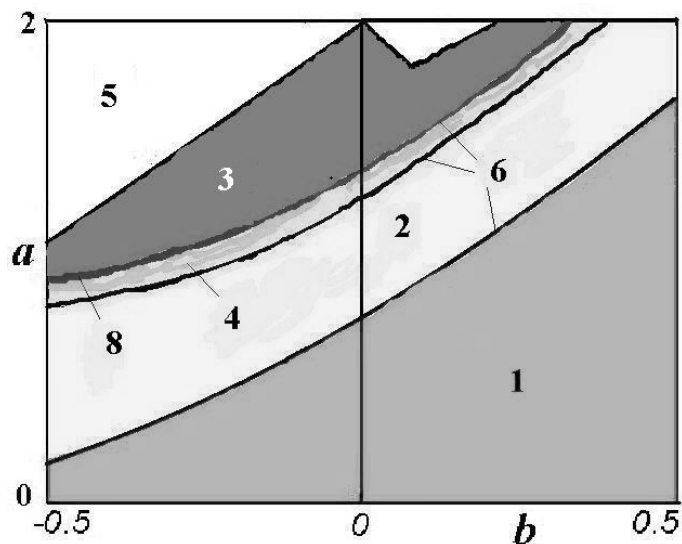


Рис. 5.21. Карта режимів відображення Ено

Логістичне відображення може бути представлено в інших формах запису, що зводяться до (5.41) заміною змінних, наприклад,

$$x_{n+1} = rx_n(1 - x_n), \quad x_{n+1} = 1 - rx_n^2. \quad (5.42)$$

Розглянемо, як поводить себе відображення (5.41) з ростом параметра r . Нерухома точка відображення $x_0 > 0$ (цикл періоду 1, або 1-цикл) має координату $x_0 = -\sqrt{r + 1/4}$ і стійка при $r \in [0; 3/4]$ (мультиплікатор нерухомої точки μ_1 дорівнює $-2x_0$). При $r=r_1=3/4$ мультиплікатор приймає значення $\mu_1=-1$. Має місце біфуркація подвоєння періоду 1-циклу. Народжується стійкий цикл періоду 2, що складається з точок x_1, x_2 . Координати точок 2-циклу дорівнюють $x_{1,2} = 1/2 \pm \sqrt{r - 3/4}$, а мультиплікатор циклу є $\mu_2 = f'_x(x_1)f'_x(x_2) = 4(1 - r)$. 2-цикл стійкий в області значень параметра $r \in [3/4; 5/4]$. При $r=r_2=5/4$ мультиплікатор приймає значення $\mu_2=-1$ і відбувається наступна біфуркація подвоєння періоду. Народжується 4-цикл і так далі. Отримуємо послідовність

біфуркаційних значень параметра $r_1=3/4$, $r_2=5/4$, $r_3\approx 1.368099$, $r_4\approx 1.394046$, $r_5\approx 1.399637, \dots$, що накопичується до критичної точки $r_{кр}\approx 1.40115 \dots$ При $k\rightarrow\infty$ швидкість збіжності біфуркаційних значень прагне до деякої кінцевої границі

$$\delta = \lim_{k\rightarrow\infty} \frac{r_{k+1} - r_k}{r_{k+2} - r_{k+1}} \approx 4,669201\dots \quad (5.43)$$

Фазо-параметрична діаграма режимів відображення (5.41), яка наведена на рис. 5.22, характерна для систем з каскадом подвоєнь періодів, що призводить до хаосу.

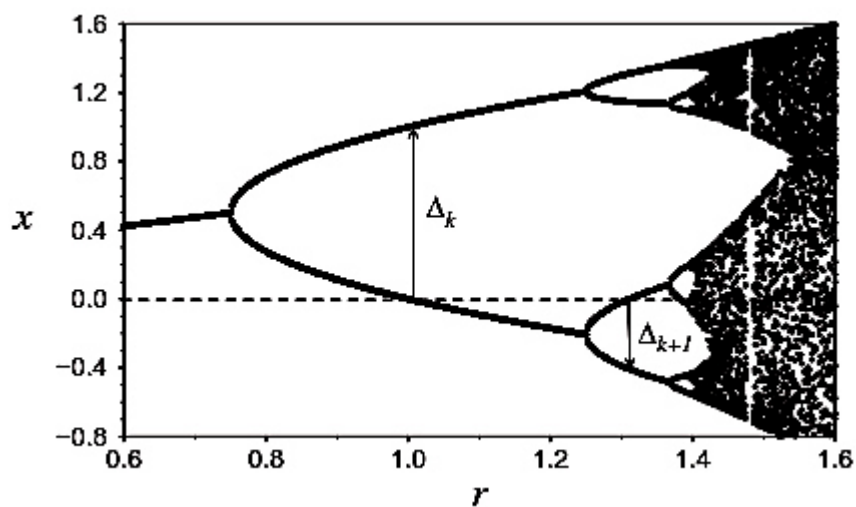


Рис. 5.22. Фазо-параметрична діаграма режимів відображення

Діаграма дає наочне уявлення про дроблення масштабу динамічної змінної і наявності властивостей *скейлінга*, тобто *масштабної інваріантності*, коли один і той же елемент зображення повторюється у все більш дрібному масштабі. Позначивши відстані між подібними точками гілок дерева Δ_k (як показано на рис. 5.22), можна ввести масштабні множники $a_k = \Delta_k / \Delta_{k+1}$, які в границі сходяться до деякого значення

$$a = \lim_{k\rightarrow\infty} \frac{\Delta_k}{\Delta_{k+1}} \approx -2,5029\dots \quad (5.44)$$

Як показали чисельні дослідження, величини δ і a не залежать від конкретного виду відображення. Головне, щоб воно було унімодальне (мало один екстремум) і щоб екстремум був квадратичним.

5.4.2 Жорсткі переходи до хаосу. Криза і перемежана.

З розвитком уявлень про динамічний хаос було встановлено, що перехід від періодичних коливань до хаосу може відбуватися стрибком, в результаті однієї єдиної біфуркації. Такий механізм виникнення хаосу називають *жорстким*. Він супроводжується явищем *перемежаності*. Перемежаністю називають режим чергування в часі майже регулярних коливань (*ламінарна фаза*) з інтервалами хаотичної поведінки (*турбулентна фаза*), що спостерігається відразу за порогом виникнення хаосу. Типовий вид коливань в режимі перемежаності наведено на рис. 5.23.

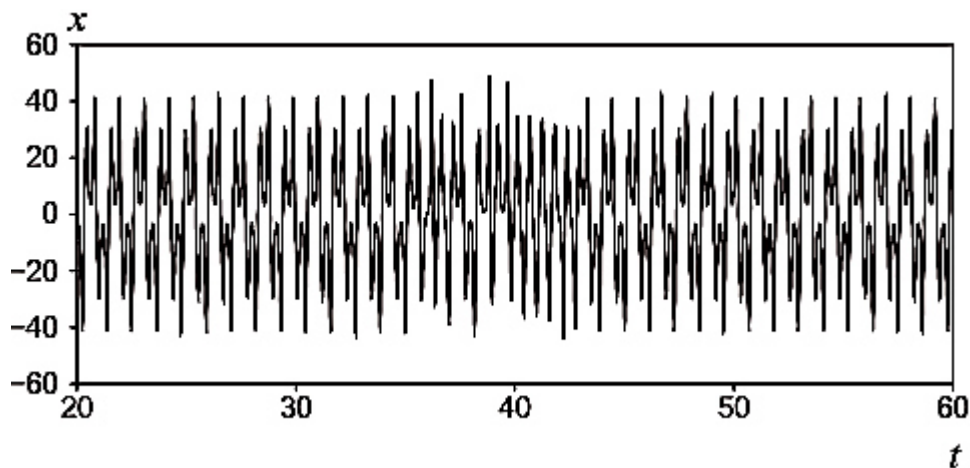


Рис. 5.23. Перемежуваність в системі Лоренца (при $r = 166.1$, $\sigma = 10$, $b = 8/3$)

Жорсткий перехід до хаосу і явище перемежуваності були вперше розглянуті в роботах І. Помо і П. Маневіля, тому відповідний біфуркаційний механізм виникнення хаосу отримав назву сценарію *Помо-Маневіля*.

При жорсткому виникненні хаосу єдина біфуркація періодичного режиму призводить до різкої якісної перебудови структури фазового простору, включаючи структуру басейну тяжіння атратора. Подібні біфуркації атраторів називають *кризами*. Як вже зазначалося в п. 5.2, типовими кризами періодичного режиму (граничного циклу) є такі локальні біфуркації корозмірності 1: дотична (сідло-вузлова) біфуркація, субкритична біфуркація

подвоєння періоду та субкритична біфуркація народження тора (біфуркація Андронова-Хопфа в відображенні). У разі дотичної біфуркації стійкий граничний цикл зникає, зливаючись із сідловим. У двох інших випадках граничний цикл продовжує існувати і після біфуркації, але стає нестійким (сідловим).

Нехай при $\alpha < \alpha_{кр}$ система має атрактор – граничний цикл C . В результаті будь-якої з перерахованих біфуркацій в точці $\alpha = \alpha_{кр}$ атрактор C перестає існувати. При $\alpha > \alpha_{кр}$ фазові траєкторії з локальної околиці зниклого атрактора C повинні потрапляти на якийсь інший атрактор, або вже існуючий в системі при $\alpha < \alpha_{кр}$, або такий, що виникає в результаті біфуркації. Нехай ДС вже мала інший атрактор. Тоді в результаті біфуркації спостерігається просте перемикання з одного режиму на інший. Перемежуваність при цьому не виникає, навіть якщо новий режим є хаотичним. Справа в тому, що в цьому випадку криза граничного циклу не є причиною, яка породжує хаотичний атрактор, а сам атрактор захоплює локальну околицю циклу C . Траєкторії йдуть з цієї околиці і не повертаються. Які ж умови, при яких криза граничного циклу призводить до виникнення перемежованого хаосу? Очевидно, це відбувається в тому випадку, коли в біфуркаційній точці $\alpha = \alpha_{кр}$ вже існує хаотична множина, яка при $\alpha > \alpha_{кр}$ стає прилягаючою і включає в себе локальну околицю циклу C так, що фазова траєкторія на хаотичному атракторі час від часу в цю околицю повертається. Умовою реалізації такої поведінки системи може з'явитися наявність у сідлових граничної множини, що бере участь в кризі атрактора C , гомоклінічної траєкторії.

Як приклад на рис. 5.24 представлена дотична біфуркація циклів, яка веде до хаотичної перемежуваності. Сідловий цикл має пару грубих гомоклінічних траєкторій. У точці біфуркації $\alpha = \alpha_{кр}$ утворюється негруба сідло-вузлова орбіта з гомоклінічною структурою в її околиці. Траєкторії віддаляються від неї і наближаються до неї уздовж двояко асимптотичних гомоклінічних кривих (їм відповідають точки перетину різноманіть в перерізі, зображеному на рис. 5.24). При $\alpha > \alpha_{кр}$ негруба замкнута орбіта зникає, а непрягуюча гомоклінічна

структура стає притягуючою. У фазовому просторі ДС виникає хаотичний атрактор. Траєкторії на ньому густішають в області, де існувала сідло-вузлова орбіта, подовгу повторюючи рух на ній, що відповідає ламінарній фазі перемежуваного хаосу.

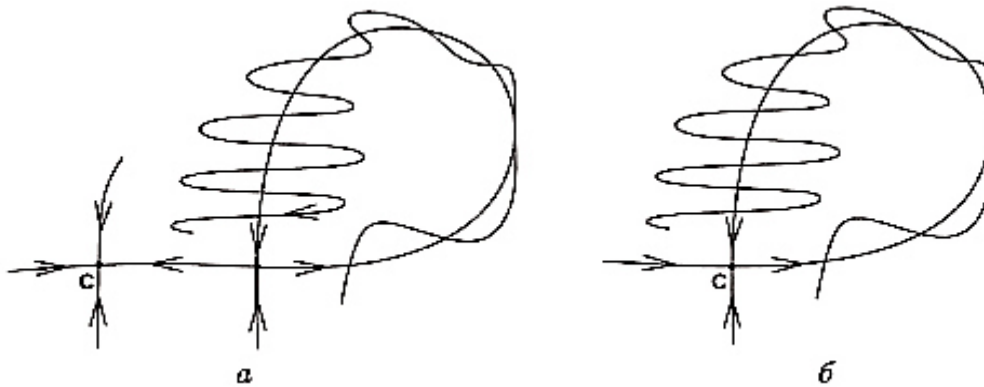


Рис. 5.24. Якісний вид перетинів Пуанкаре для дотичній біфуркації стійкого і сідлового циклів, що приводить до виникнення хаосу через перемежуваність: *a* – до біфуркації, *б* – в точці біфуркації

5.4.3. Перехід до хаосу через руйнування квазіперіодичних коливань.

Згідно зі сценарієм Рюеля-Такенса-Ньюхауса перехід від квазіперіодичних коливань до хаосу відбувається після народження третьої частоти, коли на тривимірному торі стає можливим поява нестійких по Ляпунову хаотичних траєкторій. Однак дослідження конкретних ДС показало, що не менш типовим є перехід до хаосу через руйнування двочастотного квазіперіодичного руху. При цьому двовимірний тор T^2 в фазовому просторі повинен руйнуватися, після чого траєкторії потрапляють на множину з фрактальною розмірністю $2+d$, $d \in [0; 1]$, що утворюється в його околиці і називається *тор-хаосом*. Такий сценарій можна розглядати як особливий випадок переходу до хаосу через квазіперіодичний рух.

На відміну від сценарію Фейгенбаума, перехід T^2 в хаос вимагає двопараметричного аналізу. Це пов'язано із тією обставиною, що характер квазіперіодичного режиму залежить від числа обертання θ , що визначає базові

частоти коливань. Якщо θ раціонально, то має місце резонанс на торі (й, відповідно, періодичні коливання). При ірраціональному значенні числа обертання рух на торі буде ергодичним. Спостерігати перехід від тора до хаосу при фіксованому значенні числа обертання можна тільки контролюючи, як мінімум, два параметри системи одночасно. На лінії народження тора, що задається біфуркаційною умовою $\mu_{1,2} = \exp j\phi$, де $\mu_{1,2}$ – пара комплексно-сполучених мультиплікаторів граничного циклу, число обертання визначається як $\theta = \phi/2\pi$. Області резонансів на площині двох керуючих параметрів мають форму язиків, що спираються гострим кінцем на відповідні точки лінії народження тора. Ці області називають *язиками* або *дзьобами Арнольда* (в честь В.І. Арнольда, який досліджував структуру резонансних областей). Для кожного обраного шляху руху у просторі параметрів характерна своя послідовність біфуркацій, пов'язаних з виникненням і зникненням різних резонансів на торі.

5.4.4. Грубі і негрубі динамічні системи.

Розглянемо клас автономних ДС з безперервним часом і розмірністю фазового простору $N > 3$. Крім грубих систем, аналогічних системам Андронова-Понтрягіна на площині, з'являється клас грубих систем з нетривіальною гіперболічністю, тобто систем з хаотичною динамікою. Хаотичні атрактори грубих гіперболічних систем – це в строгому математичному сенсі дивні атрактори. Зазвичай такі атрактори представляють собою якусь математичну ідеалізацію і в експериментах, як правило, не спостерігаються. У більшості випадків системи з нерегулярною динамікою виявляються негрубими. Математиками доведено, що грубі гіперболічні системи не є всюди щільними на множині динамічних систем з $N > 3$. Структурна нестійкість (негрубість) пов'язана з виникненням грубих двояко асимптотичних траєкторій: сепаратрисних петель, гомоклінічних і гетероклінічних кривих, що утворюються при нетрансверсальному перетині різноманіть сідлових циклів і

інших сідлових множин. Строгий математичний опис негрубих систем зі складною динамікою викликає великі труднощі, оскільки навіть при малих змінах параметрів спостерігається безліч біфуркацій. Теоретично число біфуркацій в будь-якому інтервалі значень керуючого параметра може бути нескінченним і описати їх всі неможливо. Крім того, негрубі системи дуже чутливі до точності завдання початкового стану і до наявності шуму (навіть якщо це шум округлення при проведенні чисельних експериментів). При таких умовах нас, як правило, вже не цікавлять окремі біфуркації і деталі фазового портрета. Важливим стає загальний характер поведінки з урахуванням флуктуацій і статистичні характеристики режиму, які можна вимірювати експериментально. Негрубі дисипативні ДС можна розділити на два класи: майже гіперболічні (*квазігіперболічні*) ДС та ДС з квазіатракторами. Квазігіперболічні системи характеризуються атракторами, які найбільш близькі за своїми характеристиками до грубих гіперболічних. Основною відмінною рисою таких атракторів є відсутність стійких циклів і торів, включених в атрактор. Квазіатрактори принципово містять стійкі цикли (тори), які суттєво впливають на їх властивості.

Поняття атрактора ДС в останні роки набуло важливість не тільки у зв'язку з відкриттям ефекту детермінованого хаосу. З поняттям атрактора в нелінійній теорії коливань пов'язують ефект автоколивань в ДС. Якщо за основу визначення автоколивань на відміну від Андронова прийняти головне: незалежність усталеного режиму від початкових умов, то будь-який тип атрактора автономної ДС буде являти собою математичний образ автоколивань системи в її фазовому просторі. Властивості автоколивальних режимів при такому розумінні будуть визначатися структурою і властивостями відповідних їм атракторів. Так, наприклад, стійким регулярним автоколиванням відповідатимуть стійкі граничні цикли і тори, а хаотичним автоколиванням – відповідні хаотичні атрактори.

5.4.5. Квазігіперболічні атрактори. Атрактори типу Лоренца.

Умови гіперболічності атрактора для реальних ДС, як правило, не виконуються. Разом з тим відомі ДС, атрактори яких є близькими до гіперболічних. Такі атрактори є хаотичними, не включають стійкі регулярні атрактори і зберігають ці властивості при збуреннях. З математичної ж точки зору, для таких систем порушується умова суворої трансверсальності і утворюються негрубі двояко асимптотичні криві.

Розглянемо приклад квазігіперболічного атрактора в диференціальній системі – системі Лоренца. Для атрактора Лоренца порушується одна з вимог гіперболічності (умова суворої трансверсальності). Атрактори типу Лоренца виявлені в ряді систем і являють собою типовий приклад квазігіперболічних атракторів. Доведено, що атрактор Лоренца включає тільки сідлові траєкторії, при варіації параметрів біфуркації в ньому відсутні стійкі точки або цикли не виникають. Атрактор Лоренца розглядається як класичний приклад квазігіперболічного хаосу.

Рівняння Лоренца були отримані з рівнянь Нав'є-Стокса в задачі про теплову конвекцію і мають вигляд

$$\begin{aligned}\dot{x} &= -\sigma(x - y), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= xy - bz,\end{aligned}\tag{5.45}$$

де σ , b і r – керуючі параметри. До рівнянь типу (5.45) зводяться деякі моделі лазерів, а також модель дискового динамо.

Відзначимо, що в системі (5.45) режим квазігіперболічного хаосу реалізується в кінцевій області значень її керуючих параметрів. На рис. 5.25 представлена біфуркаційна діаграма системи. Існуванню атрактора Лоренца відповідає заштрихована область в параметричному просторі.

Фазовий портрет атрактора Лоренца представлений на рис. 5.26,а. Поза зазначеної області властивості хаотичного атрактора будуть іншими: атрактор Лоренца трансформується в квазіатрактор.

Назвемо типові властивості атрактора Лоренца. Спектр ЛХП не змінюється при варіації початкових умов, оскільки атрактор Лоренца є єдиним, басейном тяжіння якого служить весь фазовий простір; спектр ЛХП практично не змінюється, якщо варіювати керуючі параметри системи в області існування атрактора Лоренца.



Рис. 5.25. Біфуркаційна діаграма системи Лоренца на площині параметрів r і σ для $b = 8/3$

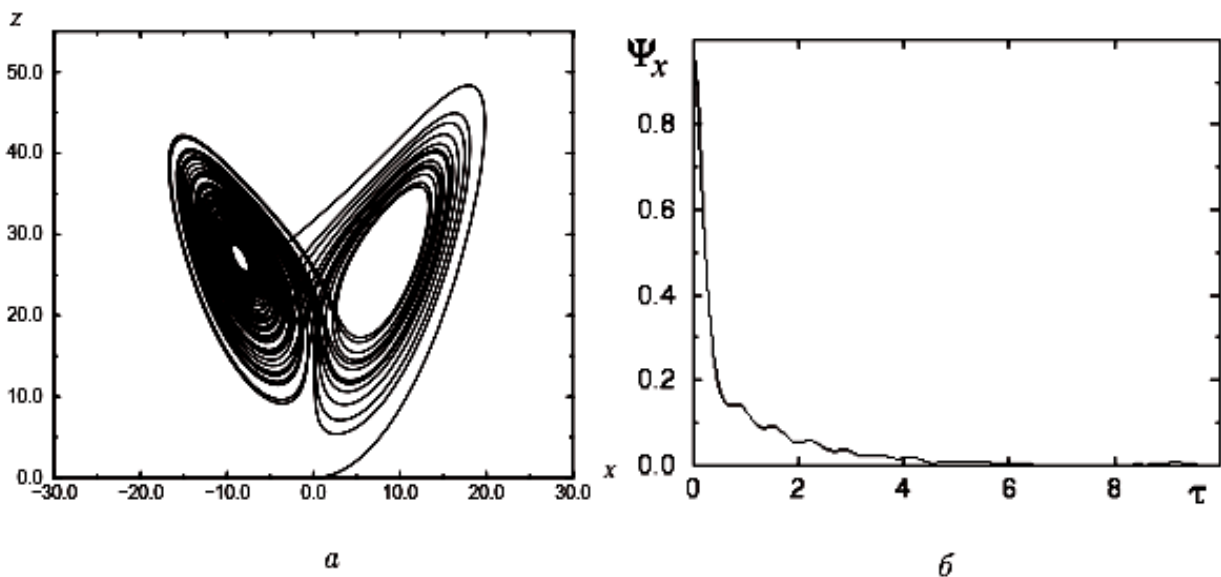


Рис. 5.26. Фазовий портрет (а) і автокореляційна функція (б) атрактора Лоренца (при $\sigma = 10$, $r = 28$ і $b = 8/3$)

Ці властивості наочно ілюструють грубість атрактора Лоренца з точки зору експерименту: структура атрактора зберігається при варіації параметрів і початкових умов, біфуркації атрактора відсутні. Наслідком хаотичної динаміки є характерний вид автокореляційної функції, яка експоненційно спадає зі збільшенням часу практично монотонно, що ілюструє рис. 5.26,б.

5.4.6. Квазіатрактори і їх властивості.

У системах з квазіатракторами реалізуються режими детермінованого хаосу, що характеризуються експоненційною нестійкістю траєкторій і фрактальною структурою атрактора. З цієї точки зору характеристики зазначених режимів автоколивань ідентичні основним характеристикам грубих гіперболічних атракторів і квазігіперболічних атракторів. Однак існує дуже суттєва відмінність, яка вимагає розуміння щоб уникнути невірного трактування експериментальних результатів. Відмінною рисою квазіатракторів є одночасне співіснування рахункової множини різних хаотичних і регулярних притягуючих підмножин в обмеженому елементі обсягу фазового простору системи при фіксованих значеннях її параметрів. Ця сукупність всіх співіснуючих граничних множин траєкторій в обмеженій області G_0 фазового простору, куди прагнуть всі або майже всі траєкторії з області G_1 , що включає G_0 , і називається квазіатрактором ДС. Звідси випливає надзвичайно складна структура вкладених басейнів їх тяжіння. Але цим складність не обмежується. При кінцевій варіації параметрів системи реалізуються каскади різних біфуркацій як регулярних, так і хаотичних атракторів. Відповідно здійснюється біфуркаційна перебудова їх басейнів тяжіння. Причиною складності квазіатракторів є ефекти гомоклінічного торкання стійких і нестійких різноманіть сідлових траєкторій або виникнення петлі сепаратриси сідло-фокуса, які мають місце на множині значень параметрів ненульової міри. Іншими словами, порушується умова трансверсальності різноманіть.

Якщо при цьому врахувати, що басейни тяжіння співіснуючих граничних множин можуть мати фрактальну структуру і складати надзвичайно вузькі області в фазовому просторі, то стає зрозуміло, наскільки важливі проблеми точності розрахунків на комп'ютері і впливу флуктуацій.

Квазіатрактори реалізуються, наприклад, у відображенні Ено.

5.5. Реконструкція динамічних систем

Часто доводиться стикатися зі складною ситуацією, коли детальні відомості про реальну систему або відсутні зовсім, або явно недостатні. Єдина інформація про властивості системи міститься лише в експериментальній залежності однієї з координат стану системи в часі. Така залежність $a(t)$, виміряна протягом кінцевого часу t_0 , називається *спостережною* (або *реалізацією*) системи а при дискретизації з кроком Δt : $a(i\Delta t)=a_i$, $i=1,\dots,N$; $N=\lceil t_0/\Delta t \rceil$, вона носить назву *одновимірного часового ряду*. Робиться припущення про те, що спостережувана $a(t)$ є детерміновано визначеною, тобто являє собою одновимірну проекцію фазової траєкторії, яку породжує деяка ДС. Одним із завдань реконструкції ДС є відновлення її моделі, рішення якої з певним ступенем точності відтворює одновимірну спостережувану $a(t)$ на заданому інтервалі часу t_0 і для $t>t_0$. Ця проблема належить до класу обернених задач, рішення яких не може бути однозначним.

Розглянемо основні проблеми, з якими пов'язано вирішення завдання реконструкції. Перша обумовлена необхідністю введення будь-яким чином координат стану системи. Адаже нам відома залежність у часі (на кінцевому інтервалі) лише однієї з координат реальної системи $a(t)$. Якби нам були відомі рівняння, які генерують досліджуваний сигнал $a(t)$, то можна було б перейти до рівнянь у варіаціях, розглядаючи нескінченно малі збурення. Але в нашому розпорядженні знаходиться лише часова залежність однієї з координат стану, і всю інформацію про складний режим коливань ми повинні витягти тільки з неї.

Не менш складною проблемою є відновлення моделі ДС. Доведеться шукати відповідь на цілий ряд серйозних питань, зокрема, який вид модельного оператора еволюції, який в разі ЗДР визначається правими частинами системи n диференціальних рівнянь першого порядку?

Раніше вважалося, що для вивчення динаміки ДС в термінах фазового простору необхідне знання усіх координат, що визначають її стан. Однак у 1980 р. була опублікована робота Н. Пакарда, в якій показано, що фазовий портрет ДС може бути відновлений по скалярному часовому ряду a , якщо в якості відсутніх координат вектора стану використовується той же самий ряд a , взятий з деяким запізненням. У 1981 р. була доведена теорема, яка стверджує, що по одновимірній реалізації $a(t)$ ДС, що має аттрактор A , який належить гладкому d -мірному різноманіттю, методом затримки можна отримати n -мірну реконструкцію A_R вихідного аттрактора як множини векторів $x(t)$ в R^n при $n \geq 2d+1$ (теорема Такенса):

$$x(t) = \Lambda_n(a(t)) = \{a(t), a(t+\tau), \dots, a(t+(n-1)\tau)\} = \{x_1, x_2, \dots, x_n\}. \quad (5.46)$$

Згідно з теоремою, відображення $\Lambda_n: A \rightarrow A_R$ є гладким і оборотним на A_R майже при довільній затримці τ (якщо $N \rightarrow \infty$).

Спробуємо розібратися в змісті теореми Такенса. Вона обґрунтовує введення в якості нових координат стану системи значень реалізації $a(t)$, взятих через деякий інтервал часу τ :

$$x_1(t) = a(t), x_2(t) = a(t+\tau), x_3(t) = a(t+2\tau), \dots \quad (5.47)$$

Оскільки на комп'ютері аналізується ряд значень змінної $a(t)$ в дискретні моменти часу $i\Delta t$, множина векторів, яка реконструюється, також є дискретною $x(i\Delta t)$, а величина τ має вигляд $\tau = k\Delta t$, де k – ціле число. Іншими словами, на практиці рівність (5.47) може бути переписана наступним чином:

$$x_i = \{a_i, a_{i+k}, \dots, a_{i+(n-1)k}\}, \quad (5.48)$$

де нижній індекс i відповідає моменту часу $i\Delta t$.

Техніка реконструювання полягає у виборі значень затримки τ , розмірності простору вкладення n і в формуванні масиву векторів $x(i\Delta t)$. Передбачається, що повний час спостереження $T_{obs} = N\Delta t$ і число точок N досить

великі, щоб по траєкторії можна було судити про найважливіші властивості атрактора, який нас цікавить.

Як приклад розглянемо рівняння моделі Лоренца (5.45) при значеннях параметрів $\sigma=10$, $r=28$, $b=8/3$, відповідних режиму динамічного хаосу. Проекція фазового портрета хаотичного режиму на площину (x, y) зображена на рис. 5.27,*а*. За часовою залежністю $x(t)$ рівнянь (5.45), що розглядається в якості аналізованої реалізації, можна здійснити реконструкцію методом затримки (5.46) – рис. 5.27,*б,в*. Відповідно до теореми Такенса ми очікуємо, що по відновленому атрактору (див. рис. 5.27,*в*) можуть бути обчислені такі характеристики аналізованого режиму динаміки, як фрактальна розмірність.

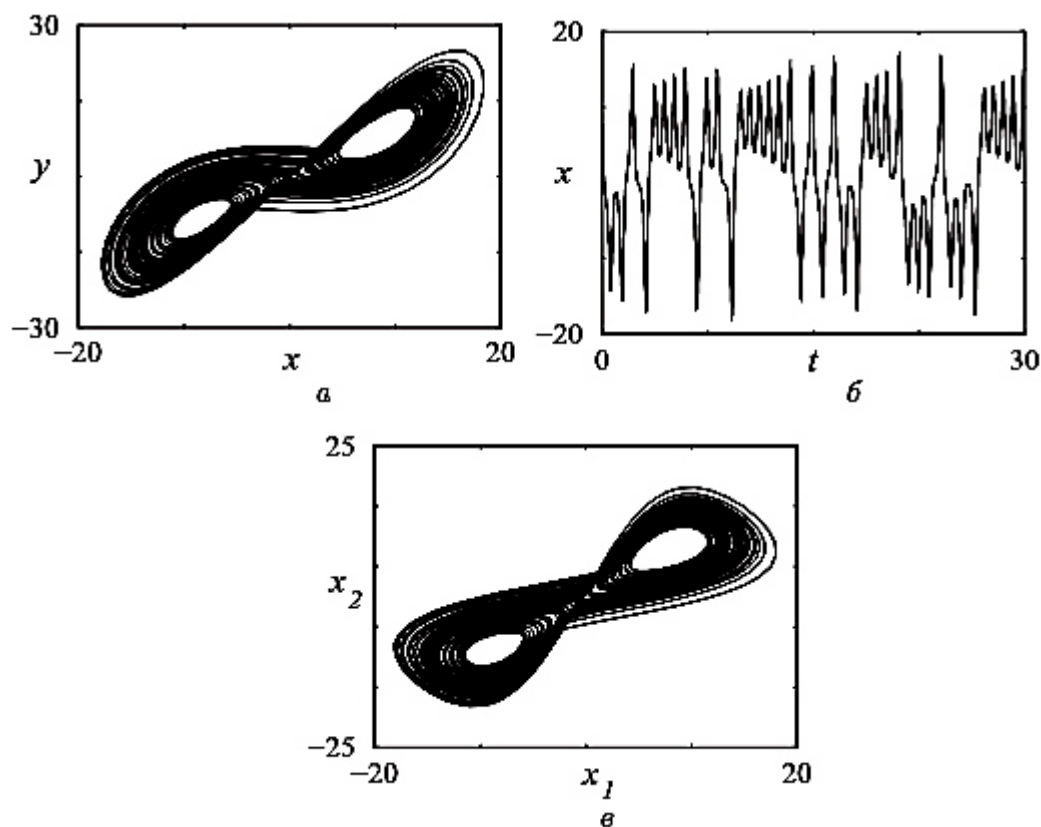


Рис. 5.27. Проекція фазового портрета хаотичного атрактора Лоренца (*а*), часова залежність координати $x(t)$ (*б*) і реконструйований атрактор (*в*)

Через роботу з кінцевим числом точок N необхідно ретельно вибирати параметр τ , оскільки якість реконструкції буде помітно відрізнятися при його

варіації. Якщо τ занадто мале (рис. 5.28,а), то координати i та $i+1$ точок фазової траєкторії практично не відрізняються одна від одної. З іншого боку, якщо час затримки дуже великий (рис. 5.28,б), координати виявляються некорельованими, і реконструйований аттрактор не відображає справжньої динаміки. На основі експериментів встановлено, що оцінка раціонального часу затримки може бути отримана з розрахунків автокореляційної функції $\psi(\tau_0) = \langle a(t)a(t + \tau_0) \rangle$, яка для складних неперіодичних процесів буде спадаючої в часі τ_0 . Значення τ_0 , відповідне часу досягнення першого нуля функції, використовується в експериментальних дослідженнях для оцінки затримки τ в (5.46) при введенні нових координат стану.

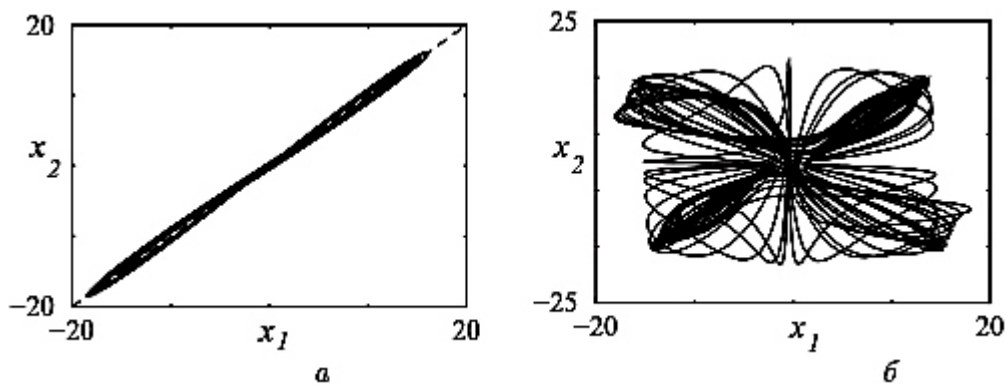


Рис. 5.28. Результат реконструкції при малій (а) і великій (б) затримці

Для визначення розмірності простору вкладення n необхідно знати розмірність аттрактора d реальної ДС (формула Мане):

$$n \geq 2d + 1. \quad (5.49)$$

Тобто, розмірність n – це розмірність простору вкладення, необхідна для реконструкції аттрактора. Таким чином, для вирішення завдання реконструкції необхідно визначити розмірність простору вкладення і задати набір векторів стану ДС в цьому новому фазовому просторі.

5.5.1. Визначення розмірності вкладення і реконструкція атрактора.

Припустимо, що в результаті експерименту отримані значення фізичної величини a , тобто набір $a_i = a(i\Delta t)$, $i=1, \dots, N$. Будемо вважати, що часовий ряд a_i породжується деякою ДС з безперервним або дискретним часом, представляючи собою дискретизовану з кроком Δt одновимірну проекцію фазової траєкторії. Ця траєкторія належить атрактору системи, розмірність якого дорівнює d . Згідно Такенсу, задати вектор стану можна за методом (5.46).

Спочатку необхідно визначити розмірність атрактора d . Для опису складної геометрії хаотичних атракторів доцільно розглядати розмірність Хаусдорфа, яка обчислюється наступним чином. Вважається, що аналізована множина S в просторі R^n покривається кубиками $\{B_i\}$ з величиною ребра, що не перевищує деяке значення ε . При цьому кожна точка множини S повинна обов'язково потрапити в той чи інший кубик. Тоді міра Хаусдорфа l_δ вводиться наступним чином:

$$l_\delta(S) = \lim_{\varepsilon \rightarrow 0} \inf_{K(\varepsilon)} \sum_{B_i \in K(\varepsilon)} |B_i|^\delta. \quad (5.50)$$

Тут \inf – мінімальне значення (нижня границя) по всіх можливих покриттях $K(\varepsilon)$ множини S кубиками $\{B_i\}$; $|B_i|$ – величина ребра кубика ($|B_i| \leq \varepsilon$). Зазначена границя залежить від параметра δ . Розмірність Хаусдорфа d_H є таким значенням δ , при якому величина $l_\delta(S)$ є кінцевою:

$$\begin{cases} \delta > d_H(S) \Rightarrow l_\delta(S) = 0, \\ \delta < d_H(S) \Rightarrow l_\delta = +\infty. \end{cases} \quad (5.51)$$

Згідно з цим визначенням, d_H може приймати нецілі значення. У загальному випадку, якщо розмірність є нецілою, її називають фрактальною. Відповідно, об'єкти, які характеризуються нецілою розмірністю, називають *фракталами*. Наявність нецілої розмірності є типовою особливістю хаотичних атракторів.

Поняття розмірності Хаусдорфа добре визначено з точки зору математики, але її надзвичайно складно обчислити. Тому зазвичай

використовують більш «практичні» визначення фрактальних розмірностей. Одним з таких «практичних» визначень є *ємність* (або *ємна розмірність* D_0). Нехай S – деяка множина в просторі R^n , яка покривається кубиками розміру ε . Якщо позначити через $N(\varepsilon)$ число кубиків, необхідних для покриття всієї множини, то *ємність* являє собою границю наступного вигляду:

$$D_0 = \lim_{\varepsilon \rightarrow 0} \frac{\lg N(\varepsilon)}{\lg(1/\varepsilon)}. \quad (5.52)$$

Визначення D_0 безпосередньо за формулою (5.52) часто ускладнюється дуже повільною збіжністю відношення $\lg N(\varepsilon)/\lg(1/\varepsilon)$ до границі $\varepsilon \rightarrow 0$ (якщо мова йде про розрахунок *ємності* об'єкта в фазовому просторі розмірності $n > 2$). Крім того, D_0 не залежить від ймовірності відвідування тих чи інших областей фазового простору, тобто не враховує статистичні властивості потоку, обумовлені динамікою аналізованої системи. Тому на практиці замість *ємності* частіше обчислюють *кореляційну розмірність* D_c , яку можна легше й швидше оцінити чисельно (в загальному випадку $D_c \leq D_0$):

$$D_c = \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\lg C(\varepsilon, N)}{\lg \varepsilon}, \quad (5.53)$$

де $C(\varepsilon, N) = N^{-2} \sum_{i \neq j} v(\varepsilon - |x_i - x_j|)$ – *кореляційний інтеграл*, ε – розмір осередку розбивки фазового простору, N – число точок, використовуваних для оцінки розмірності, v – функція Хевісайда, $x_i = x(i\Delta t)$. Для визначення D_c будують залежність $\lg C(\varepsilon, N)$ від $\lg \varepsilon$ і шукають на ній лінійну ділянку, нахил якої і визначає шукане значення розмірності. Крім того, іноді аналізують залежність $D_c(n)$ і збільшують n до тих пір, поки D_c не досягне насичення.

Відомо, що при обчисленні D_c існують обмеження на величину ε . Якщо ε наближається до розмірів атрактора ε_{\max} , то лінійна залежність $\lg C(\varepsilon, N)$ від $\lg \varepsilon$ пропадає, що пояснюється впливом кордонів атрактора, де число сусідів в кожній точці зазвичай менше, ніж в «середині». В границі, якщо $\varepsilon = \varepsilon_{\max}$, то $\lg C(\varepsilon, N) = 1$. З іншого боку, при зменшенні значення ε існує деяке ε_{\min} , таке, що

для $\varepsilon < \varepsilon_{\min}$ структура атрактора залишається невирішувемою. Як наслідок, знову порушується лінійна залежність $\lg C(\varepsilon, N)$ від $\lg \varepsilon$.

При розрахунку D_c можна виділити 3 важливі характеристики масиву даних: повний час спостереження T_{obs} , число точок N і крок між ними Δt . Вони пов'язані співвідношенням $T_{obs} = N\Delta t$, і їх слід розглядати разом, оскільки однієї характеристики недостатньо: велике T_{obs} при малому N може бути настільки ж погано, як і мале T_{obs} при великому N .

При цьому існують фундаментальні обмеження на значення кореляційної розмірності, пов'язані з кінцевою довжиною аналізованого часового ряду:

$$D_{\max} = \frac{2 \lg N}{\lg(1/r)}, \quad r = \frac{\varepsilon}{\varepsilon_{\max}}. \quad (5.54)$$

Дана формула означає, що алгоритм розрахунку розмірності не може дати значення більше, ніж D_{\max} при заданому числі точок N . Іншими словами, якщо $r=0.1$ і $N=1000$, то $D_{\max} \leq 6$; якщо $N=100000$, то $D_{\max} \leq 10$.

Метод затримки Такенса є найбільш відомим, але не єдиним способом завдання вектора стану. Альтернативою йому служить так званий *метод послідовного диференціювання*, що має певні переваги при вирішенні задачі реконструкції математичної моделі. Ідея даного методу наступна. Нехай є часовий ряд $a(i\Delta t) = a_i$, $i=1, \dots, N$. Завдання вектора стану в фазовому просторі проводиться наступним чином:

$$x(t) = \{a(t), da(t)/dt, \dots, d^{n-1}a(t)/dt^{n-1}\} = \{x_1, x_2, \dots, x_n\}. \quad (5.55)$$

Оскільки відомі значення a_i тільки в дискретні моменти часу $i\Delta t$, координати x_i вектора x визначаються шляхом чисельного диференціювання вихідного часового ряду по наближеним математичним формулам. Очевидно, що точність обчислення похідних буде визначатися малістю величини кроку дискретизації Δt . Недоліком методу є підвищена чутливість до шуму, що обмежує його застосування для просторів вкладення великої розмірності (принаймні, без проведення попередньої процедури фільтрації).

5.5.2. Розрахунок старшого показника Ляпунова по часовому ряду.

Поряд із вивченням складної геометрії хаотичних атракторів реконструкція дозволяє вирішувати задачі дослідження динамічних характеристик, до числа яких відносяться ляпуновські характеристичні показники ЛХП.

Припустимо, що задана n -вимірна ДС з безперервним часом. Щоб охарактеризувати стійкість її рішення, аналізується часова еволюція нескінченно малої n -вимірної сфери початкових умов; з плином часу ця сфера перетворюється в еліпсоїд (рис. 5.29).

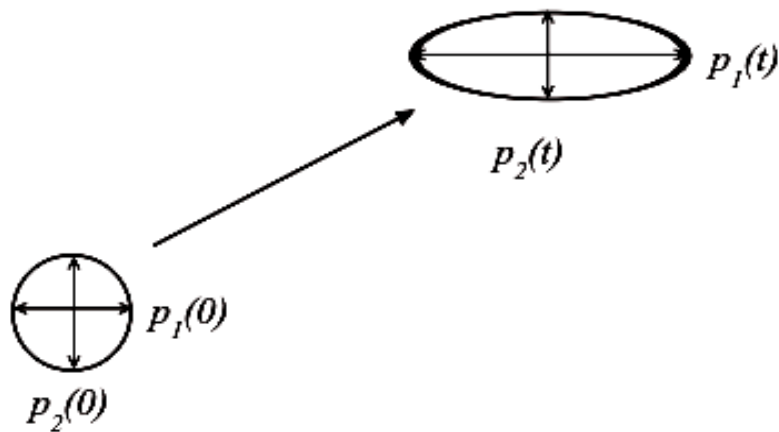


Рис. 5.29. Еволюція за часом нескінченно малої n -вимірної сфери початкових умов

Якщо говорити про спектр ЛХП, то i -й показник Ляпунова може бути визначений в термінах довжин вісей еліпсоїда $p_i(t)$:

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{p_i(t)}{p_i(0)}, \quad (5.56)$$

де λ_i впорядковані від найбільшого до найменшого. Таким чином, показники Ляпунова визначаються розширенням або стисненням сфери з різних напрямків у фазовому просторі. Оскільки орієнтація еліпсоїда безперервно змінюється з часом, змінюються і напрямки, що асоціюються з тим чи іншим показником.

Тому не можна говорити про якийсь один напрямок для кожної ляпуновської експоненти.

Слід зазначити, що якщо задається тільки одне збурення (це означає, що ми стежимо лише за головною віссю еліпсоїда), то воно буде в лінійному наближенні збільшуватись за законом $e^{\lambda_1 t}$. Для двох незалежних збурень площа утвореного ними квадрата (або прямокутника) еволюціонує за законом $e^{(\lambda_1 + \lambda_2)t}$; для трьох – еволюція обсягу описується законом $e^{(\lambda_1 + \lambda_2 + \lambda_3)t}$ і так далі. Дана властивість призводить до дещо іншого визначення спектру ляпуновських експонент: сума перших j показників визначається швидкістю експоненціального зростання j -вимірного елемента обсягу. Така інтерпретація забезпечує основу для техніки аналізу експериментальних даних (коли потрібно обчислити декілька показників).

Якщо рівняння ДС, що генерують фазову траєкторію, відомі, то визначити величину максимального показника (або повний спектр ЛХП) можна за допомогою стандартного алгоритму розрахунку ляпуновських експонент.

Розглянемо систему ЗДР:

$$dx/dt = f(x, \mu), \quad x \in R^n, \quad \mu \in R^m, \quad (5.57)$$

в якій x – вектор стану, f – нелінійна вектор-функція, μ – вектор керуючих параметрів. Дослідження даної системи на стійкість її часткового рішення $x_p(t)$ зводиться до аналізу рівнянь у варіаціях і введення поняття k -вимірного ляпуновського показника. Але якщо обмежитись тільки обчисленням старшого ЛХП, то цей алгоритм істотно спрощується. Шукана величина λ_1 визначатиме еволюцію в часі вектора збурення

$$r(t) = r_0 e^{\lambda_1 t}, \quad (5.58)$$

де r_0 – величина початкового збурення (в момент часу $t_0=0$), $r_0 = |x(t_0) - x_p(t_0)|$.

Дана формула є дуже наближеною, оскільки швидкість розбіжності траєкторій не є постійною, а залежить від вибору точки на аттракторі. Строго кажучи,

Якщо в якості способу завдання вектора стану у фазовому просторі використовується метод послідовного диференціювання, то математичну модель можна відновити в більш простому вигляді

$$\frac{dx_1}{dt} = x_2, \frac{dx_2}{dt} = x_3, \dots, \frac{dx_n}{dt} = f(x_1, x_2, \dots, x_n) \quad (5.62)$$

в силу того, що взаємозв'язок між координатами однозначно задається рівністю (5.55).

5.5.4. Приклад реконструкції динамічної системи.

Розглянемо конкретний приклад застосування описаного алгоритму. З цією метою виберемо відому динамічну систему Рьослера (R.Roessler)

$$\begin{aligned} dx/dt &= -y - z, \\ dy/dt &= x + ay, \\ dz/dt &= b - cz + xz, \end{aligned} \quad (5.63)$$

яка описує режим неперіодичних коливань. При значеннях параметрів $a=0.15$, $b=0.2$, $c=10.0$ система (5.63) характеризується режимом дивного атрактора

Використаємо в якості одновимірного часового ряду a_i залежність в часі однієї з координат $y(i\Delta t)$, отриману чисельною інтеграцією рівнянь (5.63). Будемо вважати, що вид системи (5.63) і її розмірність нам невідомі. Спостережувана $a(t)=y(t)$, задана на кінцевому інтервалі часу $0 \leq t \leq 100$, представлена на рис. 5.30,а.

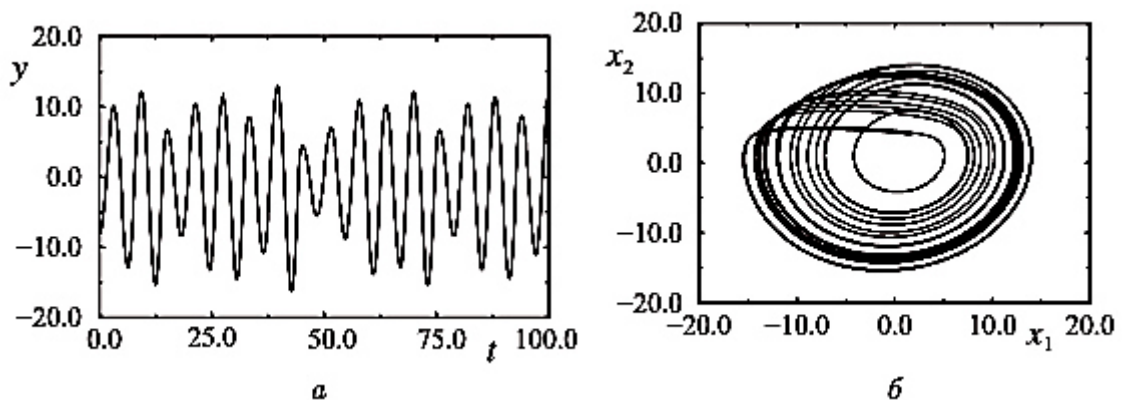


Рис. 5.30. Часова залежність координати $y(t)$ системи Рьослера (а) та реконструйований аттрактор в проекції на площину $(y(t), y(t + \tau))$ (б)

Для завдання вектора стану реконструйованої системи скористаємося теоремою Такенса (5.46). Розраховуючи по спостережуваній $a(t)$ автокореляційну функцію, знаходимо час спадання її до нуля $\tau_0 \approx 1.6$ і використаємо цю величину в якості часу затримки в (5.46). На рис. 5.30,б представлена проекція реконструйованого атрактора на площину двох змінних: $x_1(t) = y(t)$ і $x_2(t) = y(t + \tau)$.

Для визначення розмірності модельної системи потрібно розрахувати розмірність атрактора і розмірність простору вкладення. Для оцінки розмірності атрактора обчислимо його кореляційну розмірність D_c (5.53), використовуючи спеціальний алгоритм. На рис. 5.31 наведені результати розрахунку залежності D_c від $\lg \epsilon$, де ϵ – розмір осередку розбивки фазового простору. Як видно з графіків, незалежно від розмірності простору вкладення n , є «полічка» на рівні $D_c \approx 1.9$, який і приймаємо за значення шуканої розмірності.

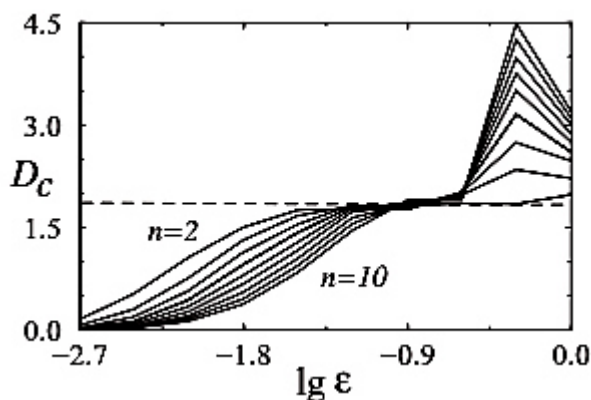


Рис. 5.31. Результати розрахунку кореляційної розмірності D_c при варіюванні розмірності простору вкладення n

Таким чином, реконструйований атрактор має розмірність $D \approx 2$ і може бути «вкладений» в тривимірний фазовий простір. Це означає, що ми можемо шукати модельну ДС у вигляді системи ЗДР третього порядку ($n=3$). Шукану систему запишемо у формі Коші, використовуючи поліноміальну апроксимацію (5.60) і обмежившись значенням $n=3$ і $\nu=2$.

Результати інтегрування модельної ДС представлені на рис. 5.32 у вигляді залежності $x_1(t)$.

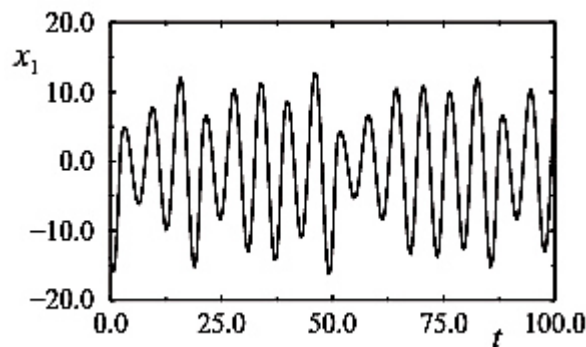


Рис. 5.32. Залежність $x_1(t)$, отримана чисельним інтегруванням реконструйованої системи

Порівняння даних рис. 5.32 з даними рис. 5.30,*a* показує якісну подібність реального і модельного коливальних процесів. Однак важливим, звичайно, є кількісні відповідності. Чи можливо за допомогою реконструйованої системи здійснювати прогноз еволюції системи в часі за межами інтервалу, на якому нам відома спостережувана? З цією метою проведемо наступний експеримент. Візьмемо в якості початкового значення координату останньої точки спостережуваної (рис. 5.32,*a*) в момент часу $t_0=100$. Далі інтегруємо як вихідну, так і модельну системи з початковими умовами при $t=t_0$ і порівняємо результати для $t>t_0$. На рис. 5.33 наведені відповідні графіки залежностей $y(t)$ для тестової системи (5.63) і $x_1(t)$ для реконструйованої ДС. Пунктирною лінією тут показано результат інтегрування системи (5.63), суцільною лінією – рішення реконструйованої модельної системи.

Як випливає з рис. 5.33, прогноз еволюції системи в часі здійснюється з деякою похибкою, яка з часом наростає. Конкретний час прогнозу можна вказати, задавши точність передбачення. З результатів рис. 5.33 випливає, що якщо обмежитися похибкою $\pm 5\%$, то час передбачення в нормованих одиницях становитиме приблизно $T=12$, тобто близько двох базових квазіперіодів коливальних системи.

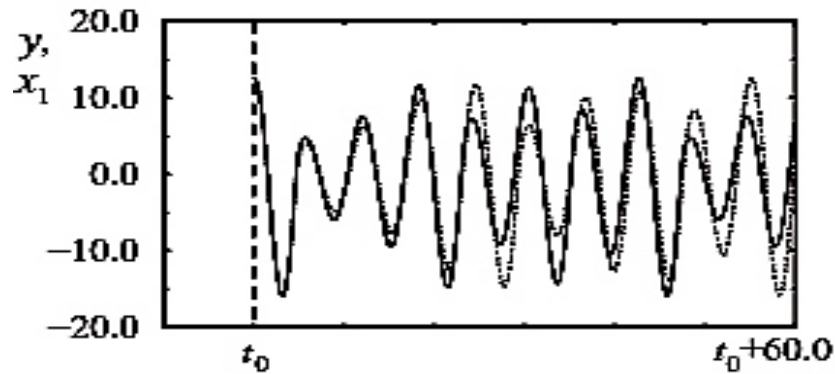


Рис. 5.33. Прогноз поведінки системи (5.62) після закінчення спостереження за сигналом, що генерується цією системою (час t_0)

Обраний для ілюстрації приклад є далеко не простим, оскільки реконструюються хаотичні автоколивання. З цим пов'язані і похибки реконструкції, що обмежують час передбачення. Якщо в якості спостерігаючої ми маємо більш простий тип руху (наприклад, періодичні коливання), реконструкція призводить до зменшення похибок і збільшення часу прогнозу.

Контрольні запитання

1. Визначення динамічної системи. Класифікація динамічних систем.
2. Що таке регулярні і хаотичні атрактори?
3. Як оцінити стійкість динамічної системи?
4. Що таке «м'які» і «жорсткі» біфуркації?
5. Розкрийте властивості детермінованість і хаос.
6. Розкрийте властивості стійкість і нелінійність.
7. Розкрийте властивості нестійкість і нелінійні обмеження.
8. Ймовірнісні властивості детермінованих систем.
9. Детермінований хаос – властивість матеріального світу.
10. Дивні хаотичні атрактори.
11. Дивні нехаотичні і хаотичні недивні атрактори.

12. Перехід до хаосу через послідовність біфуркацій подвоєння періоду.
13. Жорсткі переходи до хаосу. Криза і перемежана.
14. Перехід до хаосу через руйнування квазіперіодичних коливань.
15. Атрактори типу Лоренца.
16. Квазіатрактори і їх властивості.
17. Визначення розмірності вкладення і реконструкція атрактора.
18. Розрахунок старшого показника Ляпунова по часовому ряду.
19. Реконструкція динамічної системи.

Глава 6. ФРАКТАЛЬНИЙ АНАЛІЗ

6.1. Основні визначення

Порівняно давно в математиці виник образ об'єкта, більш об'ємного, але, тим не менш, схожого із лінією. Деяким вченим було важко примиритися з поняттям лінії, яка не має ширини, тому поступово ними стали вивчатися геометричні форми і структури, що мають дробову просторову розмірність. На зміну безперервним кривим, що мають усі свої похідні, прийшли ламані або дуже порізані криві. Яскравим прикладом такої кривої є траєкторія броунівської частки. Так в науці виникло поняття *фрактала*.

6.1.1. Поняття фракталів.

Фракталами називаються геометричні об'єкти: *лінії, поверхні, просторові тіла, що мають сильно порізану форму і мають властивість самоподібності*. Слово фрактал походить від латинського слова «fractus» і перекладається як дробовий, ламаний. Самоподібність як основна характеристика фрактала означає, що він більш-менш одноманітно влаштований в широкому діапазоні масштабів. Так, при збільшенні маленькі фрагменти фрактала виходять дуже схожими на великі.

З математичної точки зору *фрактал – це, перш за все, множина з дробовою розмірністю*. Фрактал за першим визначенням Мандельброта – *це множина, хаусдорфова розмірність якої перевершує її топологічну розмірність*. За другим визначенням фрактал – *це геометрична структура, частини (фрагменти) якої в якійсь мірі подібні самій структурі*. Можна також сказати, що математичне поняття фрактала виділяє об'єкти, що володіють структурами різних масштабів, як великими, так і малими, і, таким чином, він відображає ієрархічний принцип організації матерії в природі. В основі цього поняття міститься одна важлива ідеалізація дійсності: фрактальні об'єкти

самоподібні, тобто їх вигляд не зазнає істотних змін при розгляді їх під мікроскопом з будь-яким збільшенням.

Ми знаємо, що лінія має один вимір, поверхня двовимірна, а просторова фігура тривимірна. Фрактал же – це не лінія і не поверхня, а, якщо так можна висловитись, щось середнє. Розмірність об'єкта показує, за яким законом зростає його внутрішня область. Аналогічним чином з ростом розмірів зростає «обсяг» фрактала, але його розмірність – величина не ціла, а дробова. Тому межа фрактальної фігури не лінія: при великому збільшенні стає видно, що вона розмита і вся складається зі спіралей і завитків, які повторюють в малому масштабі саму фігуру. Така геометрична регулярність називається *масштабною інваріантністю* або *масштабною самоподібністю*, *скейлінгом* (від англ. *scaling*). Вона-то і визначає дробову розмірність фрактальних фігур. Рідина, газ, тверде тіло – три звичних для нас стани однорідної речовини, що існує в тривимірному світі. Але яка розмірність хмари або клубу диму, точніше їх кордонів, що розмиваються турбулентним рухом повітря? Виявилось, що вона більше двох, але менше трьох. Кровоносна система людини, наприклад, має розмірність близько 2,7. Усі об'єкти з нечіткою, невпорядкованою, хаотичною, зламанною структурою виявилися фракталами або такими, що складаються з фракталів.

Зауважимо, що зв'язок між хаосом і фракталами далеко не випадковий – він висловлює їх глибоку спільність. Фрактали, як геометричні поняття, можуть бути успішно застосовані при описі траєкторій хаотичних систем. Більш того, можна сказати, що фрактали – геометричні образи хаосу, а фрактальна геометрія – геометрія хаосу, безладу. Таким чином, ключовими поняттями теорії фракталів є дробова розмірність і масштабна самоподібність.

Термін «фрактал» введений бельгійським математиком Бенуа Мандельбротом і позначає множину, що має дробову фрактальну розмірність. Для пояснення *фрактальної розмірності* необхідно ввести поняття *топологічної розмірності*. Під топологічною розмірністю D_T множини в лінійному просторі розуміють число лінійно незалежних координат у просторі.

Наприклад, окружність і лінія мають топологічну розмірність 1; коло і квадрат – 2; куля і куб – 3. Фрактальна розмірність множини це D -розмірність того простору, яке повністю заповнюється множиною. Для зв'язку фрактальної і топологічної розмірності використовують показник Херста H , який обчислюється за формулою: $H=D-D_T$. Фракталом називають множину, фрактальна розмірність якої не співпадає з топологічною.

6.1.2. Класифікація фракталів.

Геометричні фрактали самі наочні. У двовимірному випадку їх отримують за допомогою деякої ламаної (або поверхні в тривимірному випадку), званої *генератором*. За один крок алгоритму кожен з відрізків, що становлять ламану, замінюється на ламану-генератор, у відповідному масштабі. В результаті нескінченного повторення цієї процедури, виходить геометричний фрактал.

Розглянемо тріадну криву Коха. Побудова кривої починається з відрізка одиничної довжини (рис.6.1) – це 0-е покоління кривої Коха. Далі кожна ланка (в нульовому поколінні один відрізок) замінюється на який *утворюючий елемент* ($n=1$ на рис.6.1) і виходить наступне покоління кривої Коха – це крива з чотирьох прямолінійних ланок, кожна довжиною по $1/3$. Отже, для отримання кожного наступного покоління, всі ланки попереднього покоління необхідно замінити зменшеним утворюючим елементом. Крива n -го покоління при будь-якому кінцевому n називається *предфракталом*. На рис. 6.1 представлені п'ять поколінь кривої. При $n \rightarrow \infty$ крива Коха стає фрактальним об'єктом.

Алгебраїчні фрактали – найбільша група фракталів. Отримують їх за допомогою нелінійних процесів в n -вимірних просторах. Найбільш вивчені двовимірні процеси. Інтерпретуючи нелінійний ітераційний процес, як дискретну ДС, можна користуватись термінологією теорії цих систем: *фазовий портрет, сталий процес, аттрактор* тощо.

Відомо, що нелінійні ДС мають декілька стійких станів (атракторів) в фазовому просторі. Якщо фазовим є двовимірний простір, то фарбуючи області тяжіння аттракторів різними кольорами, можна отримати *кольоровий фазовий*

портрет цієї системи (ітераційного процесу). Змінюючи алгоритм вибору кольору, можна отримати складні фрактальні картини з химерними кольоровими візерунками. Несподіванкою для математиків стала можливість за допомогою примітивних алгоритмів породжувати дуже складні нетривіальні структури.

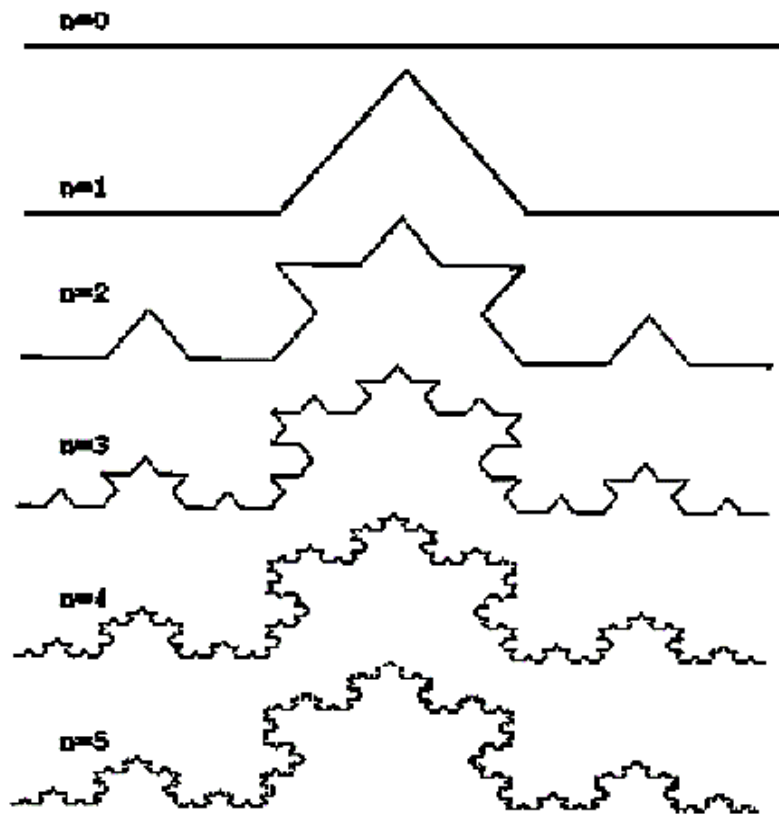


Рис. 6.1. Побудова тріадної кривої Коха

Як приклад розглянемо множину Мандельброта (рис. 6.2 і 6.3). Алгоритм її побудови досить простий і заснований на простому ітеративному виразі $Z[i+1]=Z[i] \cdot Z[i]+C$, де Z і C – комплексні змінні. Ітерації виконуються для кожної стартової точки C квадратної або прямокутної області – підмножині комплексної площини. Ітераційний процес продовжується до тих пір, поки $Z[i]$ не вийде за межі кола радіуса 2, центр якого лежить в точці $(0,0)$, або після досить великої кількості ітерацій $Z[i]$ зійдеться до будь-якої точки кола. В залежності від кількості ітерацій, протягом яких $Z[i]$ залишалась всередині

кола, можна встановити колір точки C (якщо $Z[i]$ залишається в колі протягом досить великої кількості ітерацій, ітераційний процес припиняється і ця точка растра забарвлюється в чорний колір).

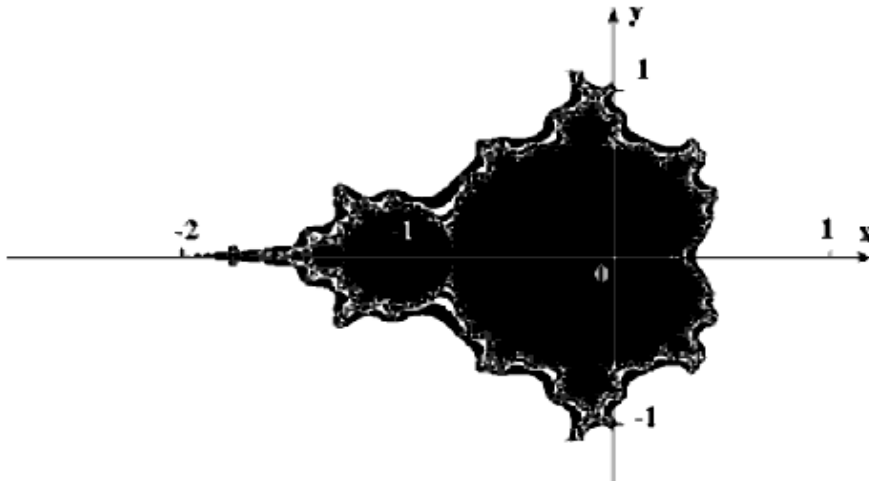


Рис. 6.2. Множина Мандельброта

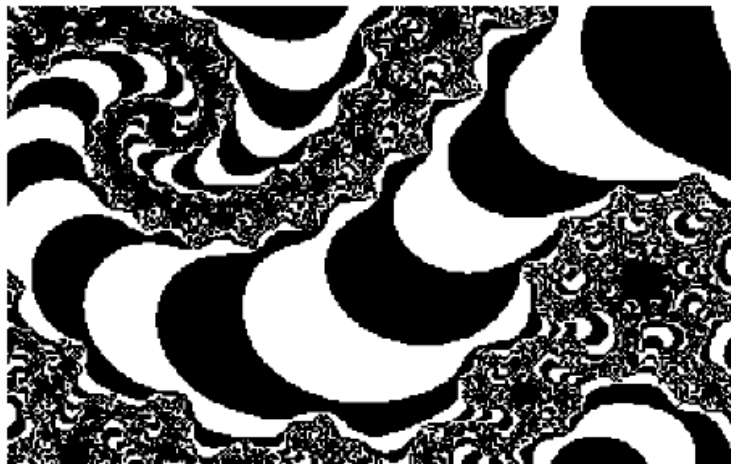


Рис. 6.3. Ділянка межі множини Мандельброта, яка збільшена в 200 разів

Вищеописаний алгоритм дає наближення до так званої множини Мандельброта. Йому належать точки, які протягом нескінченного числа ітерацій не йдуть в нескінченність (точки мають чорний колір). Точки, що належать межі множини (саме там виникають складні структури), йдуть в нескінченність за кінцеве число ітерацій, а точки, що лежать за межами множини, йдуть в нескінченність через декілька ітерацій (білий фон).

Стохастичні фрактали утворюються шляхом багаторазових повторень випадкових змін будь-яких параметрів. При цьому виходять об'єкти дуже схожі на природні – несиметричні дерева, порізані берегові лінії тощо. Двовимірні стохастичні фрактали використовуються при моделюванні рельєфу місцевості і поверхні моря.

Найбільш яскравими прикладами стохастичних фракталів є:

- траєкторія броунівського руху на площині і в просторі (доведено, що розмірність границі броунівського руху на площині дорівнює $4/3$);
- еволюції Шрамма-Льовнера – конформно-інваріантні фрактальні криві, що виникають в критичних двовимірних моделях статистичної механіки;
- різноманітні види рандомізованих фракталів, які утворюються за допомогою рекурсивної процедури, в яку на кожному кроці вводиться випадковий параметр. Дуже часто цей різновид фракталів використовують в комп'ютерній графіці для зображення плазми.

Існують і інші класифікації фракталів, наприклад розподіл фракталів на детерміновані (алгебраїчні і геометричні) і недетерміновані (стохастичні).

6.2.Стохастичні шуми

Чисто детерміністичний опис ДС є неповним, оскільки в будь-якій реальній дисипативній системі неминуче присутній шум. Тоді, дослідження системи передбачає вивчення властивостей ансамблю реалізацій і вимагає ймовірнісного підходу.

Причини, за якими повинна враховуватись стохастичність, дуже різноманітні. З одного боку, змінні ДС, як правило, описують властивості системи, що складається з великого числа мікрочастинок. Тому змінні на макрорівні постійно відчувають теплові флуктуації. До шуму в системі неминуче призводить кінцеве число макрозмінних, що враховуються при

математичному моделюванні ДС, а також квантовий характер мікрочасток. Шум, пов'язаний із зазначеними причинами, зазвичай називають *внутрішнім*.

З іншого боку, ДС описує реальний процес тільки на деякому певному рівні. При цьому проводять розмежування між системою і навколишнім середовищем. Тоді взаємодія системи з навколишнім середовищем описується за допомогою *зовнішнього* шуму, що впливає на систему.

Фазова траєкторія детермінованої ДС однозначно визначається початковим станом. У разі стохастичної поведінки стан $x(t)$ системи відображається в часі, наприклад, за допомогою стохастичних диференціальних рівнянь, які явно враховують випадкові джерела

$$\dot{x}_\xi(t) = f(x_\xi, \xi(t)). \quad (6.1)$$

Тут $\xi(t)$ – послідовність випадкових чисел. В результаті оператор еволюції в часі залежить від конкретного вибору $\xi(t)$, і, отже, є стохастичним (випадковим) процесом.

6.2.1. Рівняння Ланжевена.

Поль Ланжевен ввів в розгляд стохастичні диференціальні рівняння, додавши випадкову силу в рівняння руху броунівської частки. Ланжевен припустив рівність нулю середнього значення і кореляції між положенням частки і випадкової сили в один момент часу. У цих припущеннях він розглянув ансамбль систем і отримав звичайні диференціальні рівняння для середніх величин. В межі сильного тертя результати інтегрування виявили дифузійну поведінку частки, тобто лінійне збільшення середнього квадрата зміщення частки з плином часу. Базуючись на рівномірному розподілі енергії, Ланжевен вивів флуктуаційно-дісипаційні співвідношення, що зв'язують дифузію з коефіцієнтом тертя Стокса, які були отримані Ейнштейном трьома роками раніше на основі кінетичного рівняння для щільності ймовірності броунівських часток.

Нехай $x(t)=[x_1(t), \dots, x_n(t)]$ є вектором, складеним з компонент, які задають стан динамічної системи. У загальному вигляді рівняння Ланжевена може бути записано як:

$$\dot{x}_i(t) = f_i(x_1, \dots, x_n, t) + g_i(x_1, \dots, x_n, \xi_1(t), \dots, \xi_m(t), t). \quad (6.2)$$

У цьому рівнянні виділяються дві різні частини, що задаються у загальному випадку нелінійними функціями: детермінована і стохастична частини: $f(\mathbf{x}, t) = [f_1(\mathbf{x}, t), \dots, f_n(\mathbf{x}, t)]$ і $g(\mathbf{x}, \xi, t) = [g_1(\mathbf{x}, \xi, t), \dots, g_n(\mathbf{x}, \xi, t)]$, причому $g_i(\mathbf{x}, \xi = 0, t) = 0$. Багатовимірний випадок процесу $\xi(t) [\xi_1(t), \dots, \xi_m(t)]$ є сукупністю випадкових сил (шумів), що впливають на систему. Надалі вважається, що $\langle \xi_i(t) \rangle = 0$.

У разі внутрішнього шуму, який присутній в системі, що знаходиться в термодинамічній рівновазі, інтенсивності джерел шуму задовольняють особливим флуктуаційно-дісипаційним співвідношенням, що включає протидіючі флуктуаціям дисипативні сили.

Ситуація спрощується в разі зовнішнього шуму (який виникає поза системою): дві частини рівнянь Ланжевена можуть розглядатись як незалежні, і характеристики випадкових сил стають додатковими незалежними параметрами системи. Зазвичай використовується спрощення, що шум входить в рівняння Ланжевена лінійно:

$$\dot{x}_i(t) = f_i(x, t) + \sum_{j=1}^m g_{i,j}(x, t) \xi_j(t). \quad (6.3)$$

Крім того, якщо для даного j всі коефіцієнти $g_{i,j} = const$, то говорять про *адитивний шум*, оскільки інтенсивність впливу ξ_j на всі змінні стану $x_i(t)$ не залежить від миттєвих значень $x_i(t)$. Протилежна ситуація, коли $g_{i,j}$ залежить від змінних стану системи, називається *мультиплікативним шумом*.

6.2.2. Гаусові процеси.

Гаусові процеси представляють собою великий підклас випадкових процесів. Спільна щільність ймовірності задається розподілом Гауса

$$p_n(x_1, t_1, \dots, x_n, t_n) = \frac{\exp\left(-\frac{1}{2} \sum_{i,j=1}^n b_{i,j} (x_i - a(t_i))(x_j - a(t_j))\right)}{\sqrt{(2\pi)^n \text{Det}[R(t_i, t_j)]}}. \quad (6.4)$$

де

$$a(t_i) = \langle x(t_i) \rangle \quad (6.5)$$

є середнє значення, що у загальному випадку залежить від часу, а $R(t_i, t_j)$ – кореляційна матриця з елементами

$$R_{i,j} = \langle (x(t_i) - a(t_i))(x(t_j) - a(t_j)) \rangle = \langle x(t_i)x(t_j) \rangle - a(t_i)a(t_j) \quad (6.6)$$

$b_{i,j}$ – елементи оберненої матриці R^{-1} . Таким чином, для визначення гаусового процесу необхідно знати тільки два перших моменти випадкового процесу.

6.2.3. Вінерівський процес.

Броунівський рух можна моделювати гаусовим процесом. Якщо W_i є випадкова величина, що визначає положення броунівської частки в момент часу t_i , то її багатовимірний розподіл:

$$p_n(W_1, t_1, \dots, W_n, t_n) = \prod_{i=1}^n \frac{1}{\sqrt{4\pi D(t_i - t_{i-1})}} \exp\left(-\frac{1}{4D} \frac{(W_i - W_{i-1})^2}{t_i - t_{i-1}}\right), \quad (6.7)$$

де $W_0=0$ и $t_0=0$. Такий процес $W(t)$ називається *вінерівським процесом* на честь Норберта Вінера. Він генерує ансамбль безперервних траєкторій таких, що не диференціюються, але з ймовірністю 1. Очевидно, що даний процес є марковським, оскільки співмножники з $i \geq 2$ в добутку (6.7) є вірогідністю переходу між двома послідовними станами:

$$p_n(W_1, t_1 | W_{i-1}, t_{i-1}) = \frac{1}{\sqrt{4\pi D(t_i - t_{i-1})}} \exp\left(-\frac{1}{4D} \frac{(W_i - W_{i-1})^2}{t_i - t_{i-1}}\right). \quad (6.8)$$

Вінерівський процес не є стаціонарним ні в широкому, ні в строгому сенсі. При заданій початковій умові $p(W_0, t_0 = 0) = \delta(W_0)$ для $t \geq 0$ можна знайти щільність ймовірності:

$$p_1(W, t) = \int dW_0 p(W, t | W_0, t_0) \delta(W_0) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{W^2}{2Dt}\right), \quad (6.9)$$

яка явно залежить від часу. Перший момент вінерівського процесу з заданою початковою умовою тотожно дорівнює нулю: $\langle W(t) \rangle = 0$, але другий момент зростає лінійно з часом:

$$\langle W(t_1)W(t_2) \rangle = 2D_{\min}(t_1, t_2), \quad (6.10)$$

так що для $t_1 = t_2 = t$ маємо $\langle W^2(t) \rangle = 2Dt$. Таким чином, запроваджений параметр D визначає швидкість росту дисперсії. У разі броунівського руху ця швидкість співпадає з коефіцієнтом дифузії.

6.2.4. Білий гаусовий шум.

Вінерівський процес може бути представлений як сума послідовних незалежних прирощень:

$$W(t) = \sum_{k=0}^{n-1} \Delta W_{t_{k+1}-t_k}. \quad (6.11)$$

Виникає питання, якими мають бути властивості процесу $\xi(t)$, щоб можна було представити вінерівський процес у вигляді стохастичного інтеграла

$$W(t) = \int_0^t ds \xi(s). \quad (6.12)$$

Формально у цьому випадку $\xi(t)$ є похідною за часом від вінерівського процесу $\Delta t \rightarrow 0$. Однак ця похідна не існує в середньоквадратичній границі при $\Delta t \rightarrow 0$, оскільки гаусовий розподіл $\Delta W / \Delta t$ розходиться. Інтеграл від $\xi(t)$ за часом має сенс, тільки в тому випадку, якщо його замінювати сумами за прирощеннями, і використовувати властивості прирощень.

Тим не менш, використання $\xi(t)$ із зазначеними властивостями в рівнянні Ланжевена часто зустрічається у фізичній літературі. Обчислення, вироблені для стаціонарних в широкому сенсі процесів, визначають властивості, якими повинен володіти процес $\xi(t)$. Це процес, який має гаусовий розподіл із середнім значенням, рівним 0. Кореляційна (коваріаційна) функція (6.10) вінерівського процесу може бути отримана, якщо припустити, що коваріаційна функція процесу $\xi(t)$ має вигляд $c_{\xi\xi}(t_1 - t_2) = 2D\delta(t_1 - t_2)$. Отже, значення процесу $\xi(t)$ по (6.12) не корелюють в різні моменти часу, що відповідає незалежності збільшень. Спектр потужності $G_{\xi\xi}(\omega)$ постійний на всіх частотах: $G_{\xi\xi}(\omega) = 2D$. З цієї причини процес $\xi(t)$ називають *білим шумом*.

Білий шум не має будь-яких фізичних аналогів. При моделюванні реальних процесів білим шумом виникають певні математичні складнощі, але є й величезна перевага. Наприклад, тільки в тому випадку, коли випадковий компонент в рівнянні Ланжевена (6.4) є білим шумом, результуючий процес $x(t)$ буде марковським. *Це важливе спрощення втрачається для будь-яких джерел шуму, що відрізняються від моделі білого шуму.*

Можливе фізичне обґрунтування для використання білого шуму засноване на поділі масштабів часу випадкових збурень і чисто детермінованої динаміки.

Далі в цій главі будемо позначати гаусовий білий шум як $\xi(t)$ і вважати, що $\langle \xi(t)\xi(t + \tau) \rangle = \delta(\tau)$, а для прирощення вінерівського процесу будемо, відповідно, вважати $\langle \Delta W_{\Delta t}^2 \rangle = \Delta t$. Інтенсивність білого гаусового шуму, а також значення дисперсії прирощення можуть бути введені в явному вигляді, як мультиплікативний множник перед $\xi(t)$ або $\Delta W_{\Delta t}$.

6.2.5. Пуассонівський процес.

Нехай випадкова функція $N(t)$ визначає, скільки раз деяка випадкова подія відбудеться на інтервалі часу $[0, t]$, причому $N(0)=0$.

Ймовірність виявити стан N в момент часу t задається розподілом Пуассона:

$$P_N(t) = \frac{(\gamma t)^N}{N!} \exp(-\gamma t), \quad (6.13)$$

де $t \geq 0$. Такий процес $N(t)$ називається *пуассонівським*. Всі його властивості визначаються тільки значенням γ . Пуассонівський процес, очевидно, не є гаусовим. Його перший момент, а також усі більш високі моменти виражаються через γ . Розглянемо середнє і другий момент:

$$\langle N(t) \rangle = \gamma t, \quad \langle (N(t) - \langle N(t) \rangle)(N(t') - \langle N(t') \rangle) \rangle = \gamma \min(t, t'). \quad (6.14)$$

Розподіл Пуассона (6.13) є рішенням керуючого рівняння

$$\dot{P}_N(t) = -\gamma P_N(t) + \gamma P_{N-1}(t); \quad N \geq 1, \quad \dot{P}_0(t) = -\gamma P_0(t), \quad (6.15)$$

що описує однокроковий процес («процес народження») $N \rightarrow N+1$ з ймовірністю переходу в одиницю часу $W(N \rightarrow N+1) = \gamma$. Керуюче рівняння являє собою диференціальну форму рівняння Чепмена-Колмогорова для дискретної множини станів. Воно описує еволюцію в часі ймовірностей станів. Слід зазначити, що для отримання $P_N(t + \Delta t)$ потрібно знати тільки $P_N(t)$. Отже, пуассонівський процес $N(t)$ є марковським.

6.2.6. Білий дробовий шум.

Подібно вінерівському процесу представимо пуассонівський процес за допомогою інтеграла:

$$N(t) = \int_0^t ds \xi_{SN}(s). \quad (6.16)$$

Підінтегральна функція ξ_{SN} являє собою випадковий процес, званий *дробовим шумом* і що не є гаусовим. Дробовий шум являє собою послідовність δ -імпульсів, що виникають у випадкові моменти часу t_i :

$$\xi_{SN}(t) = \sum_i \delta(t - t_i). \quad (6.17)$$

Моменти часу t_i мають пуассонівський розподіл по i . Кожен дельтоподібний пік відповідає стрибку значення $N(t)$ на величину $\Delta N = 1$. Пуассонівський дробовий шум, також званий процесом Кемпбелла, є білим шумом.

6.2.7. Кольоровий шум.

Як уже зазначалося, білий шум є математичною абстракцією. Саме тому так важливо вивчити вплив кінцевих часів кореляції випадкових сил на керовану шумом ДС. Щоб уникнути проблем, які виникають з не марковськими процесами, можна уявити кольорове джерело шуму (тобто шум з кінцевою кореляцією) як результат фільтрації нормально розподіленого білого шуму.

Простий фільтр низьких частот описується наступним рівнянням Ланжевена:

$$\dot{y}(t) = -\frac{1}{\tau_c} y + \frac{\sqrt{2D}}{\tau_c} \xi(t), \quad (6.18)$$

з гаусовим білим шумом $\xi(t)$, що має нульове середнє значення і $\langle \xi(t) \xi(t + \tau) \rangle = \delta(\tau)$. Процес (6.18), називається *процесом Орнштейна-Уленбека*, який в асимптотичній границі є стаціонарним гаусовим процесом зі стаціонарною щільністю ймовірності

$$P(y) = \sqrt{\frac{\tau_c}{2\pi D}} \exp\left(-\frac{\tau_c}{2D} y^2\right), \quad (6.19)$$

автоковаріаційною функцією

$$c_{yy}(\tau) = \frac{D}{\tau_c} \exp\left(-\frac{\tau}{\tau_c}\right) \quad (6.20)$$

і спектром потужності

$$S_{yy} = \frac{2D}{1 + \omega^2 \tau_c^2}. \quad (6.21)$$

У багатьох публікаціях процес Орнштейна-Уленбека використовується в якості джерела кольорового шуму, що впливає на ДС.

6.3. Фрактальні шуми

6.3.1. Самоподібні степеневі закони.

Розглянемо однорідну степеневу функцію

$$f(x) = cx^\beta, \quad (6.22)$$

де c і β – постійні. При $\beta=1$ функція $f(x) = cx$ при $c<0$ описує відновлюючу силу лінійної пружини, а при $\beta=-2$ рівняння (6.22) стає законом всесвітнього тяжіння Ньютона $f(x) = cx^{-2}$. Ці прості степеневі закони, які в безлічі зустрічаються в природі, є, насправді самоподібними: якщо піддати x перетворенню подібності (помноживши його на деяку константу), то функція $f(x)$ як і раніше буде пропорційна x^β , хоча й з іншим коефіцієнтом пропорційності.

Таким чином, однорідні функції володіють цікавою властивістю *масштабної інваріантності*: при зміні масштабу вони відтворюють самі себе. Така інваріантність може пролити світло на деякі темні куточки фізики, біології та інших наук.

Масштабна інваріантність обумовлена тим, що однорідні степеневі закони не мають природних масштабів; в них немає місця характерній одиничній мірі (такій, як одинична довжина, одиниця часу або одинична маса). Тому такі закони називають *масштабно-незалежними*.

6.3.2. Фрактальний метод нормованого розмаху Херста (R/S-аналіз).

Стандартна гаусова статистика добре працює при деяких обмежуючих припущеннях. Центральна гранична теорема (закон великих чисел) стверджує, що по мірі проведення дедалі більшої кількості випробувань граничний розподіл випадкової системи буде мати нормальний розподілом. Досліджувані

події повинні бути незалежними та ідентично розподілені. Але що робити, якщо для системи не виконуються ці умови? На щастя, існує непараметрична методологія, відкрита ще у 1951 р. Х.Е. Херстом, знаменитим британським гідрологом. Він розробив метод нормованого розмаху (R/S -аналіз), який використовується для розрізнення випадкового часового ряду і фрактального ряду. Нижче опишемо цю методику на прикладі резервуара річки Ніл.

Протягом кожного проміжку часу t такий резервуар приймає приплив $\xi(t)$ з озера, в той час, як регульований обсяг води (стік) спускається з водосховища. Необхідно знайти потрібну кількість води в резервуарі, щоб щорічно можна було спускати з нього кількість води, яка дорівнює середньому припливу за цей період.

Середній приплив за період в τ років дорівнює

$$\langle \xi \rangle_r = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t). \quad (6.23)$$

Тоді $X(t)$ – накопичене відхилення припливу $\xi(t)$ від його середнього значення є сумою

$$X(t, \tau) = \sum_{u=1}^t (\xi(u) - \langle \xi \rangle_r). \quad (6.24)$$

Розмах відхилень буде визначатися як

$$R(\tau) = \max_{1 \leq t \leq \tau} X(t, \tau) - \min_{1 \leq t \leq \tau} X(t, \tau). \quad (6.25)$$

Стандартне відхилення можна отримати за формулою квадратного кореня з дисперсії

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum (\xi(t) - \langle \xi \rangle_r)^2}. \quad (6.26)$$

Як виявив Херст, для багатьох часових рядів, що спостерігаються, нормований розмах R/S дуже добре описується емпіричним співвідношенням у вигляді степеневого закону:

$$R / S = (a\tau)^H, \quad (6.27)$$

де H – показник Херста. Тут слід зазначити, що розмах іменується нормованим, оскільки він за задумом Херста повинен ділитися на квадратний корінь з дисперсії. Це дозволяє застосовувати метод до самих різних систем. Показник Херста є стійкою мірою деяких статистичних явищ, для яких дисперсія такою не є.

6.3.3. Класифікація фрактальних шумів.

В основі класифікації шуму лежить однорідний степеневий закон (6.22). Класифікацію проводять в залежності від значення показника ступеня β :

- $\beta=0$ – білий шум;
- $\beta=1$ – рожевий шум;
- $\beta=2$ – коричневий шум;
- $\beta=3$ – чорний шум.

У білому шумі будь-яке його значення в момент часу t абсолютно не залежить від свого минулого – воно завжди несподівано (рис. 6.4,*a*). Навпаки, в «коричневій» музиці (броунівський рух – це теж «коричневий» шум) тільки *інкременти* не залежать від свого минулого, в результаті чого шум виходить стомливо одноманітним.

На рис. 6.4,*б* представлений зразок шуму з гіперболічним степеневим спектром f^{-1} . Такі функції відомі також під назвою *рожевого* шуму, тому що вони займають проміжне положення між коричневим (броунівським) (f^{-2}) і білим шумом ($1/f^0$) (див., відповідно, рис. 6.4,*в* і рис. 6.4,*a*). Оскільки спектр потужності будь-якого шуму, який підкорюється однорідному степеневому закону (f^β), самоподібний, відповідна часова діаграма також повинна бути самоподібна. Дійсно, якщо масштаб уздовж вісі частот змінити в r раз, то за законом взаємності Фур'є масштаб уздовж вісі часу відповідної часової діаграми зміниться в $1/r$ раз. Зрозуміло що, в разі шуму (і інших ймовірнісних

явищ) самоподоба носить лише *статистичний* характер: збільшений фрагмент не є точною детермінованою копією форми сигналу до зміни масштабу.



Рис. 6.4. Часові діаграми шумів:

а – білий шум з $1/f^0$ -спектром потужності; *б* – «рожевий» шум з $1/f^1$ -спектром; *в* – «коричневий» шум з $1/f^2$ -спектром; *г* – чорний шум з $1/f^3$ -спектром

Рожевий, білий і коричневий шум є ідеальними зразками статистично самоподібних процесів. Явища, спектри потужності яких представляють собою однорідні степеневі функції, не мають власних масштабів часу і частоти: вони *масштабно-незалежні*. Тут немає таких понять як характеристичний час або характеристична частота: те, що відбувається в одному часовому або частотному інтервалі, відбувається при *будь-якому* масштабуванні часу або частоти. Якщо такі шуми записати на магнітну стрічку і програти на різних швидкостях, то звучати вони будуть однаково.

Степеневі закони аж ніяк не обмежені цілочисельними показниками, як у випадку білого, рожевого і коричневого шумів. У природі в достатку зустрічаються прості степеневі закони з дробовими показниками, що описують найрізноманітніші явища: розливи річок, розподіл галактик у Всесвіті тощо. З цих законів виділимо головне: часто складні функції двох і більше змінних поведуться поблизу «критичних точок» як прості степеневі закони. Наприклад, функцію двох змінних $f(x, y)$ дуже часто можна представити в наступному загальному вигляді:

$$f(x, y) = x^\alpha g\left(y / x^\beta\right), \quad (6.28)$$

де функція $f(x,y)$ замінена функцією тільки однієї змінної g . Для будь-якого інтервалу змінних, на якому функція g постійна, функція $f(x,y)$ наближено може бути представлена простим степеневим законом від x .

Подібне представлення через степеневі закони і їх показники – виявляється надзвичайно плідним при аналізі різних критичних явищ.

Виявляється, і рожевий, і чорний шуми поширені досить широко. Рожеві процеси виникають у багатьох фізичних ситуаціях і знаходять дивовижні естетичні застосування в музиці та інших видах мистецтва.

Чорні спектри описують розвиток в часі багатьох природних і протиприродних катастроф, таких як розливи річок, засухи, ринки з тенденцією до зниження курсів і різні аварійні ситуації – наприклад, перебої в подачі електроенергії. Через свої чорні спектри подібні неприємності нерідко трапляються по кілька разів поспіль.

6.3.4. Рожевий шум.

Рожевий шум, званий також $1/f$ -шумом, або *флікер-шумом*, має однакову потужність в смугах частот шириною в октаву або в будь-яких постійних інтервалах в *логарифмічній* частотній шкалі. Завдяки цій властивості він знаходить безліч застосувань. Так, рожевий шум є одним з улюблених тестових сигналів в дослідженнях слуху і акустичних дослідженнях в цілому, оскільки він близький до багатьох природних шумів. Рожевий шум має також властивість порушувати на приблизно рівних по довжині ділянках основної перетинки у нашому внутрішньому вусі рівні по амплітуді коливання, тим самим стимулюючи постійну кількість закінчень слухових нервів, що передають звукові сигнали в мозок. Таким чином, рожевий шум являє собою *психоакустичний* еквівалент білого шуму.

Рожевий шум зустрічається також в самих різних фізичних системах, у тому числі в напівпровідникових пристроях. Однією з причин повсюдності $1/f$ -шумів є те, що їх породжують паралельні релаксаційні процеси, які в достатку зустрічаються в природі. У релаксаційному процесі (електрони, замкнені між

стінок потенційної ями в напівпровіднику) замкнена частка переходить в збуджений стан, в якому і залишається протягом експоненційно розподіленого інтервалу часу з часом релаксації τ . Спектр потужності такого процесу $P_\tau(f)$ є не що інше, як лоренцева резонансна крива, центрована на частоті 0:

$$P_\tau(f) = \frac{4\tau P_0}{1 + (2\pi f\tau)^2}. \quad (6.29)$$

Повна потужність P_0 релаксаційного процесу, тобто інтеграл від $P_\tau(f)$ по всьому діапазону позитивних частот, не залежить від τ .

Багато фізичних, хімічних або біологічних систем мають не один релаксаційний час τ , а цілий спектр таких часів, залежних від значень енергії потенційних бар'єрів E , які протягом деякого часу утримують замкнену в потенційній ямі систему в збудженому стані. Відношення між часом релаксації і енергією бар'єру E має такий вигляд:

$$\tau = \tau_0 e^{E/kT}, \quad (6.30)$$

де T – абсолютна температура, а k – постійна Больцмана. Припустимо, що ці енергії рівномірно розподілені в інтервалі $[E_1, E_2]$. Тоді розподіл часу релаксації $p(\tau)$ може бути отриманий зі співвідношення (6.30) за допомогою елементарних правил перетворення ймовірностей. Він виявляється гіперболічним по τ :

$$p(\tau) = \frac{kT}{E_2 - E_1} \cdot \frac{1}{\tau}, \quad \tau_1 \leq \tau \leq \tau_2, \quad (6.31)$$

де $\tau_{1,2} = \tau_0 \exp(E_{1,2} / kT)$.

Накладення великої кількості незалежних релаксаційних процесів на спектри потужності типу (6.29) і часи релаксації, розподілені по формулі (6.31), призводить до спектру

$$P(f) = \frac{2kTP_0}{\pi(E_2 - E_1)f} [\arctg(2\pi f\tau_2) - \arctg(2\pi f\tau_1)], \quad (6.32)$$

де різниця в квадратних дужках, не дивлячись на дещо громіздкий вигляд, практично постійна в інтервалі частот

$$\frac{1}{\pi\tau_2} < f < \frac{1}{4\pi\tau_1}. \quad (6.33)$$

Тут необхідно підкреслити, що інтервал частот (6.33) може бути дуже широким і в численних ситуаціях *дійсно* буває досить широкий.

Розподіл часу релаксації за широкими діапазонами значень спостерігались в багатьох фізичних і біологічних явищах.

Перемежуваність, що виникає при дотичній біфуркації в логістичній параболі й інших ітерованих нелінійних відображеннях є також механізмом гіперболічної поведінки, що породжує широкодіапазонні $1/f$ спектри.

Порівняно простий метод генерування рожевого або $1/f$ шуму на комп'ютері полягає у тому, щоб скласти декілька релаксаційних процесів зі спектром потужності типу (6.29) (характеристика фільтра нижніх частот першого порядку) зі значеннями часу релаксації τ , що утворюють самоподібну прогресію з коефіцієнтом подібності 10 (або менше – для кращої збіжності). При такому підході достатньо лише трьох значень часу релаксації для того, щоб покрити частотний діапазон шириною майже в три десятих порядку (рис. 6.5).

Релаксаційний процес з дискретними значеннями часу x_n можна задавати за допомогою наявного в комп'ютері генератора випадкових чисел, який дозволяє отримувати незалежні випадкові числа r_n (нульові міри), що підставляються потім в рекурентне співвідношення

$$x_{n+1} = \rho x_n + \sqrt{1 - \rho^2} r_n, \quad x_0 = 0. \quad (6.34)$$

Тут ρ – необхідний коефіцієнт кореляції між сусідніми випадковими значеннями. З часом релаксації τ цей коефіцієнт пов'язаний співвідношенням $\rho = \exp(-1/\tau)$.

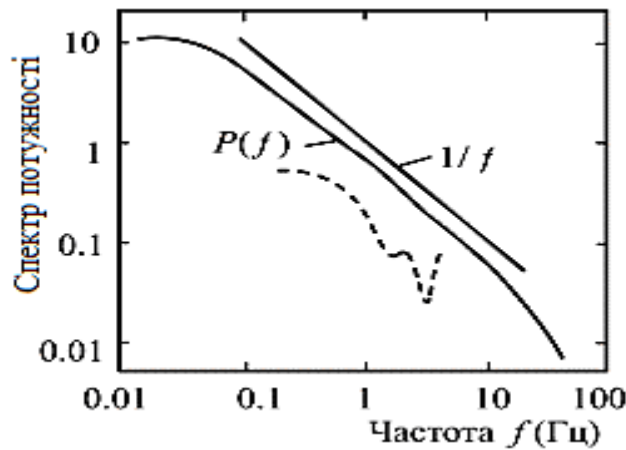


Рис. 6.5. Спектр потужності рожевого шуму:

суцільна крива отримана накладенням трьох релаксаційних процесів;
штрихова – накладенням значень, що випали на трьох гральних кістках

Таким чином, для набору значень часу релаксації, кожен з яких в 10 разів перевершує попередній ($\tau=1,10,100,\dots$), коефіцієнти кореляції отримують обчисленням послідовних коренів десятому ступені (тобто $\rho=0.37,0.9,0.99 \dots$).

Якщо не потрібна висока точність, то генераторами випадкових чисел можуть послужити три гральні кістки: першу кістку кидаємо для кожного нового значення рожевого шуму, другу кістку кидаємо через раз, а третю – через три рази на четвертий. Сума очок, що випали на всіх трьох кістках, утворює випадкову величину із середнім значенням 10,5 і дисперсією (потужністю шуму) 8,75, що є грубим наближенням рожевого шуму в обмеженому діапазоні частот.

6.3.5. Коричневий шум і самоподібні процеси на фондовій біржі.

Самоподоба між денними, тижневими і місячними цінами фондової біржі спостерігається аж до 30-ти секундних інтервалів. Час від часу в даних про коливання біржового курсу відбувається нехарактерний скачок. Спеціалісти називають такі скачки цін «інноваційними процесами».

Тенденції та флуктуації біржових курсів були свого часу надзвичайно детально проаналізовані з точки зору таких понять теорії інформації як перехресна ентропія і взаємна інформація.

Наразі аналіз стану ринку, поряд з іншими економічними додатками ентропійних принципів, належить до числа найбільш розроблених розділів теорії інформації. А оскільки біржові операції проводяться зараз на швидкодіючих машинах, керованих миттєвим зворотним зв'язком, виникла необхідність серйозного переосмислення правил усіма зацікавленими сторонами: правлінням біржі, аналітиками та інвесторами.

При розгляді біржових курсів в першому наближенні можна вважати, що їх реальні рівні складаються під впливом незалежних збільшень. Отриманий в результаті «ціновий шум» має спектр потужності, обернено пропорційний квадрату частоти. Такі випадкові процеси часто називають броунівськими (або коричневими) шумами через їх схожості з броунівським рухом – хаотичним мерехтінням зважених у воді часток квіткового пилку, відкритим під мікроскопом шотландським ботаніком Броуном. (В броунівському русі інноваційний процес складається з незалежних поштовхів, одержуваних зваженими частками з боку молекул тієї рідини, в якій вони плавають).

Іншим, ще більш чистим прикладом коричневого шуму можуть служити флуктуації капіталу гравця, для якого роль інноваційного процесу виконують незалежні кидки гральних кісток.

6.3.6. Чорні шуми і розливи Нілу.

Процеси з яскраво вираженою статистичною стійкістю ставлять перед дослідниками складні задачі, і пропоновані інтерпретації таких процесів часто виявляються невірними.

Амплітуди чорного шуму, загасають при великих частотах як f^{-3} . При малих частотах ми спостерігали б розбіжність спектра, що означало б

нескінченну енергію процесу. Однак, які не великі були б реальні катастрофи, їх енергія завжди кінцева. Кінцевий також і час спостереження T .

Тоді реалістичний спектр потужності $P(f)$ з асимптотичною залежністю f^{-3} , побудований за даними за період T , міг би виглядати наступним чином:

$$P(f) = \frac{T^4 f}{1 + T^4 f^4} \quad (f > 0). \quad (6.35)$$

Графік цього спектру для періоду спостережень $T=1$ (наприклад, рік) представлений на рис. 6.6. Припустимо, що спостереження тривають протягом двох років. «Новий» спектр потужності, побудований за розширеними спостереженнями, зображений на рис. 6.6 штриховою лінією. Таким чином, весь цей «річний» надлишок потужності, зображений на рис. 6.6 темним кольором, утворюється за допомогою простого продовження періоду спостережень з 1 до 2 років.

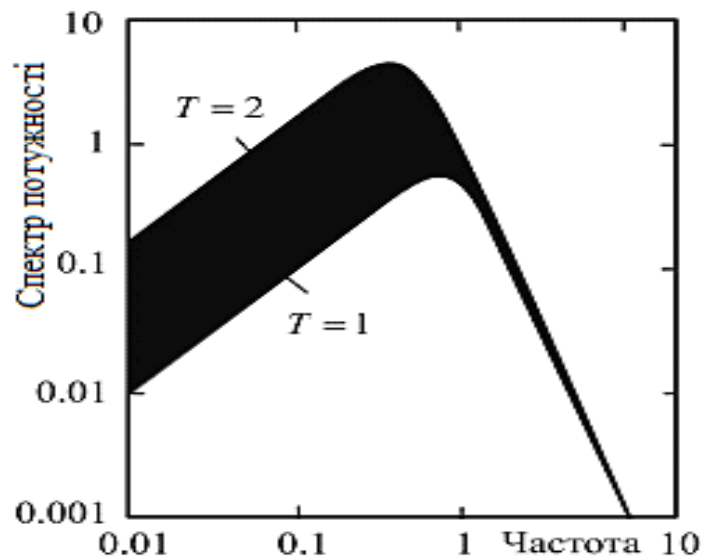


Рис. 6.6. Залежність спектра потужності від часу спостереження

Щоб охарактеризувати чорні шумові процеси, введемо нову міру розходження, яка була запропонована Харольдом Едвіном Херстом і Бенуа Мандельбротом. Величина, про яку йдеться, називається нормованим розмахом

R/S (що є розвитком методу R/S -аналізу Херста, розглянутого в п. 6.3.2) і по суті є розмахом $R(t)$ даних на часовому інтервалі t (після вирахування будь-якого лінійного тренда) поділений на стандартне відхилення вибірки $S(t)$.

Для білого гаусового шуму відношення R/S при великих t прагне до постійної. У певному сенсі і R , і S служать мірою розмаху даних, але R «розглядає» дані лінійно, а S – після зведення в квадрат. Для деяких процесів нормований розмах R/S не дає нової інформації і асимптотично прагне до постійної, тобто пропорційний Δt^0 . Однак в разі численних геофізичних записів, що відносяться до розливів річок, і безлічі інших даних відношення R/S поводиться інакше.

Для броунівської функції (спектр потужності якої пропорційний f^{-2}) величина R/S пропорційна $\Delta t^{1/2}$, що відображає довготривалу залежність, яка криється за коричневими процесами, та називається стійкістю.

Статистика коливань рівня води в Рейні (недалеко від Базеля) ось уже на протязі досить довгого терміну демонструє тенденцію до аналогічної поведінки з $R/S \approx \Delta t^{0,55}$. Наприклад, мінімальні рівні води в Нілі по записах за період з 622 по 1469 р. утворюють залежність $R/S \approx \Delta t^{0,9}$ (рис. 6.7) і показник ступеня 0.9 відображає високу ступінь стійкості.

Показник Херста, який визначається як

$$H = \ln(R/S) / \ln(\Delta t), \quad (6.36)$$

служить зручною мірою стійкості статистичного явища. У разі повної відсутності стійкості (білий шум) показник $H=-0.5$, тоді як при безсумнівній її наявності (коричневий шум) $H=0.5$.

Цікаво, що між показником Херста H і спектральним показником існує просте співвідношення: $\beta = 2H + 1$. Отже, «нільський шум» має спектр потужності, пропорційний $f^{-\beta} = f^{-2,8}$, що як і великий показник Херста $H=0.9$, передбачає довгострокову стійкість, при якій для стримування повеней і

запобігання руйнівним наслідкам необхідно споруджувати надзвичайно високі перепони (такі, як велика гребля в Асуані).

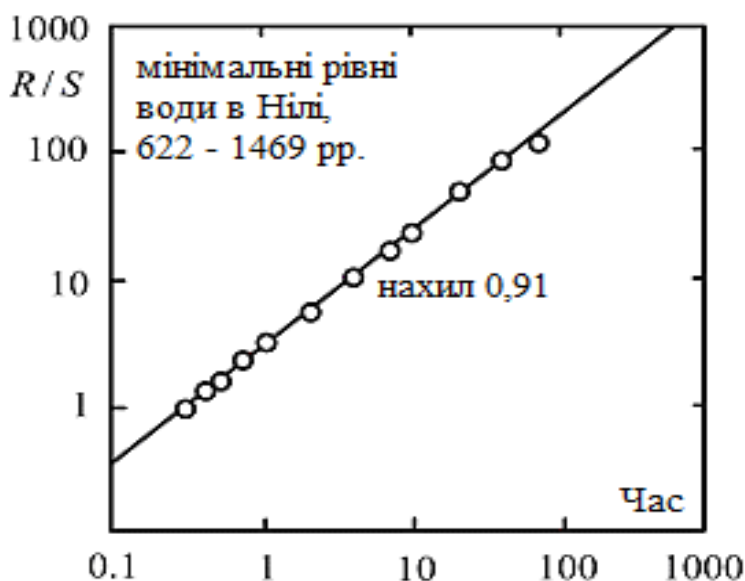


Рис. 6.7. Коливання виміряного мінімального рівня води в Нілі

6.4. Фрактальна розмірність

6.4.1. Загальне поняття фрактальної розмірності.

Ми схильні думати про всі об'єкти, які мають глибину, як про «тривимірні». З точки зору математики це невірно. Лінія, прокреслена в тривимірному просторі, має глибину, але ця лінія залишається одновимірною. Істинно тривимірний об'єкт – суцільне тіло, яке не має отворів або тріщин на своїй поверхні. Ось чому уявлення природних форм за допомогою евклідової геометрії є настільки важким. Більшість реальних об'єктів не суцільні в класичному, евклідовому сенсі – вони мають проломи в порожнині. Вони просто розташовуються в тривимірному просторі.

Нездатність евклідової геометрії описати більшість природних об'єктів обмежує нашу здатність зрозуміти те, як об'єкт влаштований. Для випадку часових рядів класична геометрія не може надати суттєвої допомоги в розумінні походження їх структури, якщо тільки це не випадкове блукання –

система настільки складна, що передбачити її поведінку неможливо. У термінах статистики число ступенів свободи, або факторів впливу на систему дуже велике.

Фрактальна розмірність, яка описує об'єкт (або часовий ряд) заповнює свій простір, є продуктом всіх тих факторів впливу на систему, які й породжують цей об'єкт (або часовий ряд). Часовий ряд буде тільки тоді випадковий, коли він є наслідком великої кількості рівно можливих подій. У термінах статистики – він має велику кількість ступенів свободи. Невипадковий часовий ряд буде відображати невідповідну природу впливів. Скачки даних будуть відповідати стрибкам факторів, що впливають, відображаючи притаманну їм кореляцію. Іншими словами, часовий ряд буде *фракталом*.

Зазвичай ми поміщуємо об'єкт в простір, більший ніж фрактальна розмірність цього об'єкта. Ми вважаємо, що кулька зім'ятого паперу є тривимірною, хоча вона і не заповнює весь відведений їй тривимірний простір. Це простір, що розглядається як об'єкт, називається *розмірністю вкладення*, або *топологічною розмірністю*. Коли об'єкти мають розмірність між двома і трьома, ми схильні думати про них як про тривимірні. Прикладами можуть служити гори і хмари.

Ми думаємо про берегову лінію як про двовимірну, в той час як в дійсності її розмірність менше. Часовий ряд відноситься до тієї ж категорії об'єктів. Тільки випадковий часовий ряд, який би суцільно покрити площину, був би істинно двовимірним.

Одна з характеристик фрактальних об'єктів полягає у тому, що вони залишають собі свою власну розмірність, будучи поміщені в простір розмірності, більше ніж їх фрактальна розмірність. Випадкові розподіли (білий шум) не мають цієї характеристики. Білий шум заповнює свій простір подібно до того, як газ заповнює обсяг. Якщо певну кількість газу помістити в контейнер більшого обсягу, газ просто розтечеться в більшому просторі, оскільки молекули газу ніщо не пов'язує між собою. З іншого боку, тверде тіло має молекули, зчеплені одна з одною. Аналогічно цьому у фрактальному

часовому ряді положення точок визначені кореляціями, але таких кореляцій не існує у випадковому ряді. У фракталі, подібному трикутнику Серпінського, кожна точка корельована з точкою, нанесеною до неї. Якщо ми збільшимо розмірність простору вкладення трикутника, то кореляції залишаться незмінними і будуть стягувати точки в групи. Розмірність трикутника залишиться незмінною, так само як залишилася б незмінною розмірність часового ряду.

У випадковому часовому ряді немає кореляцій точок. Ніщо не утримує точки в тому ж сусідстві, зберігаючи їх розмірність. Замість того вони цілком заповнюють відведений їм простір.

Фрактальна розмірність визначається тим, як об'єкт або часовий ряд заповнює простір. Фрактальний об'єкт заповнює простір нерівномірно, оскільки його частини залежні, або корельовані. Щоб визначити фрактальну розмірність, ми повинні визначити, яким чином об'єкт групується в єдине ціле в своєму просторі.

Берегові лінії є хорошим прикладом, особливо якщо провести паралель між ними та часовими рядами. Мандельброт висунув постулат про те, що ми ніколи не зможемо виміряти дійсну довжину берегової лінії, оскільки вимірювана довжина залежить від довжини використовуваної для вимірювання лінійки.

Припустимо, наприклад, що ми хочемо виміряти довжину узбережжя. Ми почнемо з самої північної точки і будемо міряти, накладаючи на поверхню землі лінійку метрової довжини. Ми будемо складати метрові збільшення, рухаючись вниз по березі, і прийдемо до якогось числа. Потім ми повторимо цю процедуру, використовуючи півметрову лінійку. На цей раз ми зможемо вловити більше деталей, оскільки наша лінійка коротше. Оскільки ми зможемо врахувати більшу кількість бухт і фіордів, ми в результаті отримаємо велику довжину узбережжя. Якщо ми вкоротити лінійку ще в два рази, то отримаємо ще більше деталей і ще більшу довжину. Чим коротше буде ставати лінійка,

тим довше берегова лінія. Виходить, що довжина берегової лінії залежить від розмірів лінійки!

З огляду на те, що це справедливо для усіх берегових ліній, довжина як міра не годиться для порівняння берегових ліній. Замість неї Мандельброт запропонував використовувати фрактальну розмірність. Берегові лінії являють собою зазубрені криві, тому їх фрактальна розмірність більше одиниці (тобто їх евклідової розмірності); те, наскільки вона більше одиниці, залежить від ступеня зазубреності. Чим вона більше, тим ближче розмірність берегової лінії до двох – розмірності площини.

Фрактальна розмірність розраховується за допомогою вимірювання цієї властивості зазубреності. Ми підраховуємо кількість кіл певного діаметра, яке необхідно для покриття берегової лінії. Ми збільшуємо їх діаметр і знову рахуємо їх кількість. Продовжуючи цю процедуру, ми знайдемо, що кількість кіл і їх радіус пов'язує показова залежність:

$$N(2r)^D = 1, \quad (6.37)$$

де N – кількість кіл, r – радіус кола, D – фрактальна розмірність.

Рівняння (6.37) може бути приведене до відношенню логарифмів

$$D = \frac{\log N}{\log(r/2)}. \quad (6.38)$$

Фрактальна розмірність показує нам, як форма або часовий ряд заповнюють простір. Спосіб заповнення об'єктом простору визначається тими силами, які визначили його формування. Для берегової лінії такими силами виступає геологічна активність, яка обумовлює її формування: тиск вітру, вулканічні явища тощо. Зауважимо, що метод кіл для визначення фрактальної розмірності незручний в практичному відношенні.

6.4.2. Узагальнені фрактальні розмірності.

Розглянемо фрактальний об'єкт, що займає певну обмежену область \mathfrak{Z} розміру L в евклідовому просторі з розмірністю d . Нехай на якомусь етапі його

побудови він представляє собою множину з $N > 1$ точок, яось розподілених в цій області. Ми будемо припускати, що в кінці кінців $N \rightarrow \infty$. Прикладом такої множини може служити трикутник Серпінського, побудований методом випадкових ітерацій. Кожен крок ітераційної процедури додає до цієї множини одну нову точку.

Розіб'ємо всю область на кубічні осередки зі стороною ε і обсягом d . Далі нас будуть цікавити тільки зайняті осередки, в яких міститься хоча б одна точка. Нехай номер зайнятих осередків i змінюється в межах $i=1,2,3,\dots,N(\varepsilon)$, де $N(\varepsilon)$ – сумарна кількість зайнятих осередків, яка, звичайно, залежить від розміру осередку.

Нехай $n_i(\varepsilon)$ є кількістю точок в осередку з номером i , тоді величина

$$p_i(\varepsilon) = \lim_{N \rightarrow \infty} \frac{n_i(\varepsilon)}{N} \quad (6.39)$$

є ймовірність того, що навмання взята точка з нашої множини знаходиться в осередку i . Іншими словами, ймовірності p_i характеризують відносну заселеність осередків. З умови нормування ймовірності випливає, що

$$\sum_{i=1}^{N(\varepsilon)} p_i(\varepsilon) = 1. \quad (6.40)$$

Введемо тепер в розгляд узагальнену статистичну суму $Z(q, \varepsilon)$, яка характеризується показником ступеня q який може приймати будь-які значення в інтервалі $-\infty < q < +\infty$:

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon). \quad (6.41)$$

Спектр узагальнених фрактальних розмірностей D_q , що характеризують даний розподіл точок в області \mathfrak{Z} , визначається за допомогою співвідношення:

$$D_q = \frac{\tau(q)}{q-1}, \quad (6.42)$$

де функція $\tau(q)$ має вигляд

$$\tau(q) = \lim_{\varepsilon \rightarrow 0} \frac{\ln Z(q, \varepsilon)}{\ln \varepsilon}. \quad (6.43)$$

Якщо $D_q = D = \text{const}$, тобто не залежить від q , то дана множина точок є звичайним, регулярним фракталом, який характеризується лише однією величиною – фрактальною розмірністю D . Навпаки, якщо функція D_q якось змінюється зі зміною q , то множина точок, яка розглядається, є *мультифракталом*.

Таким чином, *мультифрактал* у загальному випадку характеризується деякою нелінійною функцією $\tau(q)$, що визначає поведінку статистичної суми $Z(q, \varepsilon)$ при $\varepsilon \rightarrow 0$:

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon) \approx \varepsilon^{\tau(q)}. \quad (6.44)$$

У разі звичайного регулярного фрактала з фрактальною розмірністю D у всіх зайнятих осередках міститься однакова кількість точок

$$n_i(\varepsilon) = \frac{N}{N(\varepsilon)}, \quad (6.45)$$

тобто фрактал є *однорідним*. Тоді очевидно, що відносні населеності усіх осередків, $p_i(\varepsilon) = 1/N(\varepsilon)$, теж однакові, і узагальнена статистична сума набирає вигляду

$$Z(q, \varepsilon) = N^{1-q}(\varepsilon). \quad (6.46)$$

Тепер врахуємо, що згідно із визначенням фрактальної розмірності D , число зайнятих осередків при досить малому ε поводитьсь таким чином

$$N(\varepsilon) \approx \varepsilon^{-D}. \quad (6.47)$$

Підставляючи це в формулу (6.46) і порівнюючи з (6.44), ми приходимо до висновку, що у разі звичайного фрактала функція

$$\tau(q) = (q - 1)D \quad (6.48)$$

є лінійною. Тоді усі $D_q = D$ справді не залежать від q . Для фрактала, всі узагальнені фрактальні розмірності D_q якого збігаються, часто використовується термін *монофрактал*.

Якщо розподіл точок по осередках неоднаковий, то фрактал є неоднорідним, тобто являє собою мультифрактал, і для його характеристики необхідний цілий спектр узагальнених фрактальних розмірностей D_q , число яких, у загальному випадку, нескінченно.

Так, наприклад, при $q \rightarrow +\infty$ основний внесок в узагальнену статистичну суму (6.41) вносять осередки, які містять найбільшу кількість часток n_i в них й, отже, характеризуються найбільшою ймовірністю їх заповнення p_i . Навпаки, при $q \rightarrow -\infty$ основний внесок в суму (6.41) дають самі розріджені осередки з малими значеннями чисел заповнення p_i . Таким чином, функція D_q показує, наскільки неоднорідним є досліджувана множина точок \mathfrak{Z} . Надалі, для характеристики розподілу точок необхідно знати не тільки функцію $\tau(q)$, а й її похідну, яка безпосередньо обчислюється з виразів (6.43) і (6.42):

$$\frac{d\tau(q)}{dq} = \lim_{\varepsilon \rightarrow 0} \frac{\sum_{i=1}^{N(\varepsilon)} p_i^q \ln p_i}{\left(\sum_{i=1}^{N(\varepsilon)} p_i \right) \ln \varepsilon}. \quad (6.49)$$

Ця похідна, як буде показано далі, має важливий фізичний зміст. Тут же повторимо, що якщо вона не залишається постійною і змінюється з q , то це означає, що ми маємо справу з *мультифракталом*.

6.4.3. Основні властивості розмірності Хаусдорфа і інформаційна розмірність.

Розглянемо деякі основні фрактальні розмірності з ряду D_q при різних значення q . Для $q \rightarrow 0$

$$D_q = \lim_{\varepsilon \rightarrow 0} \frac{1}{q-1} \frac{\ln \left(\sum_{i=0}^{N(\varepsilon)} p_i^q \right)}{\ln \varepsilon}. \quad (6.50)$$

Звідки отримаємо

$$D_0 = - \lim_{\varepsilon \rightarrow 0} \left(\ln \left(\sum_{i=0}^{N(\varepsilon)} 1 \right) \right) / \ln \varepsilon = - \lim_{\varepsilon \rightarrow 0} \left[\frac{\ln N(\varepsilon)}{\ln \varepsilon} \right]. \quad (6.51)$$

З очевидної подібності цього виразу і формули для визначення розмірності Хаусдорфа випливає, що нуль-розмірність D_0 є не що інше як ємність атрактора $D_0=D$, тобто розмірність Хаусдорфа. При $q \rightarrow 1$ з (6.50) випливає, що

$$D_1 = - \lim_{\varepsilon \rightarrow 0} \left[\frac{K(\varepsilon)}{\ln \varepsilon} \right], \quad (6.52)$$

де

$$K_1 = - \sum_{i=0}^{N(\varepsilon)} p_i \ln p_i. \quad (6.53)$$

Тут K_1 – приріст інформації або *інформаційна ентропія*, отримана з відомих значень $\{p_i\}$, які свідчать про проходження фазової траєкторії через осередок i . Тому величину D_1 називають *інформаційною розмірністю*. Інформаційна розмірність показує, як зростає отримана інформація при $\varepsilon \rightarrow 0$. Насправді, K_1 є окремим випадком узагальненої ентропії, яка визначається з виразу

$$K_q = \frac{1}{q-1} \ln \left(\sum_{i=1}^{M(\varepsilon)} p_i^q \right), \quad (6.54)$$

для $q=1$ відповідно. Ентропія K_1 і інформаційна розмірність D_1 грають важливу роль при аналізі нелінійних ДС, особливо в описі втрати інформації при еволюції хаотичної системи в часі. В даному контексті величина K_1 називається *ентропією Колмогорова (K-ентропія)*.

На підставі аналізу виразів (6.50)-(6.53) можна зробити деякі висновки. Наприклад, для однорідних атракторів, коли всі p_i рівні між собою, тобто $p_i=1/N(\varepsilon)$, отримаємо

$$K_1 = - \sum_{i=1}^{N(\varepsilon)} \frac{1}{N(\varepsilon)} \ln \left(\frac{1}{N(\varepsilon)} \right) = \ln N(\varepsilon), \quad (6.55)$$

тобто інформаційна розмірність D_1 дорівнює розмірності Хаусдорфа D_0 . Таким чином, різниця D_0-D_1 є мірою неоднорідності дивного атрактора. Крім того, оскільки величина $K_1 \leq \ln N(\varepsilon)$, інформаційна розмірність завжди менше або дорівнює хаусдорфовій. Ці висновки можна узагальнити наступним правилом:

$$D_{q'} \leq D_q \text{ для } q' > q, \quad (6.56)$$

причому рівність справедлива тільки для однорідного атрактора.

6.4.4. Кореляційна розмірність.

Кореляційна розмірність D_2 може бути обчислена за формулою (6.51) для $q=2$:

$$D_2 = \lim_{\varepsilon \rightarrow 0} \frac{1}{q-1} \frac{\ln \left(\sum_{i=0}^{N(\varepsilon)} p_i^2 \right)}{\ln \varepsilon}. \quad (6.57)$$

Іноді, покриття фазового простору множиною багатовимірних осередків представляється досить непростим завданням. Однак для окремого випадку $q=2$ застосовується кореляційний інтеграл

$$C(\varepsilon) = \lim_{N \rightarrow \infty} \frac{1}{N(N-1)} \sum_{i,j=1}^N \theta \left[l - \left| \mathbf{x}_i - \mathbf{x}_j \right| \right], \quad (6.58)$$

де θ – функція Хевісайда, яка показує чи розділені точки \mathbf{x}_i і \mathbf{x}_j відстанню більшою за 1. Визначення кореляційної розмірності D_2 за допомогою кореляційного інтеграла (6.58) називають *методом Грасбергера-Прокаччіа*.

6.4.5. Розмірність подібності і її експериментальне визначення.

Безумовною цінністю розмірності Хаусдорфа є можливість її експериментального визначення. Деяка множина може бути виміряна d -вимірними (d – ціле) зразками зі стороною ε_1 . Тоді, кількість зразків N_1 , які покривають множину буде: $N_1 = A / \varepsilon_1^d$. Значення d має ґрунтуватися на попередніх відомостях про розмірності множини, теоретично, якщо d буде менше топологічної розмірності, то $N_1 \neq 0$, а якщо $d > R^n$, де R^n – евклідов простір, то $N_1 \rightarrow 0$.

Зразок з розміром ε_2 дасть оцінку $N_2 = A / \varepsilon_2^d$, тоді розмірністю подібності буде:

$$D = -\log_{\varepsilon_2 / \varepsilon_1} \frac{N_2}{N_1}. \quad (6.59)$$

Наприклад, лінія довжиною $A=100$ одиниць може бути покрита $N=10$ відрізками ε_2 довжиною 10 одиниць, або 100 відрізками ε_1 довжиною в 1 одиницю довжини. Тоді, розмірність лінії, згідно (6.59)

буде: $D = -\log_{10/1} \frac{10}{100} = 1$.

Прямокутник площею $A=64$ одиниці, може бути покритий 16-ю майданчиками ε_2^2 площею 4 одиниці (сторона 2) або 64 майданчиками ε_1^2 площею в 1 одиницю (сторона 1). Тоді, розмірність прямокутника

буде $D = -\log_{2/1} \frac{16}{64} = 2$.

Слід зазначити, що використання розмірності подібності зручно в разі об'єктів, які мають явну масштабну подобу. Такими об'єктами, перш за все, є само подібні фрактали – крива Коха, множина Мандельброта, Жюліа та інші. В тих же випадках, коли подобу важко встановити навіть при її наявності, зручно застосовувати наближене обчислення розмірності Хаусдорфа. Тут, як приклади, можна розглянути стохастичні само афінні фрактали, динамічні відображення – дивні атрактори тощо.

Для наближеного вимірювання розмірності Хаусдорфа, на початковій множині (вибірці) $x(t)$ встановлюється деяка «міра» – наприклад, довжина графіка вибірки. Далі вибірку треба виміряти за допомогою зразка фіксованої довжини δ (рис. 6.8). Вибірка, в даному випадку повинна бути упорядкованою.

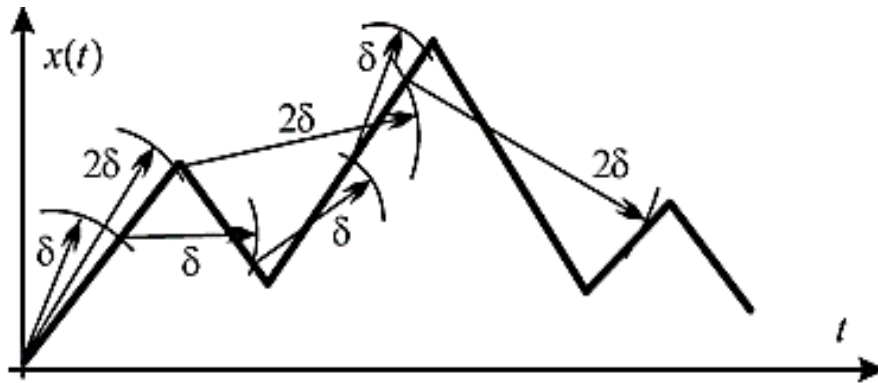


Рис. 6.8. Експериментальне вимірювання розмірності Хаусдорфа

(При аналізі зображень і сигналів, вибірки впорядковані або за координатами, або за часом).

Далі потрібно виміряти вибірку зразком довжини в 2δ . У підсумку ми отримаємо дві оцінки довжини A на «масштабі» δ – у вигляді A_1 і на «масштабі» 2δ – у вигляді A_2 .

Розмірність кривої на рис. 6.8 обчислюється за формулою:

$$D_{x(t)} = \log_2 \frac{A_1}{A_2}. \quad (6.60)$$

Якщо довжина вимірюється тільки цілим числом зразків, то розмірність може бути обчислена за формулою (6.59), що є аналогом фізичної подібності у відрізках.

Практична реалізація описаного методу, наштовхується на труднощі, обумовлені великим обсягом обчислень. Пов'язано це з тим, що для обчислення розмірності Хаусдорфа, потрібно вимірювати не просто співвідношення, а верхню границю цього співвідношення.

6.4.6. Визначення фрактальної розмірності фазової траєкторії і поверхні.

Фрактали можна розглядати як множини точок, вкладені в простір. Наприклад, множина точок, що утворюють лінію в звичайному евклідовому просторі, має топологічну розмірність $D_T=1$ і розмірність Хаусдорфа $D=1$. Евклідова розмірність простору дорівнює $E=3$. Оскільки для лінії $D=D_T$, то лінія, згідно з визначенням Мандельброта, не фрактальна, що підтверджує розумність визначення. Аналогічно множина точок, що утворюють поверхню в просторі з $E=3$, має топологічну розмірність $D_T=2$ і $D=2$. Ми бачимо, що і звичайна поверхня не фрактальна незалежно від того, наскільки вона складна. Нарешті, куля, або повна сфера, має $D=3$ і $D_T=3$. Ці приклади дозволяють визначити деякі з розглянутих нами типів множин.

Центральне місце у визначенні розмірності Хаусдорфа і фрактальної розмірності D займає поняття відстані між точками у просторі. Простий спосіб виміряти довжину кривих, площу поверхонь або обсяг тіла полягає у тому, щоб розділити простір на невеликі куби з ребром δ , як показано на рис. 6.9.

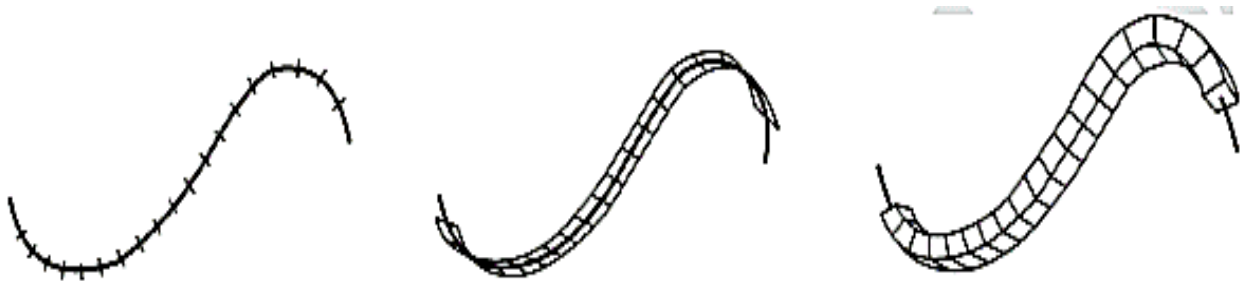


Рис. 6.9. Вимірювання «величини» кривої

Замість кубів можна було б взяти невеликі сфери діаметром δ . Якщо помістити центр малої сфери в якій-небудь крапці множини, то всі крапки, що знаходяться від центру на відстані $r < \delta / 2$, виявляться покритими цією сферою. Підраховуючи число сфер, необхідних для покриття множини точок, яка нас цікавить, ми отримуємо міру величини множини. Криву можна виміряти, визначаючи число $N(\delta)$ прямолінійних відрізків довжини, необхідних для того,

щоб покрити її. Зрозуміло, для звичайної кривої $N(\delta) = L_0 / \delta$. Довжина кривої визначається граничним переходом

$$L = N(\delta)\delta \xrightarrow{\delta \rightarrow 0} L_0\delta^0. \quad (6.61)$$

В границі при $\delta \rightarrow 0$ міра L стає такою, що асимптотично дорівнює довжині кривої і не залежить від δ .

Множині точок можна поставити у відповідність і *площу*. Наприклад, площу кривої можна визначити, вказуючи число кіл або квадратів, необхідних для її покриття. Якщо $N(\delta)$ – число цих квадратів, а δ^2 – площа кожного з них, то площа A кривої дорівнює

$$A = N(\delta)\delta^2 \xrightarrow{\delta \rightarrow 0} L_0\delta^1. \quad (6.62)$$

Аналогічно *обсяг* кривої V можна визначити як величину

$$V = N(\delta)\delta^2 \xrightarrow{\delta \rightarrow 0} L_0\delta^2. \quad (6.63)$$

Зрозуміло, що для звичайних кривих A і V обертаються в нуль при $\delta \rightarrow 0$ і єдиною мірою, що представляє інтерес, є довжина кривої.

Розглянемо далі множину точок, що утворюють поверхню (рис. 6.10).

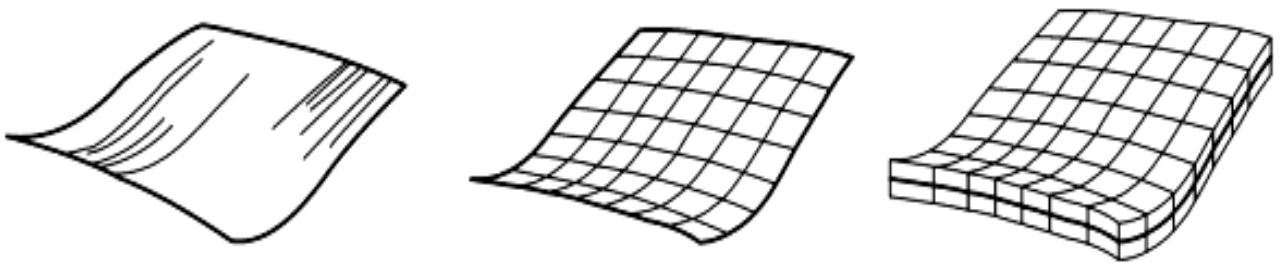


Рис. 6.10. Вимірювання «величини» поверхні

Нормальною мірою такої множини служить площа A , і ми маємо

$$V = N(\delta)\delta^2 \xrightarrow{\delta \rightarrow 0} A_0\delta^0. \quad (6.64)$$

Як неважко бачити, для звичайної поверхні число квадратів, необхідних для її покриття, визначається в границі при $\delta \rightarrow 0$ виразом

$$N(\delta) = A_0 / \delta^2, \quad (6.65)$$

де A_0 – площа поверхні.

Поверхні можна поставити у відповідність *обсяг*, утворюючи суму обсягів кубів, необхідних для покриття поверхні:

$$V = N(\delta)\delta^3 \xrightarrow{\delta \rightarrow 0} A_0\delta^2. \quad (6.66)$$

При $\delta \rightarrow 0$ цей обсяг, як і слід було очікувати, обертається в нуль.

Чи можна поверхні поставити у відповідність якусь довжину? Формально ми можемо прийняти за таку довжину величину

$$L = N(\delta)\delta \xrightarrow{\delta \rightarrow 0} A_0\delta^1, \quad (6.67)$$

яка розходитья при $\delta \rightarrow 0$. Цей результат має сенс, оскільки поверхню неможливо покрити кінцевим числом прямолінійних відрізків. Ми робимо висновок, що єдиною змістовною мірою множини точок, що утворюють поверхню в тривимірному просторі, є площа.

6.4.7. 'Визначення фрактальної розмірності для часових рядів.

Розглянемо деякі особливості обчислення фрактальних розмірностей за часовою вибіркою сигналу.

У чисельних, і у фізичних експериментах фрактальну розмірність і показники Ляпунова найчастіше знаходять, дискретизуючи сигнали послідовністю рівновіддалених (за часом) точок і обробляючи отримані дані на комп'ютері. Існує 3 основні методи:

- 1) часові дискретизації змінних у фазовому просторі;
- 2) обчислення фрактальної розмірності відображень Пуанкаре;
- 3) побудова псевдофазового простору за вимірюваннями однієї змінної (метод вкладення простору).

У першому і третьому методах змінні вимірюються через однакові інтервали часу $\{x(t_0 + n\tau)\}$, $n = 0, 1, 2, \dots$. Часовий інтервал τ обирається з таким розрахунком, щоб він становив певну частку періоду змушуючої сили або

характерного часу траєкторії. Якщо перетин Пуанкаре в методі 2) проводиться за часовою змінною, то τ – період траєкторії. У пристроях імпульсної силової електроніки, де є генератор синхроімпульсів, доцільно приймати $\tau=T$, де T – період роботи силового ключа. Якщо ж перетин Пуанкаре проводиться по яким-небудь іншим змінним у фазовому просторі, то зібрані дані відповідають різним моментам часу в залежності від конкретного типу вибраного перерізу Пуанкаре.

Основу чисельного знаходження фрактальної розмірності можна продемонструвати на наступному прикладі. Припустимо, що ми підозрюємо про існування у хаотичної системи атрактора в тривимірному фазовому просторі з фізичними змінними $\{x(t), y(t), z(t)\}$. Кожному інтервалу часу відповідає точка $\mathbf{x}=\{x(n\tau), y(n\tau), z(n\tau)\}$ в фазовому просторі.

Для обчислення усередненої поточної розмірності обирають випадковим чином декілька точок \mathbf{x}_n . Для кожної обраної точки обчислюють відстані до найближчих (в часі) оточуючих точок $\mathbf{x}_{n+1}, \mathbf{x}_{n-1}$. Наприклад, можна скористатися сумою абсолютних величин компонент вектора $\mathbf{x}_n - \mathbf{x}_m$, тобто

$$s_{nm} = |x(n\tau) - x(m\tau)| + |y(n\tau) - y(m\tau)| + |z(n\tau) - z(m\tau)|. \quad (6.68)$$

Далі підраховують число точок в кулі, кубі або іншому геометричному тілі (або фігурі) розміру ε і знаходять ймовірність як функцію параметра ε :

$$P_n(\varepsilon) = \frac{1}{N_0} \sum_m \theta(\varepsilon - s_{nm}), \quad (6.69)$$

де N_0 – загальне число точок у вибірці, θ – функція Хевісайда. Тоді усереднена поточкова розмірність є величина

$$d_n = \lim_{\varepsilon \rightarrow 0} \frac{\log P_n(\varepsilon)}{\log \varepsilon}, \quad d = \frac{1}{M} \sum_{n=1}^M d_n, \quad (6.70)$$

де границя у формулі для d_n існує. Для деяких атракторів ймовірність P_n залежить від ε не по степеневому закону, а розривно, або має злами. У цих випадках можна обчислювати модифіковану поточкову розмірність,

попередньо усереднюючи («згладжуючи») P_n . Наприклад, можна вибрати розмірність

$$C(\varepsilon) = \frac{1}{M} \sum_{n=1}^M P_n(\varepsilon), \quad d = \lim_{\varepsilon \rightarrow 0} \frac{\log C(\varepsilon)}{\log \varepsilon}, \quad (6.71)$$

яка аналогічна кореляційній розмірності.

Найменше значення ε має бути таким, щоб сфера радіуса ε або куб з ребром ε містили принаймні одну вибірку точку. Наприклад, в тривимірному фазовому просторі, якщо середній глобальний масштаб атрактора дорівнює L , то середня щільність точок становить величину

$$\rho \approx \frac{N_0 \cdot 6}{\pi L^3}, \quad (6.72)$$

тому обсяг, пов'язаний з масштабом ε , повинен бути більше, ніж ρ^{-1} або

$$\varepsilon > \frac{L}{2\sqrt[3]{N_0}}. \quad (6.73)$$

Іншим обмеженням на мінімальну величину ε є рівень «реального шуму», або невизначеність у вимірах змінних стану (x, y, z) . У будь-якому реальному експерименті існує сфера невизначеності, що оточує кожен вимірюваний пункт в фазовому просторі. Коли ε стає радіусом цієї сфери, розглянута вище теорія фрактальної розмірності, строго кажучи, стає непридатною, оскільки при менших ε не можна очікувати самоподібної структури.

6.4.8. Дисперсійна розмірність.

Даний вид розмірності зручно використовувати в разі стохастичних структур. Як обґрунтування даного методу використовується той факт, що спектр потужності $S(\omega)$ фрактальних сигналів має наступний вигляд:

$$S(\omega) \approx \omega^{-\beta}. \quad (6.74)$$

Якщо процес фрактальний і вкладений в евклідовий простір з розмірністю D_0 , то він буде мати розмірність $D = D_0 + \frac{3-\beta}{2}$, звідки $\beta = 3 + 2(D_0 - D)$.

Помічаючи, що β залежить тільки від D , можна для зручності переписати (6.74):

$$S(\omega) \approx \omega^{-D'}, \quad (6.75)$$

де D' – оцінка фрактальної розмірності.

Розглянемо випадковий процес $X(t)$ зі спектром потужності $S(\omega)$ при $0 \leq \omega \leq \Omega$, $\Omega < \infty$. Згідно з теоремою Вінера-Хінчіна маємо

$$\int_0^{\infty} S(\omega) d\omega = \int_0^{\Omega} S(\omega) d\omega = C_1(0) = \sigma_1^2. \quad (6.76)$$

Нехай згідно (6.74) вважаємо $S(\omega) \approx \omega^{-D'}$. Тоді

$$\sigma_1^2 = \frac{1}{-D'+1} \omega^{(-D'+1)} \Big|_0^{\Omega}. \quad (6.77)$$

Якщо зробити фільтрацію процесу X таким чином, щоб його смуга складала $\Omega/2$, то отримаємо:

$$\sigma_2^2 = \frac{1}{-D'+1} \omega^{(-D'+1)} \Big|_0^{\Omega/2}. \quad (6.78)$$

Не звертаючи увагу на ділення на 0 при $D' > 1$, знайдемо відношення

$$\frac{\sigma_1^2}{\sigma_2^2} = 2^{(-D'+1)}, \quad (6.79)$$

звідки

$$D' = 1 - \log_2 \frac{\sigma_1^2}{\sigma_2^2}, \quad (6.80)$$

або, якщо спектр обрізується з масштабу m до масштабу n , то

$$D' = 1 - \frac{\log \sigma_1^2 - \log \sigma_2^2}{\log m - \log n}. \quad (6.81)$$

Логарифм в (6.81) може бути взятий за будь-якою основою. Критерієм є тільки зручність обчислення. Далі можна визначити *дисперсійну фрактальну розмірність* за формулою:

$$D_{\sigma} = D_0 + \frac{3 - D'}{2}. \quad (6.82)$$

Порівняємо оцінки фрактальної розмірності вінерівського випадкового процесу – процесу з нормальними стаціонарними прирощеннями і розглянутого в главі 5 відображення Ено.

З математичної точки зору броунівський рух – безперервний гаусівський випадковий процес, $X = X(t)|_{t \geq 0}$, $X_0 = 0$, з нульовим середнім і дисперсією $\sigma^2(X_t) = t$. Автокореляційна функція його збільшень – δ -функція Дірака, що означає повну відсутність кореляцій в послідовних значеннях прирощень величини X_t і сталість спектра на всіх частотах $f(\omega) = \text{const}$ (ω – частота). Спектр такого виду називається «білим шумом» і успішно застосовується для моделювання багатьох кліматичних і гідрологічних процесів.

Однак спроба його використання для пояснення ефекту Херста зазнала невдачі. Не рятує положення й застосування випадкових процесів з кінцевим, не нульовим, як в процесі Вінера, часом кореляції. А.М. Колмогоров у 1940 р. в двох своїх роботах вперше розглянув випадкові процеси з дисперсією $\sigma^2(X_t) = t^\gamma$, $0 \leq \gamma \leq 2$ і назвав їх спіралями Вінера. Це стало початковим узагальненням вінерівського процесу. У 1968 р. Б. Мандельброт ввів поняття узагальненого броунівського руху. Безперервний гаусівський випадковий процес називається узагальненим (фрактальним) броунівським рухом з показником Херста H , якщо дисперсія цього процесу впливає зі співвідношення $\sigma^2(X_t) = t^{2H}$, тут t – час, $0 < H < 1$.

Виконаємо оцінки фрактальної розмірності для згенерованої вибірки процесу з незалежними прирощеннями зі спектром потужності, наведеними на рис. 6.11.

Представлений спектр апроксимується степеневою залежністю з деяким дробовим показником. Вимірювання, проведені за допомогою масштабування і вимірювання «довжини» послідовності дають значення $D_L = 1.48$. Вимірювання

за допомогою оцінки дисперсії дають $D_{\sigma} = 1,505$, а за допомогою вимірювання кореляційного інтеграла $D_C = 1.66$.

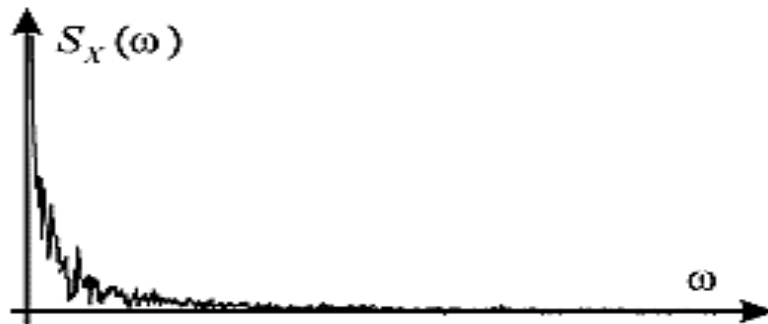


Рис. 6.11. Спектр потужності вінерівського процесу з незалежними прирощеннями

Відображення Ено має теоретичну розмірність $D=1.25$ і його спектр потужності представлений на рис. 6.12. Такий спектр погано представляється у вигляді степеневого закону, однак, оцінка $D_{\sigma} = 1.29$ напевно прийнятна з урахуванням того, що у чисельному експерименті було отримано $D_L = 1,22$. Значення дробової розмірності D_C для цього відображення виходить рівним $D_C = 1.3$.



Рис. 6.12. Спектр потужності відображення Ено

6.4.9. Визначення розмірності за максимумами.

Розглянемо чотири способи знаходження розмірності за максимумами.

А. *Елементарна оцінка D*. Розглянемо для прикладу предфрактал на основі кривої Коха (рис. 6.13).

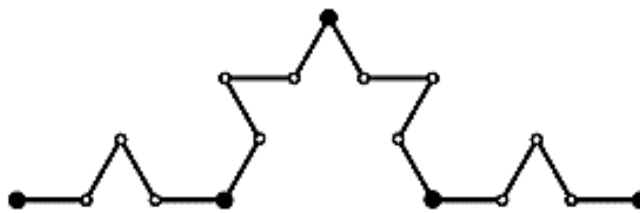


Рис. 6.13. Предфрактал з двох поколінь для кривої Коха

Для того щоб оцінити розмірність кривої Коха, потрібно знайти відношення довжин двох поколінь предфрактала. На рис. 6.13 перше (великомасштабне) покоління позначено жирними точками, а в другому поколінні (масштаб в три рази менший) дрібнішими, порожніми точками. Таким чином, жирними точками показані особливості (сингулярності), відповідні сусіднім масштабам. Тоді, розмірність кривої Коха буде $D = \log_3 4 = 1,26$. Даний розрахунок проведений з метою продемонструвати, як просто визначається розмірність у разі, коли відомі алгоритм побудови кривої і коефіцієнт скейлінга (рівний 3 в даному випадку).

Б. *Метод обліку сингулярностей при оцінці D*. Якщо який-небудь масштаб предфрактала містить особливість, то ця особливість повинна бути відтворена й на інших масштабах. У випадку з кривою Коха на роль особливостей напрошуються точки зламів лінії – на великому масштабі їх 5, а на дрібному їх 17 (див. рис. 6.13). Помітивши, що деяка «довжина - відрізок» знаходиться між двох особливостей або, по іншому, число відрізків менше числа особливостей на одиницю, отримуємо таку ж оцінку розмірності, що й

$$\text{при вимірюванні довжини } D = \log_3 \frac{(17 - 1)}{(5 - 1)} = 1,26.$$

Це підтверджує зроблений раніше висновок, що, дійсно, якщо деякий масштаб лінії матиме n особливостей, то більш детальний масштаб у фрактала з розмірністю D повинен мати $n_1 = n^D$ особливостей, звідки випливає, що

фрактальна розмірність в даному випадку більше топологічної, що, в свою чергу, збігається з іншим визначенням фрактала.

Цінність даного підходу полягає у тому, що вихідна крива не піддається ніяким перетворенням, і немає необхідності обговорювати спосіб вимірювання. Раніше оцінка проводилася по відношенню заходів на початковому зображенні і на згладженому.

В. *Метод функціоналів.* Розглянемо, спочатку, одновимірну функцію з фрактальними властивостями, тобто з деяким скейлінгом в кінцевому масштабі (рис. 6.14). Будь-яка безперервна функція приймає на деякому часовому відрізку T свої глобальні максимальне і мінімальне значення.

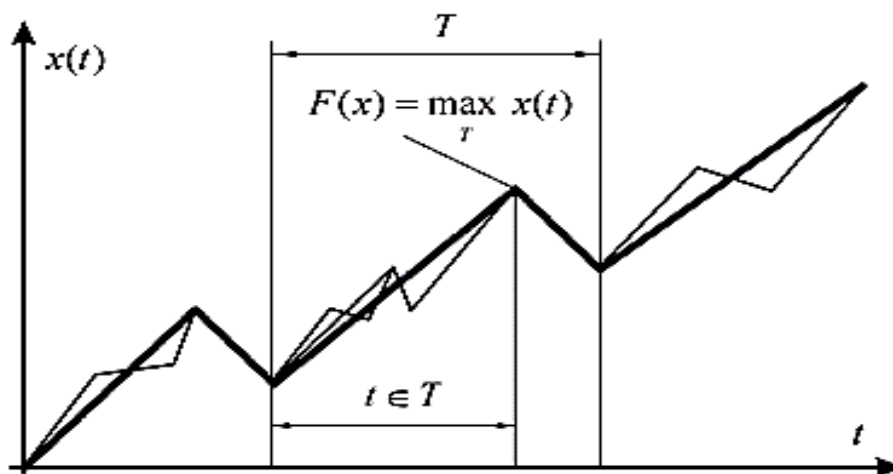


Рис. 6.14. Максимуми функції на різних масштабах

Якщо розділити відрізок T на частини, то в кожній з цих частин функція також прийме максимальне і мінімальне значення. Очевидно, що тільки деякі з цих значень будуть збігатися з глобальними максимумом і мінімумом функції на T і не один з максимумів (мінімумів) не буде перевищувати глобального максимуму (мінімуму) на T .

Самоподібна функція повинна задовольняти будь-якій умові подоби: нехай є функція $f(x)=Af(bx+c)$, де a, b, c – постійні, причому $A>1$. Умова $A>1$ дозволяє зробити висновок, що у подібній функції $\max_{t \in T} f(x) < \max_T f(x)$.

Іншими словами максимуми на більш детальних масштабах завжди менше максимумів на великих масштабах. Крім того, максимуми на великих масштабах будуть повторюватися при різних розподілах інтервалу T стільки раз, скільки було проведено розподілів, як це показано на рис. 6.14.

Для оцінки фрактальної розмірності за функціоналом, звернемося до представлення деякої функції з носієм у вигляді канторової множини (рис. 6.15).

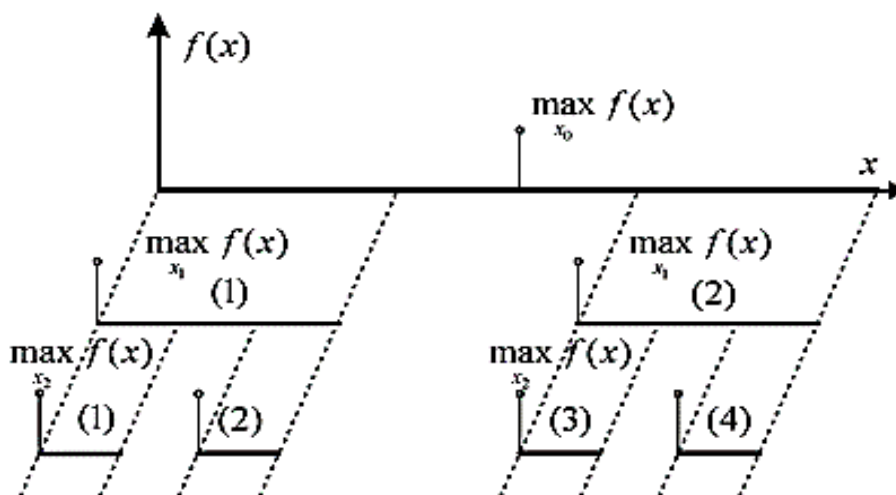


Рис. 6.15. Визначення розмірності функції з носієм у вигляді канторової множини

Співвіднесемо кількості максимумів на відрізках між двома сусідніми поколіннями предфрактала-носія. У нашому випадку маємо $N_2/N_1=2$, а фрактальна розмірність $D=\log_3 2 \approx 0.63$. Отриманий результат збігається з класичним значенням. Однак, в разі довільної множини, виникає питання – яким чином розділити максимуми за їх масштабами.

Г. *Метод триад* полягає в прибиранні сусідніх максимумів. Робиться це в такий спосіб. По-перше, вихідна послідовність розбивається на *триади* – множини, що містять три сусідні елементи послідовності. Далі, якщо триада містить максимум: $x_{i-1}^0 < x_i^0 > x_{i+1}^0$, то ця триада залишається для подальшого аналізу, а елементу x_i^0 присвоюється ранг – 1. Далі, з триад, отриманих на

першому кроці, будуються тріади з максимумами: $x_{i-1}^1 < x_i^1 > x_{i+1}^1$, і елементам x_i^1 присвоюється ранг 2 і так далі, поки у вибірці не залишиться всього три елементи.

В результаті, у нас є набори елементів x_i^j . За співвідношенням між сусідніми рангами максимумів можна оцінити фрактальну розмірність. Основа логарифма у цьому випадку дорівнює 3.

6.4.10.Визначення хаотичності режиму і наявності дивного атрактора.

Ознаки хаотичних коливань, які описані тут, носять в основному якісний характер і вимагають від дослідника застосування частки здорового глузду і досвіду. Є й кількісні ознаки хаосу, використання яких приносить певний успіх. Два найбільш поширених критерії – це показник Ляпунова і фрактальна розмірність, які використовуються у такий спосіб:

- 1) позитивний показник Ляпунова вказує на хаотичну динаміку;
- 2) фрактальна структура орбіти у фазовому просторі вказує на присутність дивного атрактора.

Перевірка із застосуванням показника Ляпунова може використовуватись як в дисипативних, так і в бездисипативних (консервативних) системах, а фрактальні розмірності мають сенс тільки в дисипативних системах.

За допомогою показників Ляпунова перевіряється чутливість системи до варіацій початкових умов. Ідея такої перевірки полягає у тому, щоб подумки виділити у фазовому просторі невелику кулю, в якій зосереджені початкові точки траєкторій, і простежити за її деформацією в еліпсоїд в ході динамічної еволюції системи. Якщо d – максимальний розмір еліпсоїда, а d_0 – початковий розмір сфери початкових умов, то сенс показника Ляпунова λ видно зі співвідношення

$$d = d_0 2^{\lambda(t-t_0)}. \quad (6.83)$$

Тут одиничного вимірювання недостатньо і результати розрахунку слід усереднити по різних дільницях фазового простору. Таке середнє може мати вигляд

$$\lambda = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(t_i - t_{0i})} \log_2 \frac{d_i}{d_{0i}}. \quad (6.84)$$

У системах з хаотичною динамікою області фазового простору витягуються, стискаються, складаються і відображаються назад на початкову область. При цьому відображенні в фазовому просторі залишаються лакуни. Це означає, що орбіти прагнуть заповнити менш ніж цілий підпростір фазового простору. Фрактальна розмірність – міра ступеня заповнення орбітою певного підпростору, і неціла розмірність – візитна картка дивного атрактора.

Практичне використання усіх введених вище фрактальних розмірностей для кількісної та якісної характеристик хаотичних коливань поки все ще залишається справою майбутнього і буде розширюватися в міру підвищення ступеня вивченості процесів нелінійної динаміки. У багатьох випадках виявляється достатнім встановити, що розмірність – неціле число і що аттрактор дійсно дивний. Однак фрактальна розмірність деяких атракторів близька до цілого числа (наприклад, для атрактора Лоренца $D_0 \approx 2,06$), тому фрактальна розмірність сама по собі ще не свідчить про хаотичну природу руху. У динамічних експериментах краще не покладатися на який-небудь один критерій хаосу, а використовувати для більшої надійності дві, три і більше ознаки, наприклад відображення Пуанкаре, спектр Фур'є, показники Ляпунова або фрактальну розмірність, перш ніж оголошувати систему хаотичною або дивною.

6.5. Мультифрактальний скейлінг-спектр

З дослідженням розподілу фізичних або будь-яких інших величин на геометричному носії пов'язані мультифрактальні заходи.

Ідея про те, що фрактальна міра може бути представлена взаємопов'язаними фрактальними підмножинами, що змінюються за степеневим законом з різними показниками, відкриває новий простір для застосувань фрактальної геометрії до фізичних систем.

6.5.1. Послідовність показників маси.

При аналізі структур різних множин \mathfrak{Z} найбільш широко використовується метод підрахунку клітин. При цьому методі E -вимірний простір спостережень розбивається на (гіпер)куби з ребром δ , після чого проводиться підрахунок числа $N(\delta)$ кубів, що містять принаймні одну точку множини \mathfrak{Z} . Ясно, що такий підрахунок дає грубу оцінку міри множини і число $N(\delta)$ не несе в собі ніякої інформації про структуру цієї множини. Наприклад, якщо берегова лінія сильно порізана і перетинає якусь клітину n_i раз, то при підрахунку ця клітина все одно дає внесок в загальне число клітин, що покривають множини \mathfrak{Z} , рівне 1, що не цілком коректне.

Чи існує спосіб, що дозволяє надавати клітинам з $n_i > 1$ більшу вагу, ніж клітинам з $n_i = 1$? Для вирішення цього питання було введено поняття – *набір показників маси для поверхні*.

Нехай множина \mathfrak{Z} , що складається з N точок, має в i -у осередку N_i точок. Ці точки можна розглядати як вибірку, яка відобразить розподіл міри на множині. Скористаємося «масою», або ймовірністю $\mu_i = N_i / N$, що відповідає i -й клітці, і побудуємо міру, яку можна записати у вигляді

$$M_d(q, \delta) = \sum_{i=1}^N \mu_i^q \delta^d = N(q, \delta) \delta^d \xrightarrow{\delta \rightarrow 0} \begin{cases} 0, & d > \tau(q), \\ \infty, & d < \tau(q). \end{cases} \quad (6.85)$$

Ця міра має *показник маси* $d = \tau(q)$, при якому вона не обертається в нуль і в нескінченність, коли $\delta \rightarrow 0$. Показник маси $\tau(q)$ для даної множини залежить від того, який порядок моменту q обраний.

Міра характеризується усією послідовністю показників $\tau(q)$, що визначають, за яким степеневим законом змінюються в залежності від δ ймовірності $\{\mu_i\}$. З формули (6.85) випливає, що зважене число клітин $N(q, \delta)$ представимо у вигляді

$$N(q, \delta) = \sum_i \mu_i^q \delta^{-\tau(q)}, \quad (6.86)$$

а показник маси визначається виразом

$$\tau(q) = - \lim_{\delta \rightarrow 0} \frac{\ln N(q, \delta)}{\ln \delta}. \quad (6.87)$$

Насамперед зазначимо, що, обираючи $q=0$ (q – порядок моменту), ми отримуємо $\mu_i^{q=0} = 1$. Отже, $N(q=0, \delta) = N(\delta)$ – це просто число клітин, що утворюють покриття множини, а $\tau(0) = D$ є фрактальною розмірністю множини. Ймовірності нормовані: $\sum_i \mu_i = 1$ і з формули (6.87) випливає, що $\tau(1) = 0$.

Вибір великих значень q , наприклад $q=10$ або 100 , в співвідношенні (6.86) сприяє підвищенню вкладу осередків з відносно великими значеннями μ_i , оскільки, якщо $q > 1$ і $\mu_i > \mu_j$, то $\mu_i^q > \mu_j^q$. Навпаки, вибір $q < 1$ сприяє підвищенню вкладу осередків з відносно малими значеннями міри μ_i на осередку. Ці границі найзручніше розглядати, вводячи похідну $d\tau(q) / dq$, яка визначається за допомогою границі

$$\frac{d\tau(q)}{dq} = \lim_{\delta \rightarrow 0} \frac{\sum_i \mu_i^q \ln \mu_i}{\left(\sum_i \mu_i^q \right) \ln \delta}. \quad (6.88)$$

Нехай μ_- – мінімальне значення μ_i в сумі. Тоді

$$\left. \frac{d\tau(q)}{dq} \right|_{q \rightarrow \infty} = - \lim_{\delta \rightarrow 0} \frac{\sum_i' \mu_i^q \ln \mu_i}{\left(\sum_i' \mu_i^q \right) \ln \delta}, \quad (6.89)$$

де штрих у знака суми вказує на те, що підсумовування проводиться тільки по осередках з $\mu_i = \mu_-$. Останній вираз запишемо у вигляді

$$\left. \frac{d\tau(q)}{dq} \right|_{q \rightarrow \infty} = - \lim_{\delta \rightarrow 0} \frac{\ln \mu_-}{\ln \delta} = -\alpha_{\max}. \quad (6.90)$$

Тут ми вводимо поняття *показника Ліпшиця-Гельдера α* . Аналогічні міркування в границі при $q \rightarrow +\infty$ приводять до висновку про те, що мінімальне значення α визначається виразом

$$\left. \frac{d\tau(q)}{dq} \right|_{q \rightarrow \infty} = - \lim_{\delta \rightarrow 0} \frac{\ln \mu_+}{\ln \delta} = -\alpha_{\min}, \quad (6.91)$$

де μ_+ – найбільше значення μ_i , що відповідає найменшому значенню α . Далі буде показано, що і в загальному випадку $\alpha = d\tau(q) / dq$.

При $q = 1$ отримаємо такий вираз для $d\tau(q) / dq$:

$$\left. \frac{d\tau(q)}{dq} \right|_{q=1} = - \lim_{\delta \rightarrow 0} \frac{\sum_i \mu_i \ln \mu_i}{\ln \delta} = \lim_{\delta \rightarrow 0} \frac{K(\delta)}{\ln \delta}. \quad (6.92)$$

Він має особливий інтерес: тут $K(\delta)$ – (інформаційна) ентропія розбиття міри по осередках розміру δ , тобто ентропія Колмогорова (K -ентропія). Цю ентропію розбиття можна записати у вигляді

$$s(\delta) = - \sum_i \mu_i \ln \mu_i = -\alpha_1 \ln \delta. \quad (6.93)$$

Показник $\alpha_1 = -(d\tau / dq)|_{q=1}$ є також фрактальною розмірністю множини, на якому зосереджена міра; він задає степеневий закон, за яким змінюється при зміні розміру δ осередку ентропія (розбиття) міри. Слід зазначити, що ентропія розбиття $K(\delta)$ при вирішенні δ може бути виражена через ентропію δ міри за формулою $K(\delta) = -K \ln \delta$.

Типову поведінку послідовності показників маси $\tau(q)$ можна проілюструвати прикладом міри на відрізку, породженої мультиплікативним біноміальним процесом. Для цього процесу, що визначається виразом

$$N(d, \delta) = \sum_{k=0}^n \binom{n}{k} p^{qk} (1-p)^{q(n-k)} = \left(p^q + (1-p)^q \right)^n \quad (6.94)$$

маємо показник маси, який згідно (6.87) визначається виразом

$$\tau(q) = \frac{\ln \left(p^q + (1-p)^q \right)}{\ln 2}. \quad (6.95)$$

Зовнішній вигляд $\tau(q)$ біноміального мультиплікативного процесу з $p=0.25$ представлений на рис. 6.16.

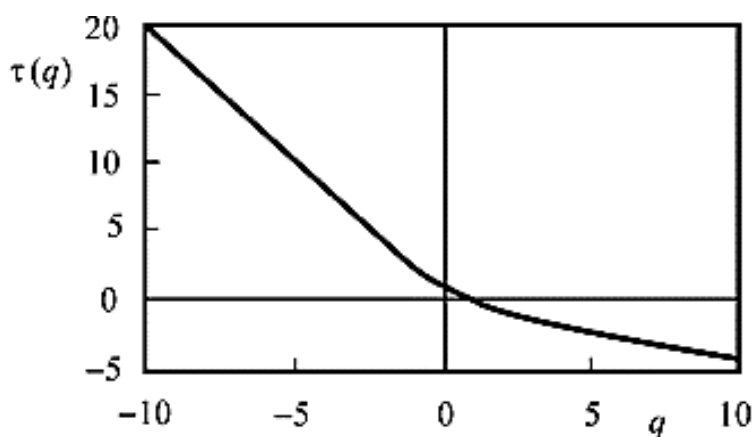


Рис. 6.16. Послідовність показників маси як функція порядку

6.5.2. Функція мультифрактального спектру.

Мультифрактал – об'єкт, який представляє собою неоднорідний фрактал. Для його опису вводиться набір узагальнених фрактальних розмірностей D_q , де q приймає будь-які значення в інтервалі $-\infty < q < +\infty$. Однак величини D_q не є, строго кажучи, фрактальними розмірностями у загальноприйнятому розумінні цього слова, через що в деякій літературі вони зветься узагальненими розмірностями.

Тому часто поряд з ними для характеристики мультифрактальної множини використовують так звану *функцію мультифрактального спектра* $f(\alpha)$ (спектр сингулярностей мультифрактала), до якої більше підходить термін фрактальна розмірність. Ми покажемо, що величина $f(\alpha)$ фактично дорівнює хаусдорфовій розмірності деякої однорідної фрактальної підмножини з початкової множини \mathfrak{Z} , яка дає домінуючий внесок в статистичну суму при заданій величині q .

Однією з основних характеристик мультифрактала є набір ймовірностей p_i , що показують відносну заселеність осередків ε , якими ми покриваємо цю множину. Чим менше розмір осередку, тим менше величина її заселеності. Для самоподібних множин залежність p_i від розміру осередку має степеневий характер $p_i(\varepsilon) \approx \varepsilon^{\alpha_i}$, де α_i є певним показником ступеня, що різний для різних осередків i . Відомо, що для регулярного (однорідного) фрактала всі показники ступеня α_i однакові і рівні фрактальній розмірності D , тоді $p_i = 1 / N(\varepsilon) \approx \varepsilon^{-D}$.

У цьому випадку статистична сума має вигляд

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon) = N(\varepsilon) \varepsilon^{Dq} \approx \varepsilon^{D(q-1)}. \quad (6.96)$$

Тому $\tau(q) = D(q-1)$ і все узагальнені фрактальні розмірності $D_q = D$ у цьому випадку збігаються й не залежать від q .

Однак для такого більш складного об'єкта, як мультифрактал, внаслідок його неоднорідності, ймовірності заповнення осередків p_i у загальному випадку неоднакові, і показник ступеня α_i для різних осередків може набувати різні значення. Як ми побачимо нижче, досить типовою є ситуація, коли ці значення безперервно заповнюють деякий закритий інтервал $(\alpha_{\min}, \alpha_{\max})$, причому $p_{\min} \approx \varepsilon^{\alpha_{\max}}$, а $p_{\max} \approx \varepsilon^{\alpha_{\min}}$.

Встановимо спершу зв'язок граничних значень α зі значеннями похідної від функції $\tau(q)$. А саме, розглянемо границі цієї похідної при $q \rightarrow \pm\infty$. Так,

якщо ми візьмемо значення $q \rightarrow +\infty$, то при виконанні підсумовування по i в виразі

$$\frac{d\tau(q)}{dq} = \lim_{\varepsilon \rightarrow 0} \frac{\sum_{i=1}^{N(\varepsilon)} p_i^q \ln p_i}{\left(\sum_{i=1}^{N(\varepsilon)} p_i^q \right) \ln \varepsilon} \quad (6.97)$$

буде істотний внесок тільки найбільш заселених осередків, кожен з яких характеризується максимальною вірогідністю заповнення p_{\max} . Залишивши в сумі тільки такі осередки (чисельністю N_{\max}), ми бачимо, що чисельник виразу (6.97) дорівнює $N_{\max} p_{\max}^q \ln p_{\max}$, а знаменник $N_{\max} p_{\max}^q \ln \varepsilon$. В результаті, з огляду на те, що $p_{\max} \approx \varepsilon^{\alpha_{\min}}$, шукана границя похідної виявляється рівною α_{\min} . Аналогічним чином, якщо $q \rightarrow -\infty$ то при підсумовуванні в виразі (6.97) необхідно враховувати лише найменш заселені осередки, які характеризуються ймовірністю p_{\min} . У цьому випадку очевидно, що похідна $d\tau/dq$ прагне до значення α_{\max} .

Величину цієї ж похідної при досить великих за модулем значеннях q можна обчислити і по-іншому. Для цього зауважимо, що у тому випадку, коли функція D_q має кінцеві границі при $q \rightarrow \pm\infty$ (рівні $D_{\pm\infty}$), функція $\tau(q)$ може бути апроксимована як $\tau(q \rightarrow \pm\infty) \approx q D_{\pm\infty}$. Таким чином, ми приходимо до важливого висновку, що

$$\left. \frac{d\tau}{dq} \right|_{q \rightarrow +\infty} = D_{+\infty} = \alpha_{\min}, \quad \left. \frac{d\tau}{dq} \right|_{q \rightarrow -\infty} = D_{-\infty} = \alpha_{\max}. \quad (6.98)$$

Тобто інтервал можливих значень α визначається граничними значеннями (при $q \rightarrow \pm\infty$) узагальнених фрактальних розмірностей D_q .

Перейдемо тепер до питання про розподіл ймовірностей різних значень α_j . Нехай $n(\alpha)d\alpha$ є ймовірністю того, що α_j знаходиться в інтервалі від α до $\alpha+d\alpha$. Іншими словами, $n(\alpha)d\alpha$ є відносним числом осередків i , що

володіють однією і тією ж мірою p_i з α_i , і які лежать в цьому інтервалі. У разі монофрактала, для якого всі α_i однакові (і рівні фрактальній розмірності D), це число, очевидно, пропорційно повній кількості осередків $N(\varepsilon) \approx \varepsilon^{-D}$, і є таким, що степеневим чином залежить від розміру осередку ε . Показник ступеня в цьому співвідношенні визначається фрактальною розмірністю множини D .

Для мультифрактала, однак, це не так, і різні значення α_i зустрічаються з ймовірністю, яка характеризується не однією і тією ж величиною D , а різними (в залежності від α) значеннями показника ступеня $f(\alpha)$, $n(\alpha) \approx \varepsilon^{-f(\alpha)}$.

Таким чином, фізичний зміст функції $f(\alpha)$ полягає у тому, що вона являє собою хаусдорфову розмірність деякої однорідної фрактальної підмножини \mathfrak{S}_α з початкової множини \mathfrak{S} , що характеризується однаковими можливостями заповнення осередків $p_i \approx \varepsilon^\alpha$. Оскільки фрактальна розмірність підмножини очевидно завжди менше або дорівнює фрактальній розмірності початкової множини D_0 , має місце важлива нерівність для функції $f(\alpha)$, а саме $f(\alpha) < D_0$.

В результаті ми прийшли до висновку, що набір різних значень функції $f(\alpha)$ (при різних α) представляють собою *спектр фрактальних розмірностей* однорідних підмножин \mathfrak{S}_α , на які можна розбити початкову множину \mathfrak{S} . Звідси стає зрозумілим термін *мультифрактал*. Його можна розуміти як деяке об'єднання різних однорідних фрактальних підмножин \mathfrak{S}_α початкової множини \mathfrak{S} , кожна з яких має своє власне значення фрактальної розмірності $f(\alpha)$.

6.5.3. Геометричні властивості мультифрактального скейлінг-спектра.

Проаналізуємо тепер поведінку функції $f(\alpha)$ для різних значень α . Оскільки $f'(\alpha) = q$, то при $q=0$ похідна функції $f(\alpha)$ обертається в нуль. Це означає, що в деякій точці $\alpha_0 = \alpha(0)$ функція $f(\alpha)$ має максимум. Значення функції в максимумі легко визначити, якщо скористатися виразом

$$D_q = \frac{1}{q-1} [q\alpha(q) - f(\alpha(q))]. \quad (6.99)$$

Поклавши в (6.99) $q=0$, ми отримаємо, що $f(\alpha_0) = D_0$, тобто максимальне значення $f(\alpha)$ дорівнює хаусдорфовій розмірності мультифрактала D_0 , тобто фрактальній розмірності носія міри. Якісно ця ситуація відображена на рис. 6.17. Там же показані межі інтервалу $(\alpha_{\min}, \alpha_{\max})$, в якому задана функція $f(\alpha)$.

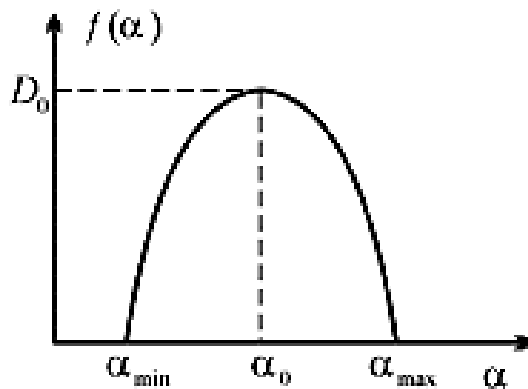


Рис. 6.17. Максимум функції $f(\alpha)$ дорівнює фрактальній розмірності D_0

Зауважимо, що обернення функції $f(\alpha)$ в нуль на цих межах (як показано на рис. 6.17) зовсім не обов'язково, і в ряді випадків $f(\alpha)$ в одній з цих точок (або в обох) може бути й відмінною від нуля. Обов'язковою умовою, однак, є обернення в нескінченність похідної $f'(\alpha)$ в цих двох точках. Це є прямим наслідком того факту, що точки α_{\min} і α_{\max} відповідають значенням $q \rightarrow \pm\infty$.

Функція $f(\alpha)$ поблизу свого максимуму може бути апроксимована параболою. Кривизна параболи визначається значенням другої похідної від цієї

функції в точці α_0 . Користуючись співвідношенням $\frac{d^2 f(\alpha)}{d\alpha^2} = \left(\frac{d^2 \tau(q)}{dq^2} \right)^{-1}$,

одержуємо, що $f''(\alpha_0) = 1 / \tau''(0)$. Диференціюючи тепер

формулу $\tau(q) = (q-1)D_q$ двічі і беручи до уваги, що $\tau'(0) = \alpha_0$, ми приходимо до виразу $\tau''(0) = 2(D_0 - \alpha_0) - D''_{q=0}$. Звідси випливає шукана апроксимація:

$$f(\alpha) \approx D_0 - \frac{(\alpha - \alpha_0)^2}{2[2(\alpha_0 - D_0) + D''_{q=0}]}. \quad (6.100)$$

Важливою властивістю функції $f(\alpha)$ є те, що вона всюди опукла. З умови опуклості функції $f(\alpha)$ очевидно, що величина, яка стоїть в квадратних дужках у цьому виразі, повинна бути завжди позитивна. Часто останній доданок $D''_{q=0}$ в цих дужках чисельно малий і їм можна знехтувати.

Розглянемо тепер випадок, коли $q=1$. Оскільки $\tau(1) = 0$, то $\alpha(1) = f(\alpha(1))$. З іншого боку, оскільки $q = df(\alpha) / d\alpha$, то похідна від функції $f(\alpha)$ в цій точці дорівнює 1: $f'(\alpha) = 1$. Диференціюючи співвідношення $\tau(q) = (q-1)D_q$ по q з урахуванням $\alpha = d\tau(q) / dq$, маємо

$$\frac{d\tau(q)}{dq} = D_q + (q-1)D'_q = \alpha(q), \quad (6.101)$$

і вважаючи $q=1$, ми отримуємо, що $\alpha(1) = D_1$. Таким чином, ми маємо

$$D_1 = \alpha(1) = f(\alpha(1)), \quad (6.102)$$

тобто інформаційна розмірність D_1 лежить на кривій $f(\alpha)$ в точці, де $\alpha = f(\alpha)$ і $f'(\alpha) = 1$. Це дає нам графічний спосіб визначення інформаційної розмірності по кривій $f(\alpha)$ (рис. 6.18).

Те, що графіки функцій α і $f(\alpha)$ торкаються один одного саме в точці (D_1, D_1) , зовсім не випадково. Нагадаємо, що $f(\alpha(1))$ – це значення фрактальної розмірності тієї підмножини з \mathfrak{S} , яка дає найбільший внесок у статистичну суму при $q=1$. Але при $q=1$ статистична сума в силу умови нормування дорівнює 1 і не залежить від розміру осередку ε . Отже, цей найбільший внесок повинен бути також біля одиниці. Тому в цьому (і тільки в

цьому) випадку ймовірності заповнення осередків $p_i \approx \varepsilon^\alpha$ обернено пропорційні числу наявних осередків $n(\alpha) \approx \varepsilon^{-f(\alpha)}$, тобто $f(\alpha) = \alpha$.

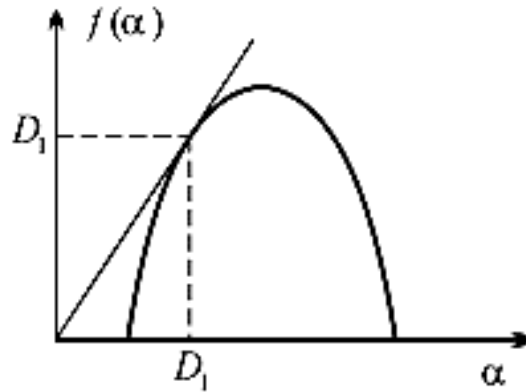


Рис. 6.18. Знаходження інформаційної розмірності $D_1 = \alpha = f(\alpha)$

Розглянемо тепер випадок $q=2$. Користуючись формулою (6.99), отримуємо $D_2 = 2\alpha(2) = f(\alpha(2))$ або $f(\alpha(2)) = 2\alpha(2) = D_2$, що відповідає геометричній побудові на рис. 6.19.

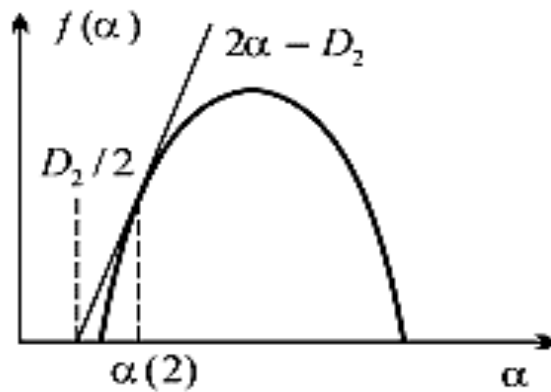


Рис. 6.19. Геометричне визначення кореляційної розмірності D_2

Як зазначалося вище, нерівність $f(\alpha) \leq \alpha$, виведена нами вище з якісних міркувань, еквівалентна твердженню, що похідна $D'_q \leq 0$. Доведемо тут це.

Для цього продиференціюємо вираз $D_q = \tau(q) / (q-1) / (q-1)$ по q .

Отримаємо $\frac{dD_q}{dq} = -\frac{\alpha - f(\alpha)}{(q-1)^2}$, звідси якраз і випливає вищенаведена нерівність

$$D'_q \leq 0.$$

6.6. Фрактальний аналіз часових рядів

Сучасні методи дослідження часових рядів припускають розгляд часових послідовностей як сукупності періодичних і хаотичних процесів. Хаотична компонента міститься як в динаміці форми сигналу, так і в зміні його частоти. Тому становить інтерес, з одного боку, зіставлення періодичних і хаотичних компонент сигналу і, з іншого боку, вивчення динаміки хаотичних компонент.

Одним з підходів вивчення хаотичної компоненти, які використовують алгоритм реконструкції *дивних атракторів*, є метод вивчення детермінованих неперіодичних процесів, для яких неможливий довгостроковий прогноз. Він дозволяє з'ясувати, наскільки складною має бути модель досліджуваного процесу або явища (скільки в ній повинно бути ступенів свободи або параметрів порядку), наскільки великий часовий інтервал, на якому можна прогнозувати поведінку досліджуваного об'єкта, встановити існування хаотичних режимів.

Дивні атрактори можуть бути описані *фрактальною розмірністю*. Методи визначення фрактальної розмірності використовуються в разі, коли передбачається наявність у системи, що вивчається, атратора.

В теорії ДС фрактальні множини займають важливе місце, оскільки рішенням більшості нелінійних задач є фрактал. Основною характеристикою фракталів є їх розмірність, яка вказує на близькість таких множин до регулярних об'єктів і дозволяє визначити число незалежних змінних, що однозначно їх описують.

Останнім часом виявлено велику кількість прикладних задач, де

фрактальна структура і розмірність служать основними характеристиками.

Одним з найбільш розроблених методів фрактального аналізу є метод розрахунку *кореляційної розмірності* (кореляційного інтеграла). Для її обчислення безперервна траєкторія замінюється множиною точок N_B в фазовому просторі, потім обчислюється відстань S_x між парами точок

$$S_x = |x_i - x_j|, \quad (6.103)$$

де x_i і x_j – координати точок.

На наступному кроці розраховується кореляційна функція як $C(r) = \lim_{N \rightarrow \infty} \frac{1}{N^2}$ (число пар (i, j) , для яких відстань $S_y < r$).

Оскільки для багатьох атракторів ця функція залежить від r при $r \rightarrow 0$ за степеневим законом, тобто $\lim C(r) = ar^d$, то кореляційну розмірність можна визначити за такою формулою:

$$d_G = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log r}. \quad (6.104)$$

Кореляційну функцію $C(r)$ також можна обчислити, описавши в фазовому просторі сферу навколо кожної точки x_i і підрахувавши число точок в кожній сфері:

$$C(r) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \Pi(r - |x_i - x_j|), \quad (6.105)$$

Де $\Pi(k) = 1$ при $k > 0$ і $\Pi(k) = 0$ при $k \leq 0$.

Наразі для порівняння фрактальних властивостей різних процесів найбільш часто використовується метод Херста, в якому для аналізу часових рядів використовується безрозмірний показник у вигляді відношення розмаху R до середньоквадратичного відхилення S (R/S).

Залежність параметра R/S від часу спостереження, побудована в подвійному логарифмічному масштабі, являє досліджуваний процес у вигляді фрактальної функції. Апроксимацією фрактальної функції прямою лінією

визначається кутовий коефіцієнт H , званий показником Херста, який використовується для обчислення основного фрактального параметра процесу – розмірності Хаусдорфа і є інтегральною характеристикою об'єкта або процесу:

$$D = 2 - H. \quad (6.106)$$

Залежно від значення фрактального показника H в методі Херста (H змінюється від 0 до 1) аналізовані процеси можуть бути класифіковані як:

- антиперсистентні, для яких характерна знакозмінна тенденція в поєднанні з відносно високим рівнем зашумленості, фрактальні лінії яких розташовані в області, де $0 < H < 0,5$;
- персистентні, для яких характерне збереження стійкої тенденції, щоспостерігається, у поєднанні з відносно низьким рівнем зашумленості, фрактальні лінії яких розташовані в області, де $0,5 < H < 1$;
- процеси, в яких тренд відсутній, а ступінь зашумленості визначається факторами, які не можна врахувати в методі Херста, якщо $H=0,5$ (зокрема, фрактальні лінії всіх стаціонарних сигналів вироджуються в пряму з $H=0,5$ і, таким чином, ніяк не розділяються).

Загальні закономірності зв'язку ступеня зашумленості сигналів і їх фрактальних властивостей, виражених показником Херста H , проілюстровані на рис. 6.20, де зображені реалізації часових рядів спостережень, які мають суттєво різні фрактальні властивості і, відповідно, різні оцінки показника Херста.

Візуально можна визначити, що стаціонарні випадкові сигнали (наприклад шум з нормальним розподілом) мають максимальну зашумленість, а зашумленість фрактальних сигналів падає зі збільшенням показника Херста H . Наведений метод дозволяє класифікувати на одній фрактальній площині стаціонарні, нестаціонарні і квазіперіодичні сигнали.

Прикладом комбінованої методики, що базується на застосуванні теорії фракталів та вейвлет-аналізу для виявлення особливостей часових рядів, є

застосування показника Херста як для частотних, так і для енергетичних спектрів аналізованих динамічних процесів.

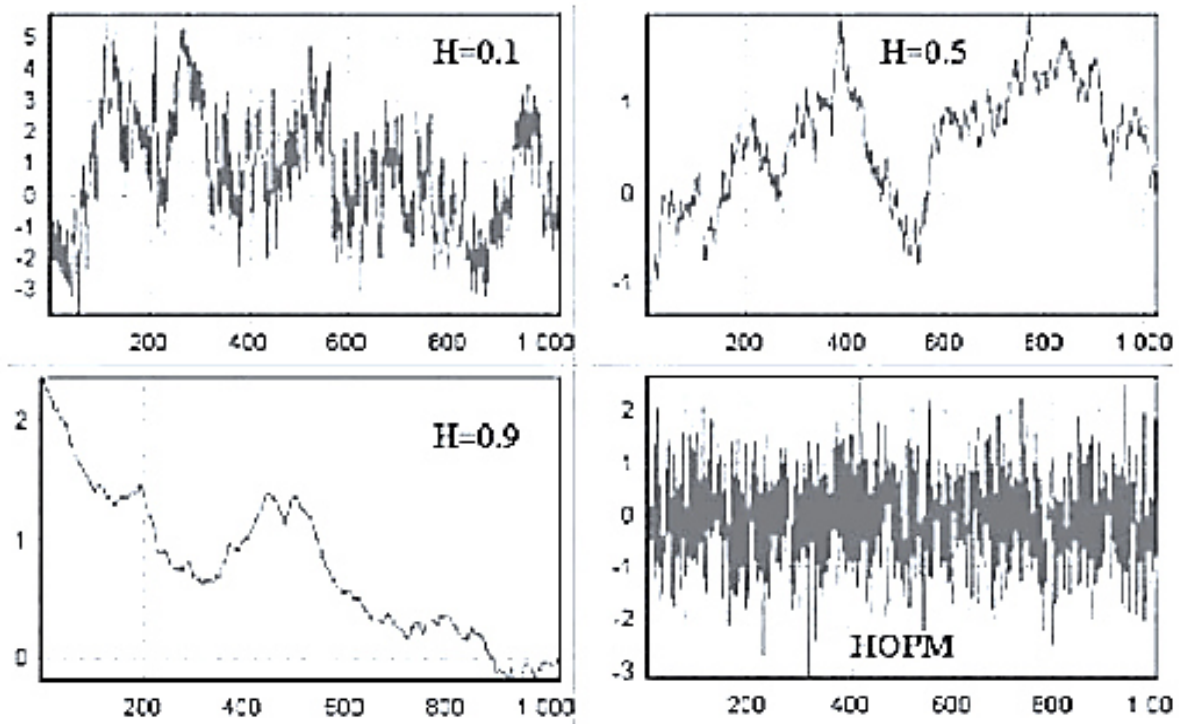


Рис. 6.20. Типові реалізації фрактальних часових рядів спостережень та білого шуму з нормальним розподілом

При цьому аналізуються множини, що характеризують частотний спектр і енергії частотних складових (вейвлет-спектр). Частотний спектр і максимальні значення енергій визначаються за допомогою вейвлет-аналізу по результуючій поверхні вейвлет-перетворення $C(a_i, b_j)$.

Розподіл частотного складу в момент часу t , що визначається максимальними коефіцієнтами вейвлет-перетворення ($C(a_i, b_j)_{\max}$) (рис. 6.21), можна охарактеризувати фрактальною розмірністю частотного спектра d_f (або показником Херста). Умовою стійкості є мінімальність зміни фрактальної розмірності частотного спектра з часом, для цього можна ввести обмежуюче значення ε_f :

$$\Delta d_f(a_i, t) \leq \varepsilon_f. \quad (6.107)$$

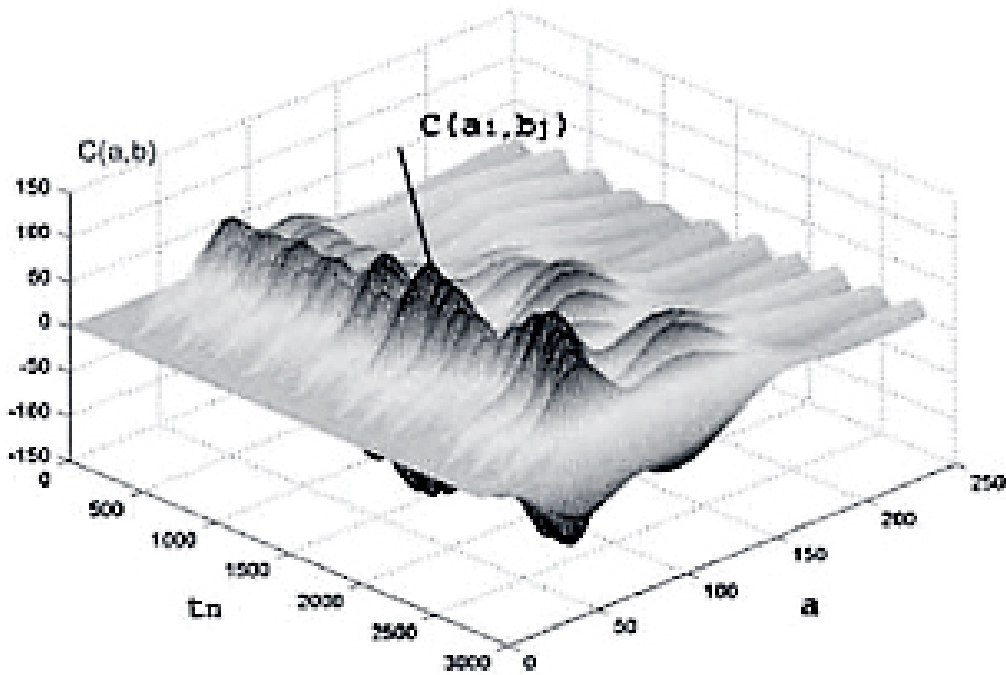


Рис. 6.21. Тривимірне зображення вейвлет-коефіцієнтів $C(a_i, b_j)$;
 t_n – час; a_i, b_j – масштабні параметри вейвлет-перетворення

Зміна енергії спектральних складових сигналу також визначається по фрактальній розмірності множини енергетичних коефіцієнтів, відповідних частотам (коефіцієнти $C(a, b)$): $\Delta d_c(a_i, t) \leq \varepsilon_c$. Частотний спектр, його розподіл і число частот визначаються за першою похідною

$$\frac{d(C(a_i, b))}{da} = 0, \quad i = 1, 2, \dots, n_b, \quad (6.108)$$

Де n_b – число розглянутих локальних максимальних значень коефіцієнтів $C(a_i, b)$.

У знайдених точках $a = a_i$ обчислюються значення коефіцієнтів $C(a_i, b)$, які визначають рівень енергії сигналу на даній частоті. Облік змін фрактальних показників d_f і d_C необхідно проводити за методикою побудови відображень Пуанкаре.

Відображення будуються з часовими інтервалами, рівними Δt ,

відповідними значущій високочастотній складовій часового ряду. Таке сканування поверхні $C(a,b)$ дозволяє визначити особливості зміни частотного складу і енергії його складових. Незначні коливання спектра і його енергії характеризуються локалізацією точок в обмеженій області фрактальної площині.

На рис. 6.22 наведено приклад оцінки структурної стійкості динаміки (часових рядів) за зміною значень коефіцієнтів $C(a,b)$ вейвлет-перетворення без введення часового інтервалу (безперервний фрактальний аналіз коефіцієнтів $C(a,b)$). Характер зміни коефіцієнтів $C(a,b)$ визначається на площині.

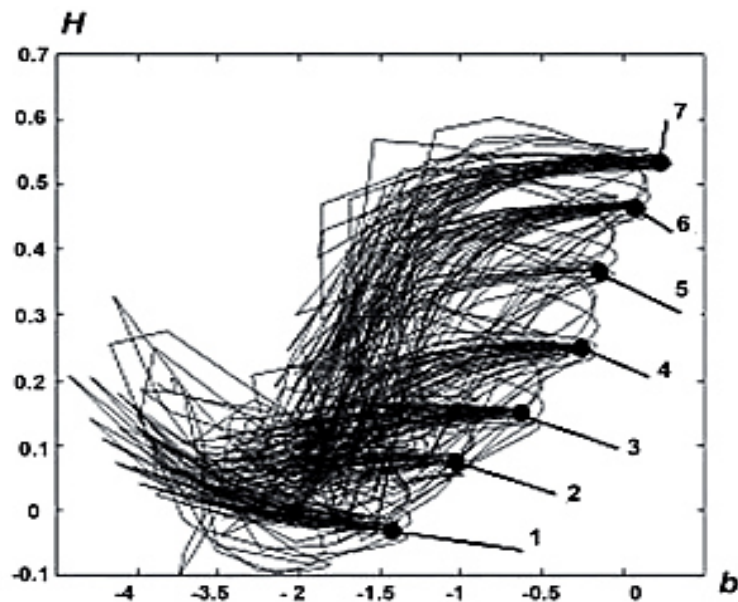


Рис. 6.22. Ілюстрація структурної стійкості часового ряду (зміна режимів функціонування системи 1...7)

Алгоритм методу можна представити у вигляді наступного процесу:

$$X(t) \Rightarrow C(a,b) \Rightarrow C(a_i,b)_{\max} \Rightarrow d(d_f(\Delta t)d_c(\Delta t)). \quad (6.109)$$

Таким чином, вейвлет-перетворення дозволяє отримати характеристики в часовій і частотній областях, а теорія фракталів є основою для опису внутрішньої структури послідовності цих даних – фрактальна оцінка дозволяє представити множини, що характеризують системи, у вигляді узагальнюючого

критерію d (або H), які можуть нести також і прогнозу функцію – оцінювати стійкість динамічного процесу.

Контрольні запитання

1. Основні поняття фракталів.
2. Класифікація фракталів.
3. Що таке гаусові процеси?
4. Що таке вінерівський процес?
5. Що таке білий шум Гауса?
6. Що таке пуассонівський процес?
7. Що таке дробовий білий шум?
8. Що таке кольоровий шум?
9. Самоподібні степеневі закони.
10. Що таке R/S-аналіз?
11. Що таке рожевий шум?
12. Що таке коричневий шум?
13. Що таке чорний шум?
14. Узагальнені фрактальні розмірності.
15. Основні властивості розмірності Хаусдорфа та інформаційна розмірність.
16. Кореляційна розмірність.
17. Розмірність подібності та її експериментальне визначення.
18. Визначення фрактальної розмірності фазової траєкторії і поверхні.
19. Визначення фрактальної розмірності для часових рядів.
20. Дисперсійна розмірність.
21. Визначення розмірності за максимумами.
22. Визначення хаотичності режиму і наявності дивного атрактора.
23. Послідовність показників маси.
24. Функція мультифрактального спектру.
25. Геометричні властивості мультифрактального скейлінг-спектру.

Глава 7. ЧАСТОТНО-ЧАСОВИЙ АНАЛІЗ

Класичне перетворення Фур'є (безперервне і дискретне) є досить корисним математичним апаратом для аналізу і синтезу сигналів, проте іноді виявляється недостатньо ефективним при обробці сигналів, що породжуються складними ДС. Перетворення Фур'є, наприклад, не відрізняє сигнали з двох синусоїд з різними частотами, один з яких представляє собою суму синусоїд, другий – синусоїди, які послідовно слідують одна за одною. В обох випадках їх спектр буде виглядати як два піки на двох фіксованих частотах. Отже, перетворення Фур'є у своєму традиційному вигляді не пристосоване для аналізу нестационарних сигналів, в тому числі, локалізованих на деякому часовому інтервалі, оскільки втрачається інформація про часові характеристики сигналу.

Спектральний аналіз реальних сигналів необхідно здійснювати як за частотою, так і за часом, оскільки на практиці найчастіше доводиться мати справу з нестационарними процесами, в яких інформативним є сам факт зміни частотно-часових характеристик сигналу. Прикладами таких сигналів є супутникові зображення Землі, рентгенограми внутрішніх органів організмів, мова і музика, турбулентні поля різної природи тощо, тобто фактично – весь обсяг інформації, з яким доводиться мати справу у повсякденному житті. Для виконання такого аналізу потрібні базисні функції, що володіють здатністю виявляти в сигналі, який аналізується, як частотні, так і його часові характеристики. Іншими словами, самі базисні функції повинні мати певні властивості, названі *частотно-часовою локалізацією*.

7.1. Частотно-часові властивості базисних функцій

7.1.1. Площина частота-час.

Для аналізу і порівняння частотно-часових локалізаційних властивостей різних базисів використовують площину частота-час. Будь-яка функція $\varphi(t)$

може характеризуватись інтервалом I_t на часовій вісі і інтервалом I_ω в частотній (Фур'є) області, в яких міститься 90% її енергії, зосередженої біля центру ваги функції $|\varphi(t)|^2$ і $|\Phi(\omega)|^2$. Тоді в цій площині функцію $\varphi(t)$ можна зобразити у вигляді прямокутника (рис. 7.1).

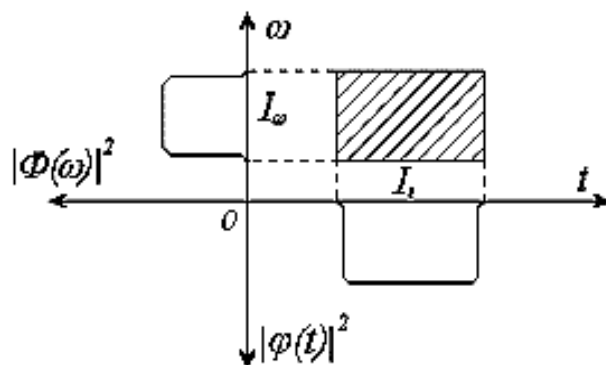


Рис. 7.1. Характеристика частотно-часової локалізації функції $\varphi(t)$

Очевидно, що зміщення функції на τ від початкового стану викличе переміщення прямокутника паралельно вісі t . Модуляція комплексною експонентою $e^{j\omega_0 t}$ цієї функції зрушує прямокутник паралельно вісі ω (рис. 7.2.).

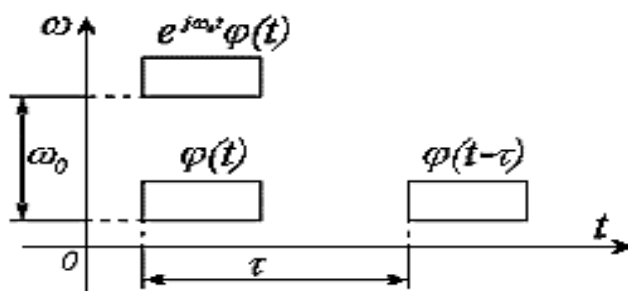


Рис. 7.2. Вплив зсуву на τ і модуляції $e^{j\omega_0 t}$ функції $\varphi(t)$ на її положення на площині частота-час

Масштабування функції (її стиснення або розтягнення) призводить до розвороту прямокутника. Дійсно, отримаємо нову функцію $\varphi_1(t)$ масштабуванням функції $\varphi(t)$ на коефіцієнт a :

$$\varphi_1(t) = \varphi(at). \quad (7.1)$$

Енергія такої функції

$$E_{\varphi_1} = \int_{-\infty}^{\infty} |\varphi_1(t)|^2 dt = \int_{-\infty}^{\infty} |\varphi(at)|^2 dt = \frac{1}{a} E_{\varphi}. \quad (7.2)$$

Отже, ширина функції $\varphi_1(t)$ дорівнює $I'_t = \left(\frac{1}{a}\right) I_t$. Відповідно до властивості масштабування Фур'є-перетворення маємо $I_{\omega} = a I'_{\omega}$. Вплив масштабування показано на рис. 7.3.

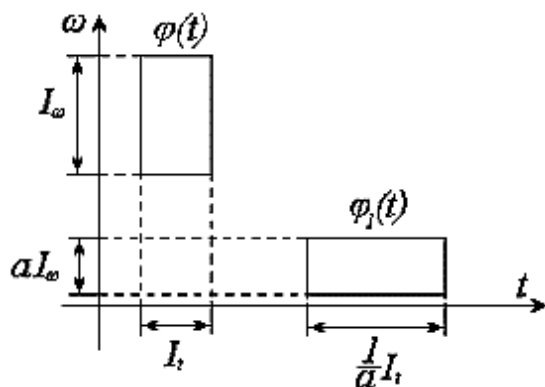


Рис. 7.3. Положення функції $\varphi(t)$ на площині частота-час при масштабуванні на коефіцієнт $a < 1$

За приклади функцій, що ілюструють ефективність їх представлення в площині частота-час, розглянемо δ -функцію Дірака і Фур'є-базис. Відомо, що δ -функція є ідеальним базисом для часового аналізу сигналів. Результатом такого аналізу є відліки, які можна розглядати як часовий спектр сигналу. На площині частота-час δ -функція $\delta(t - k\tau_0)$ виглядає як показано на рис. 7.4,а, тобто ця функція має властивість гарної часової локалізації, але поганої локалізації в спектральній області (вона має рівномірний спектр на всіх частотах). Базисні

функції $e^{j\omega t}$ Фур'є-аналізу, навпаки, мають гарну частотну локалізацію, але в часовій області вони мають нескінченну протяжність (рис. 7.4,б).

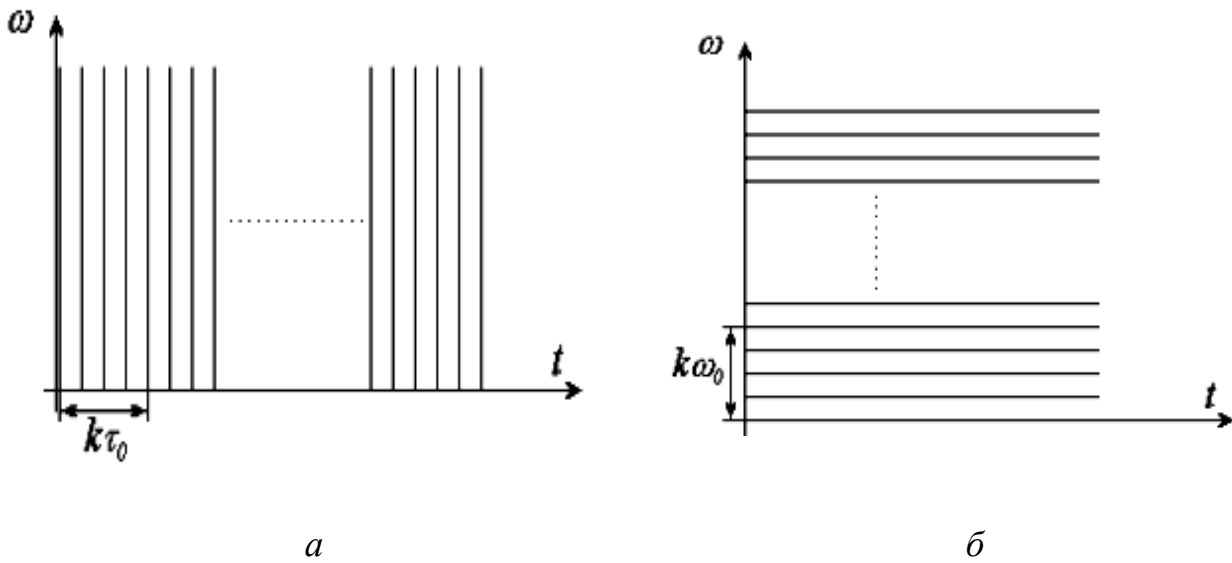


Рис. 7.4. δ -функція Дірака (а) і базисні функції Фур'є (б) на площині частота-час

7.1.2. Обмежене в часі Фур'є-перетворення.

Локальність перетворення Фур'є досягається шляхом обмеження аналізованого сигналу з допомогою рухомого вікна. Результатом такого аналізу є функція двох змінних – положення вікна τ і частоти ω :

$$F(\omega, \tau) = \int_{-\infty}^{\infty} w(t - \tau) f(t) e^{-j\omega t} dt. \quad (7.3)$$

Інакше кажучи, спектральний аналіз у вікні даних проводиться обчисленням скалярного добутку сигналу і базисної функції

$$g_{\tau}(\omega, \tau) = w(t - \tau) e^{j\omega t}, \quad (7.4)$$

або у вигляді

$$F(\omega, \tau) = \langle g_{\tau}(\omega, t), f(t) \rangle. \quad (7.5)$$

Таким чином в спектральний аналіз, крім частоти, вводиться ще один параметр – час. Для досягнення гарної часової локалізації перетворення (7.3) в

якості вікна необхідно обирати функцію Гауса. Таке Фур'є-перетворення носить назву *перетворення Габора*. Обмежене у часі перетворення Фур'є на площині частота-час показано на рис. 7.5. Тут зображено:

а) приклад базисних функцій перетворення $g_{\tau}(\omega, t) = e^{jk\omega_0 t} w(t - \tau_0)$ при зсуві τ_0 і $k = 1, 2, 3$ (дійсна частина);

б) зображення функцій в площині частота-час (заштриховані прямокутники).

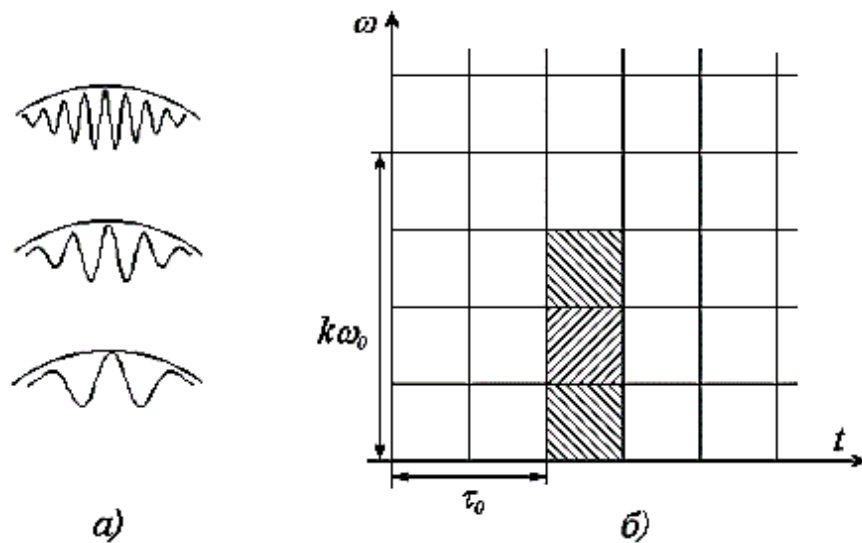


Рис. 7.5. Обмежене у часі перетворення Фур'є на площині частота-час

Як видно з рис. 7.5, при зсуві вікна або зміні частоти модуляції ширина прямокутника зберігається незмінною. Це викликано тією обставиною, що при усіх цих операціях ширина самого вікна не змінюється.

7.1.3. Принцип невизначеності.

Як впливає з властивості масштабування Фур'є-перетворення, ширина базисної функції в часовій і частотній областях взаємопов'язані: розширення функції в часовій області веде до її звуження в частотній, і навпаки. Закономірність, що пов'язує дві ці величини, називається *принципом*

невизначеності. В якості міри концентрації енергії будь-якої функції за часом і за частотою приймають добуток других центральних моментів функцій $\varphi(t)$ і її спектру $\Phi(\omega)$. Припустимо, що $\varphi(t)$ і $\Phi(\omega)$ вже центровані функції (якщо це не так, то шляхом зміщення і модуляції їх можна відцентрувати). Тоді ширини функцій можна визначити як

$$\mu_t^2 = \int_{-\infty}^{\infty} t^2 |\varphi(t)|^2 dt, \quad (7.6)$$

$$\mu_\omega^2 = \int_{-\infty}^{\infty} \omega^2 |\Phi(\omega)|^2 d\omega. \quad (7.7)$$

Принцип невизначеності стверджує, що, якщо $\varphi(t)$ убуває швидше, ніж $1/\sqrt{t}$ при $t \rightarrow \pm\infty$, то

$$\mu_t^2 \mu_\omega^2 \geq \frac{\pi}{2}, \quad (7.8)$$

де рівність справедлива тільки, якщо

$$\varphi(t) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha t^2}, \quad (7.9)$$

тобто є функцією Гауса.

7.2. Базисні функції частотно-часового аналізу

Частотно-часовий аналіз призначений для виявлення локальних частотно-часових збурень сигналу. Унаслідок короткочасності таких збурень, сам сигнал може розглядатися як заданий в просторі L_2 , тобто для одновимірних сигналів – на всій дійсній вісі $R(-\infty, \infty)$ з нормою $\|f(t)\|^2 < \infty$.

Отже, базисні функції (які отримали назву *вейвлетів*), також повинні належати L_2 і швидко спадати при $t \rightarrow \infty$. Тоді, щоб перекрити такими базисними функціями усі можливі часові положення сигналу, необхідно, щоб базисні функції являли собою набір зміщених у часі функцій. Зручніше за все, якщо цей

набір утворюється з однієї і тієї ж «материнської» функції $\psi(t)$ (прототипу), зсунутої по вісі t , тобто $\{\psi(t-b)\}$. Щоб забезпечити частотний аналіз, базисна функція повинна мати ще один аргумент – масштабний коефіцієнт, який є аналогом частоти в Фур'є-аналізі. Тоді базисні функції для частотно-часового аналізу матимуть вигляд

$$\psi\left(\frac{t-b}{a}\right) = \psi\left(\frac{t-b}{a}\right); \quad a, b, \in R, \quad (7.10)$$

де масштабний коефіцієнт a введений як дільник t , причому масштабуванню піддається також і зсув b . Це дозволяє зберегти відносну «щільність» розташування базисних функцій по вісі t при розширенні або стисненні самої функції і при $a/b = \Delta = \text{const}$ (рис. 7.6).

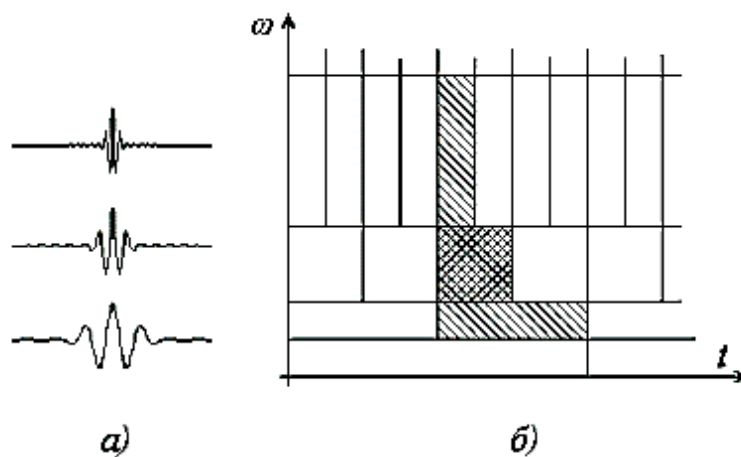


Рис. 7.6. Базисні функції вейвлет при масштабах $a = 2^k, k = 0, 1, 2$ (а)
і їх зображення в площині час-частота (б)

Таким чином, базисні функції для частотно-часового аналізу повинні мати наступні властивості.

Обмеженість, тобто належність L_2

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty. \quad (7.11)$$

Локалізація. Базисні функції вейвлет-аналізу (ВА), на відміну від перетворення Фур'є, повинні бути локалізовані, тобто визначені на кінцевому інтервалі як в часовій, так і в частотній областях. Для цього достатньо, щоб виконувались умови:

$$|\psi(t)| \leq C \cdot (1+|t|)^{-1-\varepsilon} ; |\psi(\omega)| \leq C \cdot (1+|\omega|)^{-1-\varepsilon} \quad (7.12)$$

при $\varepsilon > 0$.

Нульове середнє. Рівність нулю нульового моменту

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (7.13)$$

або, що іноді необхідно – рівність нулю моменту m -го порядку

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0. \quad (7.14)$$

Це вейвлети m -го порядку, що дозволяють аналізувати більш тонку структуру сигналу, пригнічуючи його складові, які повільно змінюються.

7.3. Безперервне вейвлет–перетворення

Введемо базис, який відповідає наведеним вище умовам

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad (7.15)$$

де множник $\frac{1}{\sqrt{|a|}}$ необхідний для збереження норми $\|\psi_{a,b}(t)\| = \|\psi(t)\|$.

Нехай $a, b \in R$, тобто приймають довільні речові значення, тоді пара перетворень, що носить назву безперервного вейвлет перетворення (ВП), та позначається як CWT (Continuous Wavelet Transform), буде мати вигляд:

$$CWT_f(a,b) = \left\langle f(t), \psi_{a,b}(t) \right\rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt, \quad (7.16)$$

$$f(t) = C_{\psi}^{-1} \int_{-\infty}^{\infty} \frac{da}{a^2} \int_{-\infty}^{\infty} CWT_f(a,b) \psi\left(\frac{t-b}{a}\right) db, \quad (7.17)$$

де нормалізуючий коефіцієнт

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty. \quad (7.18)$$

Інтегруванням (7.18) за частинами можна переконатися у тому, що ця умова завжди виконується, якщо $\Psi(\omega)=0$ при $\omega=0$ і, отже, дорівнює нулю принаймні, нульовий момент функції $\psi(t)$, тобто виконується умова (7.13).

Порівнюючи формули (7.16)-(7.17) з відповідними виразами для безперервного перетворення Фур'є

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega, \quad (7.19)$$

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt = \left\langle e^{-j\omega t}, f(t) \right\rangle_t, \quad (7.20)$$

бачимо, що роль функції $e^{j\omega t}$ тут грає функція $\psi_{a,b}(t)$, а C_{ψ} аналогічний коефіцієнту 2π , причому роль частоти грає масштабний множник $1/a$. Однак, як і в обмеженому в часі перетворенні Фур'є, базисна функція залежить ще від параметра зсуву b .

Наведемо приклади материнських вейвлетів, які формують базис (7.15). Найбільшою популярністю в даний час, мабуть, користуються функції на основі похідних функції Гауса:

$$\psi_m(t) = (-1)^m \frac{d^m}{dt^m} \exp\left\{-\frac{t^2}{2}\right\}. \quad (7.21)$$

Це викликано тією обставиною, що функція Гауса має найкращі показники локалізації як в часовій, так і в частотній областях.

При $m=1$ отримуємо вейвлет (рис. 7.7,а), який називають WAVE-вейвлет з рівним нулю нульовим моментом.

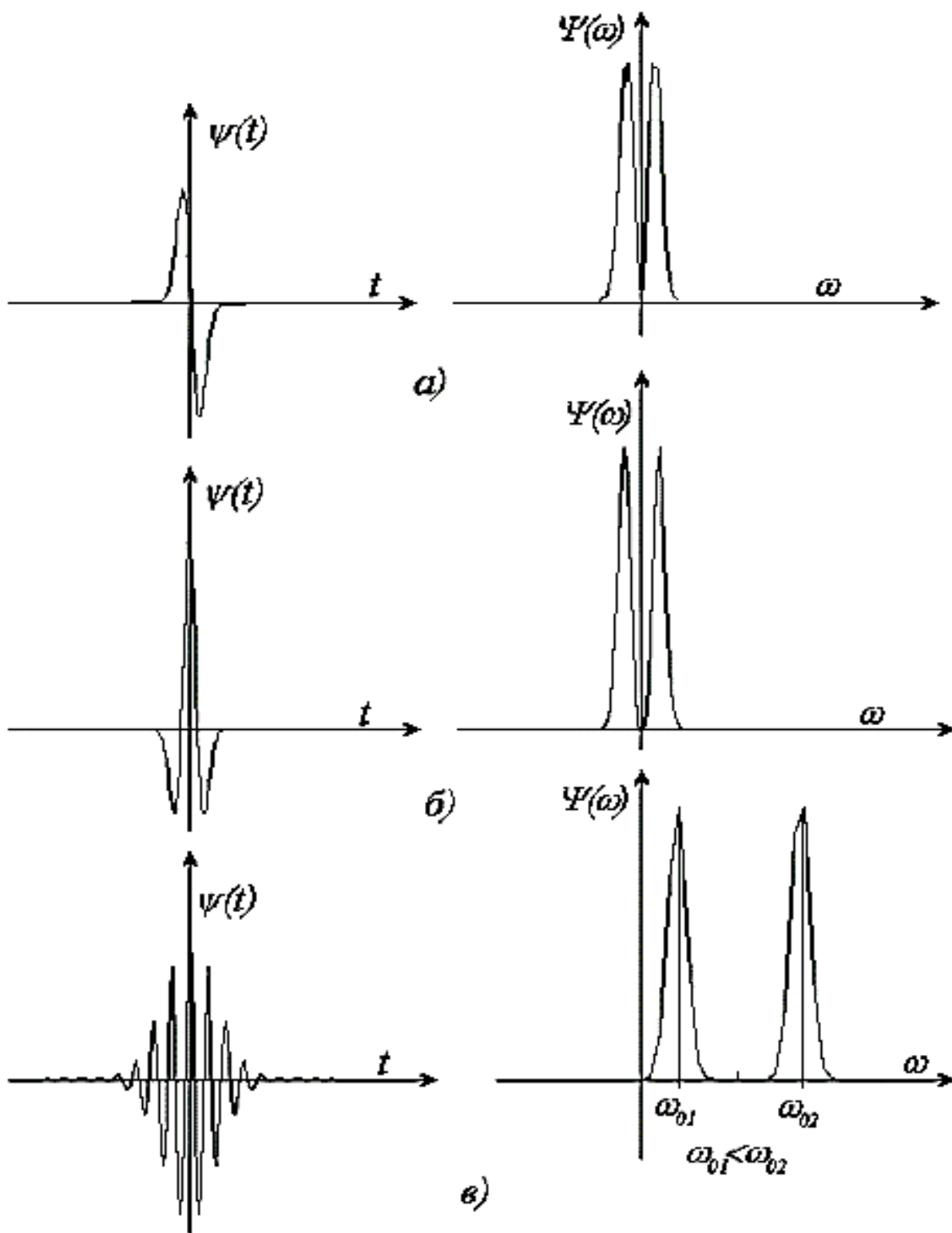


Рис. 7.7. Приклади базисних функцій WAVE-вейвлета (а), МНАТ-вейвлета (мексиканський капелюх) (б) і вейвлета Морле (в)

При $m=2$ отримуємо вейвлет (рис. 7.7,б), званий як «мексиканський капелюх» – МНАТ-вейвлет:

$$\psi(t) = \frac{2}{\sqrt{3}} \pi^{1/4} (1-t^2) e^{t^2/2}, \quad (7.22)$$

у якого нульовий і перший моменти дорівнюють нулю. Спектр Фур'є цього вейвлета більш вузький, тому він має кращу роздільність.

Функція Гауса утворює також DOG-вейвлет – різницю двох Гаусіан:

$$\psi(t) = \exp\left\{-\frac{t^2}{2}\right\} - 0,5 \exp\left\{-\frac{t^2}{8}\right\}. \quad (7.23)$$

Широке поширення отримав також вейвлет Морле (Morlet) (рис. 7.7,в),

$$\psi(t) = \exp\{j\omega_0 t\} \exp\left\{\frac{t^2}{2}\right\}. \quad (7.24)$$

де ω_0 – домінантна частота, що дозволяє варіювати вибірковістю базису. Цей вейвлет відрізняється від інших, перш за все тим, що він є комплексною функцією, у якій дійсні і уявні частини – модульовані гаусіаном гармоніки.

7.4. Властивості безперервного вейвлет-перетворення

Лінійність. Лінійність безперервного ВП впливає з лінійності скалярного добутку (7.16).

Нехай функції $f(t)$ і $g(t) \in L_2$. Тоді

$$CWT_{f,g}(a,b) = CWT_f(a,b) + CWT_g(a,b). \quad (7.25)$$

Зсув. Розглянемо безперервне перетворення вейвлет-функції $f_1(t)=f(t-b')$.

Тоді

$$\begin{aligned} CWT_{f_1}(a,b) &= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t-b') \psi\left(\frac{t-b}{a}\right) dt = \\ &= \frac{1}{\sqrt{|a|}} \int f(t_1) \psi\left(\frac{t_1 + b' - b}{a}\right) dt' = CWT_f(a, b - b'), \end{aligned} \quad (7.26)$$

тобто вейвлет-образ функції також зсувається на b' (рис. 7.8).

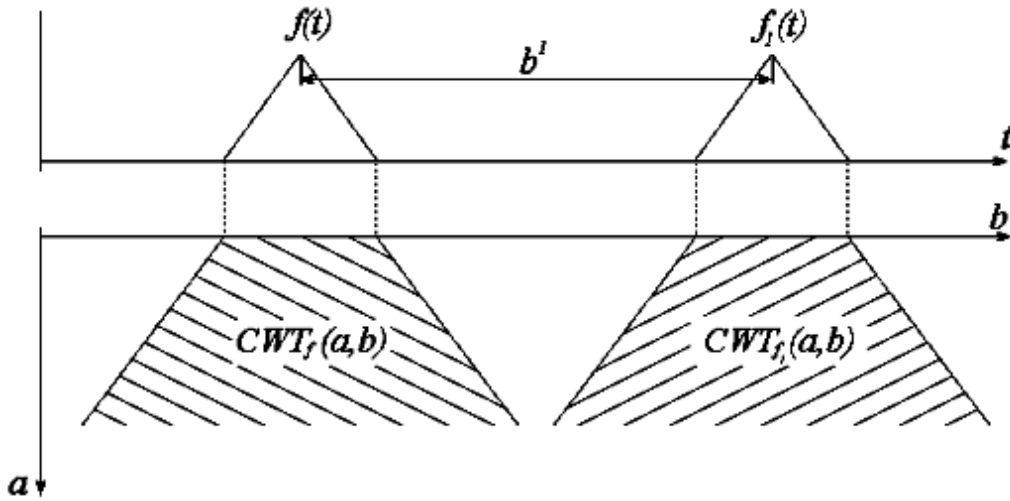


Рис.7.8. Зсув безперервного вейвлет-перетворення
(зсув функції в часовій області веде до зсуву її вейвлет-образу)

Масштабування. Розглянемо, далі, безперервне перетворення вейвлет-функції $f_1(t) = \frac{1}{\sqrt{c}} f(t/c)$, де множник $\frac{1}{\sqrt{c}}$ введений для збереження балансу енергії. Маємо:

$$\begin{aligned} CWT_{f_1}(a,b) &= \frac{1}{\sqrt{|a| \cdot c}} \int_{-\infty}^{\infty} f\left(\frac{t}{c}\right) \psi\left(\frac{t-b}{a}\right) dt = \\ &= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi\left(\frac{ct-b}{a}\right) f(t_1) dt_1 = CWT_f\left(\frac{a}{c}, \frac{t}{c}\right), \end{aligned} \quad (7.27)$$

тобто ВП також піддається масштабному перетворенню. Це означає, що, якщо функція розширюється в часовій області, то в масштабно-часовій (частотно-часовій) площині $[a, b]$ вона також розширюється (рис. 7.9).

Аналог теореми Парсеваля. Для кожної функції $f(t) \in L_2$ і її безперервного ВП справедливо наступне співвідношення

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |CWT_f(a,b)|^2 \frac{dadb}{a^2}. \quad (7.28)$$

Властивості локальності. Локальні властивості базису (7.15) в часовій і частотній областях характеризуються концентрацією енергії базисної функції

$\psi_{a,b}(t)$ в часовій і частотній областях. Тому кількісно цю властивість можна визначити величиною других центральних моментів функцій $|\psi_{a,b}(t)|^2$ і $|\Psi_{a,b}(\omega)|^2$, квадратний корінь з яких характеризує область найбільшої концентрації енергії цих функцій.

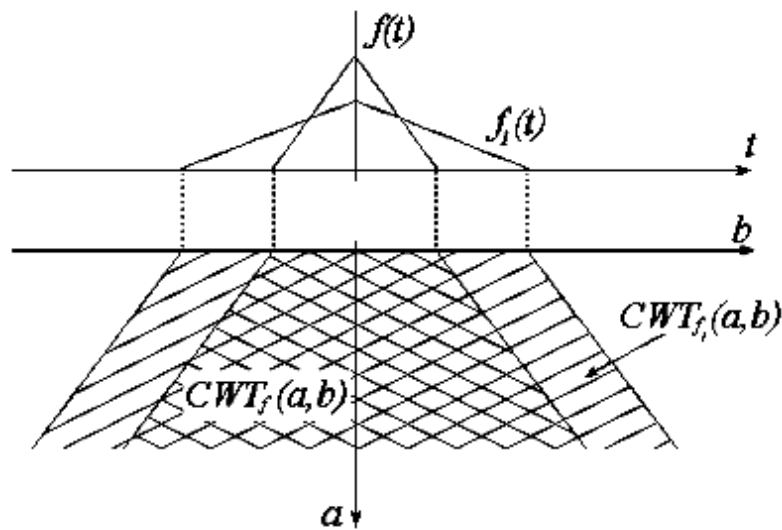


Рис. 7.9. Перетворення вейвлет-функції $f_1(t)$ ($CWT f_1(a, b)$) «перекриває» $CWT f(a, b)$) (при $c = 2$)

Позначимо через t_0 і Δ_t^2 , відповідно, величини першого і другого центральних моментів функції $|\psi(t)|^2$, тобто.

$$t_0 = \frac{1}{\|\psi\|^2} \int_{-\infty}^{\infty} t |\psi(t)|^2 dt, \quad (7.29)$$

$$\Delta_t^2 = \frac{1}{\|\psi\|^2} \left[\int_{-\infty}^{\infty} (t - t_0)^2 |\psi(t)|^2 dt \right]. \quad (7.30)$$

Тоді перший момент (центр ваги, середнє значення функції) $|\psi_{a,b}(t)|^2$ буде дорівнювати

$$\bar{t}_1 = \frac{1}{a\|\psi\|^2} \int_{-\infty}^{\infty} t \left| \psi \left(\frac{t-b}{a} \right) \right|^2 dt. \quad (7.31)$$

Після заміни змінної на $t_1 = \frac{t-b}{a}$ і перетворень, отримаємо:

$$\bar{t}_1 = \frac{a}{\|\psi\|^2} \int_{-\infty}^{\infty} t_1 \left| \psi(t_1) \right|^2 dt_1 + b \frac{1}{\|\psi\|^2} \int_{-\infty}^{\infty} \left| \psi(t_1) \right|^2 dt_1 = b + at_0. \quad (7.32)$$

Звідси випливає, що центр ваги базисної функції $\psi_{a,b}(t)$ змінюється з масштабом a : чим ширше базисна функція (більше a), тим більше відстань між центрами базисних функцій.

Визначимо другий центральний момент функції $\left| \psi_{a,b}(t) \right|^2$. Маємо

$$\begin{aligned} \Delta_{\psi}^2 &= \frac{1}{\|\psi\|^2} \left\{ \frac{1}{a} \int_{-\infty}^{\infty} (t-b-at_0)^2 \left| \psi \left(\frac{t-b}{a} \right) \right|^2 dt \right\} = \\ &= \frac{1}{\|\psi\|^2} \left\{ a \int_{-\infty}^{\infty} \left(\frac{t-b}{a} - t_0 \right)^2 \left| \psi \left(\frac{t-b}{a} \right) \right|^2 dt \right\}. \end{aligned} \quad (7.33)$$

Позначивши $t_1 = \frac{t-b}{a}$, отримаємо:

$$\Delta_{\psi}^2 = \frac{a^2}{\|\psi\|^2} \left\{ \int_{-\infty}^{\infty} (t-t_0)^2 \left| \psi(t_1) \right|^2 dt_1 \right\} = a^2 \Delta_t^2. \quad (7.34)$$

Тоді ширина (іноді кажуть – радіус) функції $\psi_{a,b}(t)$ в часовій області буде дорівнювати

$$R_{\psi} = a\Delta_t. \quad (7.35)$$

Отже, функція $\psi_{a,b}(t)$ займає «вікно» в часовій області:

$$Win_t = \left[b + at_0 - a\Delta_t, b + at_0 + a\Delta_t \right]. \quad (7.36)$$

Це «вікно» шириною $2a\Delta_t$ буде розташоване близько точки $t_w = b + at_0$. Останнє означає, з одного боку, що значення аналізованої функції $f(t)$, взяте в

деякій точці t_0 , впливає на значення коефіцієнта $CWT_f(a,b)$ з ростом масштабу у все більшому часовому діапазоні, утворюючи в площині (a,b) так званий кут впливу (рис. 7.10,*а*). З іншого боку, величина коефіцієнта $CWT_f(a,b)$ в точці (a_0,b_0) буде визначатися значеннями функції $f(t)$ в околиці точки b_0 , яка визначається тим же кутом впливу (7.10,*б*).

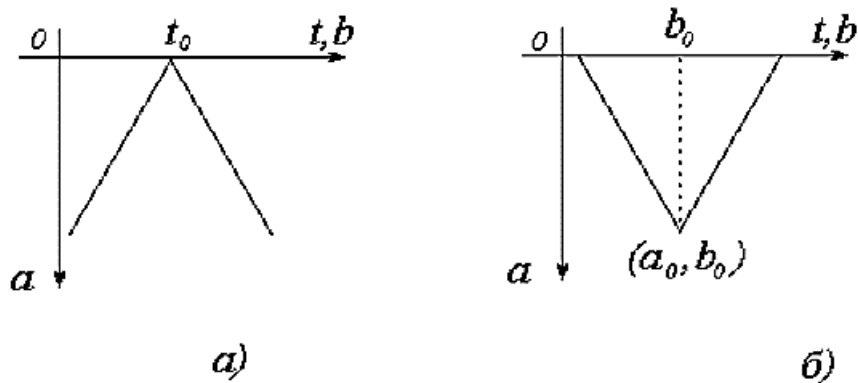


Рис. 7.10. Кут впливу на коефіцієнт $CWT_f(a, b)$ з ростом масштабу (*а*) і на величину коефіцієнта $CWT_f(a_0, b_0)$ (*б*)

З рис. 7.10, зокрема, випливає важливий висновок про те, що дрібномасштабні варіації функції $f(t)$ будуть проявлятися в вейвлет-області при малих значеннях масштабу a , в той час як при великих значеннях масштабу вони будуть згладжуватися, але при цьому будуть проявлятися великомасштабні зміни функції $f(t)$.

Визначимо далі локальні властивості вейвлета $\psi(t)$ в частотній області. Позначимо через ω_0 і $\Delta\omega^2$ величини першого і другого центральних моментів функції $\Psi(\omega)$. Як було зазначено вище, завдяки рівності нулю перших моментів вейвлету $\psi(t)$, функція $\Psi(\omega)$ описує смуговий фільтр з центральною частотою, яка дорівнює ω_0 .

Припустимо, що $\psi(t) \in L_1, f(t) \in L_1 \cap L_2$, а також $F(\omega) \in L_1$. Тоді, враховуючі, що $\Psi_{a,b} = \sqrt{a} \cdot e^{-jb\omega} \Psi(a\omega)$, відповідно до формули Парсеваля отримаємо

$$\int_{-\infty}^{\infty} f^*(t)g(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F^*(\omega)G(\omega)d\omega, \quad (7.37)$$

при $g(t) = f(t)$ маємо

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega, \quad (7.38)$$

тоді отримаємо з (7.16)

$$\begin{aligned} CWT_f(a,b) &= \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt = \\ &= \frac{1}{2\pi} \int \Psi_{a,b}^*(\omega)F(\omega)d\omega = \frac{\sqrt{a}}{2\pi} \int \Psi^*(a\omega)F(\omega)e^{jb\omega}d\omega. \end{aligned} \quad (7.39)$$

Якщо ввести функцію $\eta(\omega) = \Psi(\omega - \omega_0)$, що має нульове середнє значення і ширину $\Delta\omega$, то ВП в Фур'є-області аналогічно (7.39) можна записати у вигляді:

$$CWT_f(a,b) = \frac{a\sqrt{a}}{2\pi} \int \exp\{jb\omega\} \eta\left(a\left(\omega - \frac{\omega_0}{a}\right)\right) F(\omega) d\omega. \quad (7.40)$$

Віконна функція $\eta\left(a\left(\omega - \frac{\omega_0}{a}\right)\right) = \eta(a\omega - \omega_0) = \Psi(a\omega)$ має ширину $\frac{\omega_0}{a}$.

Тому, з точністю до константи і фазового зсуву $e^{jb\omega}$, ВП (7.40) дає локальну інформацію про спектр аналізованої функції $F(\omega)$ в полосі $\pm \frac{\Delta\omega}{a}$ з центром в

точці $\frac{\omega_0}{a}$, тобто у вікні

$$Win_{\omega} = \left[\frac{\omega_0}{a} - \frac{\Delta\omega}{a}, \frac{\omega_0}{a} + \frac{\Delta\omega}{a} \right]. \quad (7.41)$$

При цьому відношення центральної частоти $\frac{\omega_0}{a}$ до ширини вікна $2\frac{\Delta\omega}{a}$,

що дорівнює $\frac{\omega_0}{2\Delta\omega}$, залишається постійним при будь-яких значеннях a .

Частотний аналіз при постійному значенні $\frac{\omega_0}{2\Delta\omega}$ носить назву аналізу з

постійною добротністю системи. Цим ВА відрізняється від звичайного Фур'є-аналізу або Фур'є-аналізу на обмеженому часовому інтервалі. Площа частотно-

часового вікна ВА буде дорівнювати $2a\Delta_t\frac{2\Delta\omega}{a} = 4\Delta_t\Delta\omega$, тобто вона буде

постійною при всіх значеннях масштабу a . Причому, при збільшенні масштабу (зміщення в низькочастотний діапазон аналізу) вікно буде розширюватися по часовій шкалі в площині частота-час (рис. 7.11) і звужуватися у частотній шкалі. При зменшенні масштабу – навпаки.

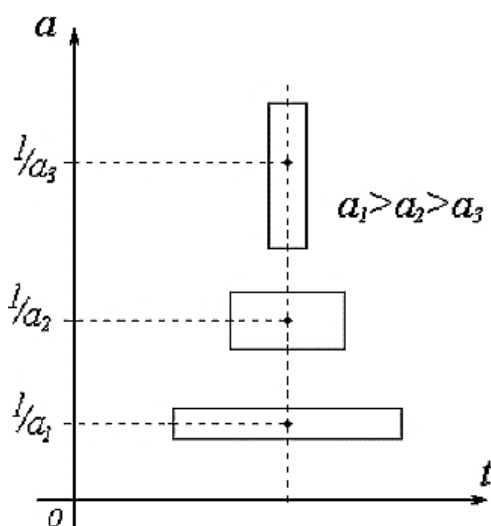


Рис. 7.11. Функції вейвлет-аналізу на площині частота-час

7.5. Дискретне вейвлет-перетворення

Представлення функції $f(t)$ через її безперервне ВП є надмірним. В задачах обробки інформації, що зустрічаються на практиці, сигнал, по-перше,

має обмежену смугу частот і, по-друге, допускаються ті чи інші похибки в одержуваних результатах. Тому використовують дискретне представлення безперервних сигналів, при яких параметри перетворення, в даному випадку a і b , набувають дискретні значення. Вейвлет-перетворення, при якому значення a і b дискретні, називають дискретним ВП (ДВП або DWT – Discrete Wavelet Transform).

7.5.1. Дискретизація масштабу.

Розглянемо спочатку випадок дискретного масштабу a і покладемо $a = a_m = a_0^m$ ($a_0 > 1$). Це рівноцінно розбиттю частотної вісі на піддіпазони (частотні смуги). Припустимо, що $\omega_0 = (a_0 + 1)\Delta\omega$ (це можна зробити завжди, помноживши функцію ψ на деякий модуляційний множник $e^{j\alpha t}\psi(t)$). Тоді частотне вікно дорівнюватиме

$$\left[\frac{\omega_0}{a_m} - \frac{\Delta\omega}{a_m}, \frac{\omega_0}{a_m} + \frac{\Delta\omega}{a_m} \right] = (a_0^{-m+1}\Delta\omega, a_0^{-m+2}\Delta\omega), \quad (7.42)$$

а центральна частота m -го вейвлета:

$$\frac{\omega_0}{a_m} = (a_0 + 1)a_0^{-m}\Delta\omega. \quad (7.43)$$

Базисом для DWT є функція, яка отримана з (7.15) при $a = a_0^m$:

$$\psi_{m,b}(t) = a_0^{\frac{m}{2}} \psi(a_0^{-m}(t-b)). \quad (7.44)$$

Якщо справедливо (7.18) і якщо $\psi(t)$ досить швидко згасає, то будь-яка функція з L_2 може бути представлена у вигляді дискретної послідовності

$$DWT_f(m,b) = \langle f, \psi_{m,b} \rangle = a_0^{\frac{m}{2}} \int_{-\infty}^{\infty} f(t) \psi(a_0^{-m}t - b) dt. \quad (7.45)$$

Для відновлення $f(t)$ по дискретним значенням (7.45) на базис $\psi_{m,b}(t)$ накладаються додаткові обмеження, а саме: Фур'є-образ вейвлета $\psi_{m,b}(t)$ повинен задовольняти співвідношенню

$$A \leq \sum_{m \in \mathbb{Z}} \left| \Psi(a_0^m \omega) \right|^2 \leq B, \quad (7.46)$$

де константи A і B такі, що $0 < A \leq B < \infty$.

Умова (7.46) в термінах радіотехніки має досить прозоре тлумачення. Дійсно, оскільки при кожному значенні масштабу a_0^m вейвлет є смуговим фільтром, то набір (сума) цих фільтрів (блок фільтрів) є деяким пристроєм з нерівномірною частотною характеристикою, яка визначається константами A і B (рис. 7.12).

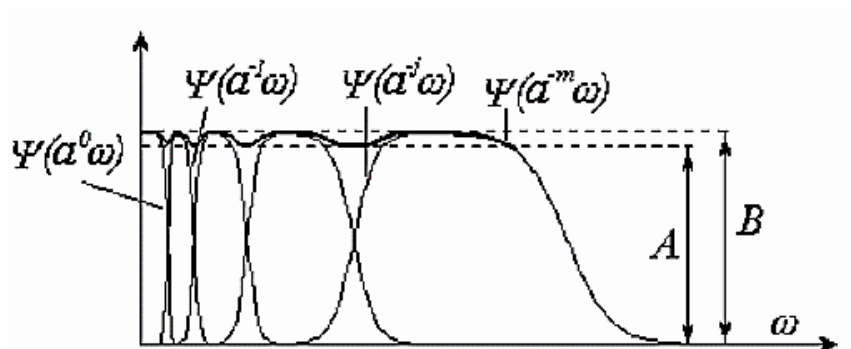


Рис. 7.12. Частотна характеристика смугових фільтрів, утворених вейвлетами $\Psi(a^m \omega)$ з різним масштабом при $m = 0, -1, -2, -3, -4$

Сигнал, наприклад звуковий, на виході такого пристрою при сильній нерівномірності частотної характеристики зазнає суттєвих спотворень. Тому для його відновлення приймають спеціальні заходи, зокрема, встановлюють фільтр, що компенсує спотворення частотної характеристики. У ВП таким фільтром є дуальний (або подвійний) вейвлет $\tilde{\psi}(t)$, Фур'є-образ якого має вигляд:

$$\tilde{\Psi}(\omega) = \frac{\Psi(\omega)}{\sum_n \left| \Psi(a_0^n \omega) \right|^2}. \quad (7.47)$$

Покажемо, що за допомогою такого вейвлета за коефіцієнтами DWT повністю відновлюється сигнал. Дійсно, використовуючи співвідношення Парсеваля (7.37) і формулу (7.47), отримаємо

$$\begin{aligned} & \sum_{m \in Z} \int_{-\infty}^{\infty} DWT_f(m, b) \left\{ a_0^{\frac{m}{2}} \tilde{\psi}(a_0^{-m}(t-b)) \right\} db = \\ & = \sum_{m \in Z} \int_{-\infty}^{\infty} DWT_f(m, b) \left\{ a_0^{\frac{m}{2}} \tilde{\psi}(a_0^{-m}(t-b)) \right\} db = \\ & = \sum_{m \in Z} \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \Psi^*(a_0^m \omega) e^{it\omega} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{it\omega} d\omega = f(t). \end{aligned} \quad (7.48)$$

З (7.46) і (7.47) можна показати, що

$$\frac{1}{B} \leq \sum_{m \in Z} \left| \tilde{\Psi}(a_0^m \omega) \right|^2 \leq \frac{1}{A}. \quad (7.49)$$

7.5.2. Дискретизація масштабу і зсуву. Фрейми.

У цьому випадку вважають дискретними величини a і b , тобто $a = a_m = a_0^m$ і $b = kb_0 a_0^m$. Частотне вікно для аналізу зберігається незмінним. Ширина часового вікна (7.36) дорівнює $2a_0^m \Delta_t$, а середнє значення $(t_0 + kb_0) a_0^m$ змінюється дискретно пропорційно m -му ступеню a_0 – масштабу вейвлета. Чим вужча функція ψ , тобто менша величина a_0^m , тим менший (на ту ж величину) крок зсуву цієї функції.

Базисними функціями для ДВП будуть функції, одержувані з (7.15) при $a = a_0^i$ і $b = kb_0 a_0^i$:

$$\psi_{ik}(t) = a_0^{-i/2} \psi\left(\frac{t - a_0^i k t_0}{a_0^i}\right) = a_0^{-i/2} \psi(a_0^{-i} t - kb_0). \quad (7.50)$$

Коефіцієнти розкладання будь-якої функції з L_2 можуть бути отримані як

$$\begin{aligned} DWT_f(i, k) &= \langle f(t), \psi_{ik}(t) \rangle = \\ &= \int_{-\infty}^{\infty} f(t) \psi_{ik}(t) dt = a_0^{i/2} \int_{-\infty}^{\infty} f(t) \psi(a_0^{-i} t - kb_0) dt. \end{aligned} \quad (7.51)$$

Вираз (7.51) є ДВП функції $f(t)$.

Щоб зворотне перетворення в часову область було справедливим, має виконуватись наступна умова:

$$A \|f\|^2 < \sum_{i, k} |\langle f, \psi_{ik} \rangle|^2 \leq B \|f\|^2 \quad (7.52)$$

для усіх $f \in L_2(Re)$, якщо константи A і B такі, що $0 < A \leq B < \infty$.

У цьому випадку формула для відновлення функції $f(t)$ за коефіцієнтами $DWT_f(i, k)$ буде мати вигляд

$$f(t) = \frac{2}{A+B} \sum_{i, k} DWT_f(i, k) \psi_{ik}(t) + \Delta, \quad (7.53)$$

де похибку відновлення Δ можна оцінити як

$$\|\Delta\| \leq O\left(\frac{B}{A} - 1\right) \|f\|. \quad (7.54)$$

Розділивши усі члени нерівності (7.52) на $\|f\|^2$, можна бачити, що константи A і B є межами нормованої на $\|f\|^2$ енергії – скалярного добутку $\langle f, \psi_{ik} \rangle$. Вони (ці константи) обмежують нормовану енергію коефіцієнтів $DWT_f(i, k)$. Звідси походить термін *фрейм* (frame), яким називають множину функцій $\{\psi_{ik}(t); i, k \in Z\}$, при яких умова (7.52) виконується.

Якщо $A=B$, то $\Delta=0$ і множину $\{\psi_{ik}(t)\}$ називають *щільним фреймом*. При цьому вираз

$$A\|f\|^2 = \sum_{i,k} |\langle f, \psi_{ik} \rangle|^2, \quad (7.55)$$

який випливає з (7.52), є узагальненням теореми Парсеваля на щільні фрейми, для котрих з (7.53) отримуємо

$$f(t) = A^{-1} \sum_{i,k} DWT_f(i,k) \psi_{ik}(t). \quad (7.56)$$

Якщо $A=B=1$, то щільний фрейм стає *ортогональним базисом*.

7.5.3. Приклади вейвлетів для дискретного перетворення.

Як було зазначено вище, функції вейвлет мають властивість частотно-часової локалізації, тобто вони обмежені як в частотній, так і в часовій областях. Нижче розглянемо два приклади: перший – спектр вейвлетів в частотній області є ідеальним смуговим фільтром, другий – самі функції вейвлет є прямокутниками. Усі вейвлети, з точки зору частотно-часових властивостей, займають проміжне положення між цими крайніми випадками.

Sinc-базис. Розіб'ємо вісь частот на інтервали (піддіапазони), як показано на рис. 7.13 при $a_0=2$. Таке розбиття називають логарифмічним, оскільки відношення верхньої і нижньої меж діапазонів постійне і дорівнює 2.

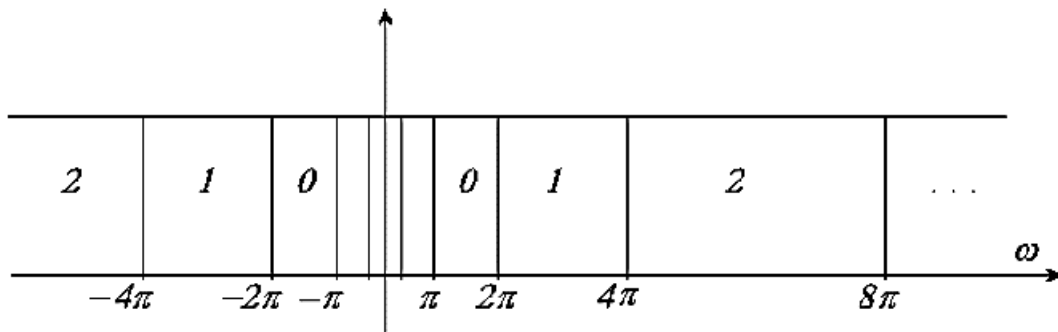


Рис. 7.13. Розбиття вісі частот в логарифмічному масштабі при $a_0=2$

Таке розбиття є ще й ідеальним, оскільки воно реалізується ідеальними смуговими фільтрами. Така ідеалізація потрібна для дослідження властивостей частотного розкладу за допомогою ідеалізованих вейвлетів, що дозволить у подальшому перейти до більш складних розкладань.

Будь-який сигнал $f(t) \in L_2$ зі спектром $F(\omega)$ може займати смугу частот, що охоплює кілька таких піддіапазонів. Тоді $F(\omega) = \sum_m F_m(\omega)$ і $f(t) = \sum_m f_m(t)$, тобто сигнал є сумою деякого числа елементарних сигналів. В даному ідеальному випадку частотні канали не перекриваються, тому має місце ортогональність цих елементарних сигналів, тобто

$$\int f_i(t) f_k(t) dt = \frac{1}{2\pi} \int F_i(\omega) F_k(\omega) d\omega = \delta(i - k). \quad (7.57)$$

Виберемо з усієї множини сигналів такі, які обмежені смугою частот 2^I , тобто такі, які мають спектр $F_I(\omega)$.

Розглянемо періодичну функцію $\Phi_I(\omega)$ таку, що

$$\Phi_I(\omega) = \sum_{k=-\infty}^{\infty} F_I(\omega - 2\pi 2^I k), \quad (7.58)$$

тобто отриману періодизацією $F_I(\omega)$ (рис. 7.14).

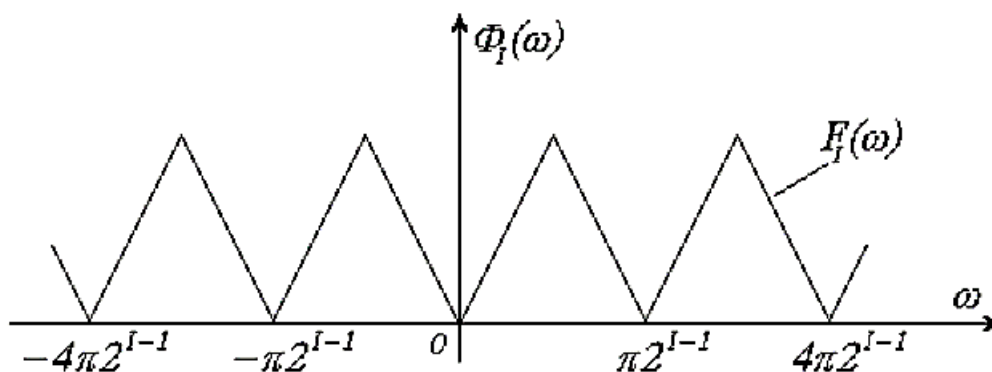


Рис. 7.14. Періодизація функції $F_I(\omega)$

Тоді спектр функції $F_I(\omega)$ при довільному I можна представити у вигляді

$$F_2(\omega) = \hat{O}_2(\omega) \left(\prod \left(\frac{\omega}{\pi 2^{i+1}} \right) - \prod \left(\frac{\omega}{\pi 2^i} \right) \right), \quad (7.59)$$

де $\prod(x)$ – функція вікна така, що

$$\prod(x) = \begin{cases} 1, & -1/2 \leq x \leq 1/2; \\ 0, & \text{інакше.} \end{cases} \quad (7.60)$$

Подивимось, як при цих умовах можна представити функцію $f(t)$ в часовій області. Для цього розкладемо періодичну функцію $\Phi_i(\omega)$ з періодом $T/2 = \pi 2^i$, в ряд Фур'є:

$$\Phi_i(\omega) = \sum_k b_i[k] e^{jk \frac{2\pi}{2\pi 2^i} \omega} = \sum_k b_i[k] e^{jk 2^i \omega}, \quad (7.61)$$

$$b_i[k] = \frac{1}{2\pi 2^i} \int_{-\pi 2^i}^{\pi 2^i} F_i(\omega) e^{-jk 2^i \omega} d\omega. \quad (7.62)$$

Підставляючи (7.61) в (7.59), і виконуючи зворотне перетворення Фур'є, отримаємо:

$$f_i(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_i(\omega) e^{i\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_k b_i[k] e^{jk 2^i \omega} d\omega \cdot \prod \left(\frac{\omega}{\pi 2^{i+1}} \right) \cdot e^{i\omega t} - \quad (7.63)$$

$$- \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_k b_i[k] e^{jk 2^i \omega} \cdot \prod \left(\frac{\omega}{\pi 2^i} \right) \cdot e^{i\omega t} d\omega.$$

Обчислимо перший інтеграл. Переставляючи операції підсумовування і інтегрування, і обмежуючи межі інтегрування із урахуванням функції вікна, отримаємо:

$$\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} b_i[k] \int_{-\pi 2^{i+1}}^{\pi 2^{i+1}} e^{jk 2^i \omega} e^{j\omega t} d\omega = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} b_i[k] \int_{-\pi 2^{i+1}}^{\pi 2^{i+1}} e^{j\omega(t+k 2^i)} d\omega =$$

$$\begin{aligned}
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} b_i[k] \int_{-\pi 2^{i+1}}^{\pi 2^{i+1}} \frac{e^{j\omega(t+k2^{-i})}}{j(t+k2^{-i})} d(j\omega(t+k2^{-i})) = \quad (7.64) \\
&= \frac{2\pi 2^{i+1}}{2\pi} \sum_{k=-\infty}^{\infty} b_i[k] \frac{e^{j(t+k2^{-i})\pi} 2^{i+1} - e^{-j(t+k2^{-i})\pi} 2^{i+1}}{2\pi j(t+k2^{-i}) 2^{i+1}} = \\
&= \sum_{k=-\infty}^{\infty} b_i[k] 2^{i+1} \frac{\sin \pi(2^{i+1}t+k)}{\pi(2^{i+1}t+k)}.
\end{aligned}$$

Аналогічно для другого інтеграла отримаємо:

$$\frac{1}{2\pi} \int_{-\pi 2^i}^{\pi 2^i} \sum_{k=-\infty}^{\infty} b_i[k] \cdot e^{jk2^{-i}\omega} \cdot e^{j\omega t} d\omega = \sum_{k=-\infty}^{\infty} b_i[k] 2^i \frac{\sin \pi(2^i t+k)}{\pi(2^i t+k)}. \quad (7.65)$$

Тоді з (7.63)

$$f_i(t) = \sum_{k=-\infty}^{\infty} b_i[k] \left(2^{i+1} \varphi(2^{i+1}t+k) - 2^i \varphi(2^i t+k) \right) = \sum_{k=-\infty}^{\infty} b_i[k] \Psi_{ik}(t), \quad (7.66)$$

де функція

$$\varphi(x) = \frac{\sin \pi x}{\pi x} = \text{sinc}(x) \quad (7.67)$$

показана на рис. 7.15.

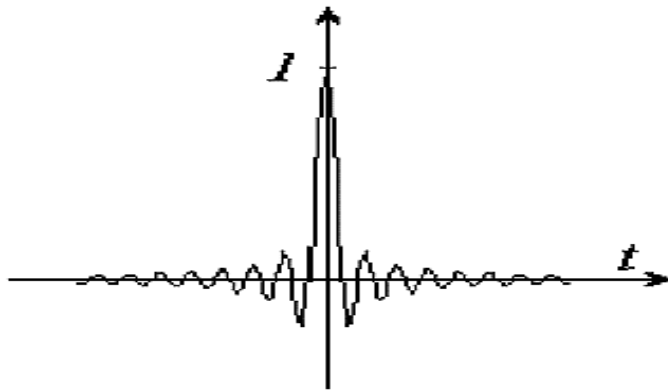


Рис. 7.15. *Sinc*-функція

Сума елементарних сигналів $f_i(t)$ дорівнює

$$f(t) = \sum_{i=0}^{M-1} f_i(t) = \sum_{i,k} a_i[k] \psi_{ik}(t), \quad a_i[k] = 2^{i/2} b_i[k], \quad (7.68)$$

де вейвлет

$$\psi_{ik}(t) = 2^{i/2} \psi(2^i t - k) \quad (7.69)$$

має вигляд:

$$\psi(x) = 2\varphi(2x) - \varphi(x). \quad (7.70)$$

Вираз (7.68) є представленням функції $f(t)$ в базисі вейвлет. В даному окремому випадку ідеальної смугової фільтрації вейвлетом є функція (7.69), утворена з материнської функції $\psi(x)$ за (7.70) з урахуванням (7.68). Такий вейвлет називається *sinc*-вейвлетом (рис. 7.16) на ім'я функції (7.68), яка його утворює.

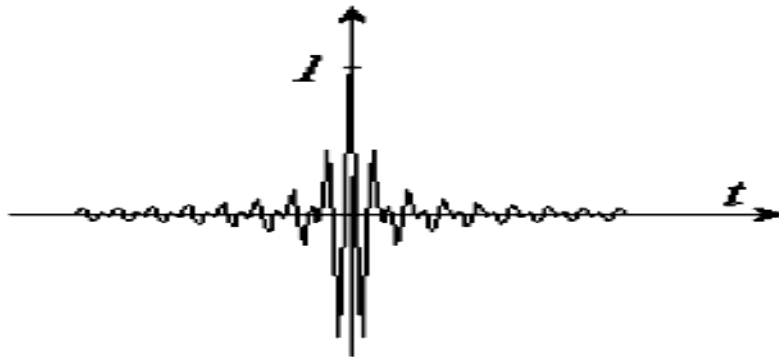


Рис 7.16 – *Sinc*-вейвлет

Множник $2^{i/2}$ при $\psi_{ik}(t)$ необхідний для збереження норми $\|\psi_{ik}(t)\|$

незалежно від величини масштабу, оскільки:

$$\|\psi_{ik}(t)\| = 2^i \int_{-\infty}^{\infty} |\psi(2^i t - k)|^2 dt = \int_{-\infty}^{\infty} |\psi(x)|^2 dx = \|\psi(x)\|. \quad (7.71)$$

Виконавши перетворення Фур'є виразу (7.69), можна бачити, що спектр Фур'є *sinc*-вейвлета є ідеальним смуговим фільтром, який, в загальному випадку, займає смугу частот від $2\pi 2^i$ до $2\pi 2^{i+1}$.

Вейвлет Хаара. Розіб'ємо тепер часову вісь на інтервали (рис. 7.17) і визначимо на одиничному інтервалі функцію

$$\psi(t) = \begin{cases} 1, & 0 < t < 1/2; \\ -1, & 1/2 \leq t < 1; \\ 0, & \text{інакше.} \end{cases} \quad (7.72)$$

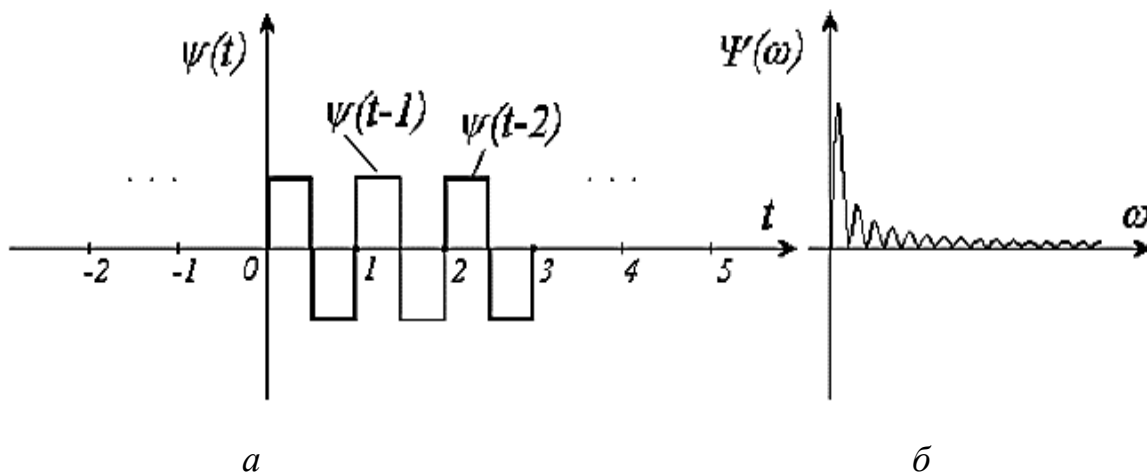


Рис. 7.17. Вейвлет Хаара $\psi(t)$ (а) і спектр Фур'є його амплітуди (б)

Ця функція є материнським вейвлетом, оскільки вона задовольняє умові (7.13). Система зсувів таких функцій $\psi(t-n)$ утворює ортонормальний базис, оскільки їх взаємна енергія дорівнює нулю при $n \neq 0$ і дорівнює одиниці при $n=0$:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1. \quad (7.73)$$

Перетворення Фур'є вейвлета Хаара має вигляд (7.17,б):

$$\psi(\omega) = je^{-j\omega/2} \frac{(\sin \omega / 4)}{\omega / 4}. \quad (7.74)$$

Функції Хаара, як і *sinc*-вейвлет, можуть бути отримані за допомогою масштабної функції (рис. 7.18):

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1; \\ 0, & \text{інакше.} \end{cases} \quad (7.75)$$

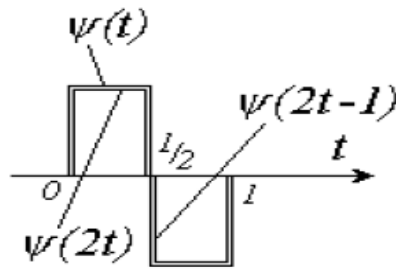


Рис. 7.18. Формування вейвлета Хаара за допомогою масштабних функцій

З наведених прикладів випливає ряд висновків:

1. Представлення вейвлет-функції у вигляді прямокутників в будь-якій з областей (частотній або часовій) веде до нескінченного розширення в протилежній сфері (порівняйте рис. 7.14 і 7.16; рис. 7.17,*а* і 7.17,*б*).

Отже, для того, щоб функції вейвлет були локалізовані одночасно в часовій і частотній областях, вони повинні спадати з ростом аргументу, принаймні, за законом зворотної пропорційності (див. формулу (7.12)).

2. Вейвлети $\psi(t)$, спектри Фур'є які є смуговими фільтрами, можуть бути виражені через масштабні функції $\varphi(t)$, спектри Фур'є яких є фільтрами нижніх частот (див. формули (7.70) і (7.75)).

3. Базисні функції для DWT можуть бути отримані з однієї материнської функції шляхом її масштабування і зсуву (див. формули (7.69) і (7.70)).

4. Будь-який сигнал $f(t)$ з L_2 може бути представлений своїм вейвлет-розкладанням (7.68), якщо число компонентів $f_i(t)$ таке, що вони займають смугу частот більшу, ніж смуга сигналу.

7.6. Швидкий алгоритм дискретного вейвлет-перетворення

ДВП функції $f(t) \in L_2$ називається наступне її розкладання (для наочності поетапно розглянемо процес розкладання):

$$f^1(t) = \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{1,k} \varphi\left(\frac{t-k}{2}\right) \cdot s_{1,k};$$

$$f^2(t) = \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot d_{2,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot s_{2,k};$$

$$f^3(t) = \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot d_{2,k} + \\ + \sum_{3,k} \psi\left(\frac{t-k}{8}\right) \cdot d_{3,k} + \sum_{3,k} \psi\left(\frac{t-k}{8}\right) \cdot s_{3,k};$$

$$f^4(t) = \sum_{1,k} \psi\left(\frac{t-k}{2}\right) \cdot d_{1,k} + \sum_{2,k} \psi\left(\frac{t-k}{4}\right) \cdot d_{2,k} + \sum_{3,k} \psi\left(\frac{t-k}{8}\right) \cdot d_{3,k} + \\ + \sum_{4,k} \psi\left(\frac{t-k}{16}\right) \cdot d_{4,k} + \sum_{4,k} \psi\left(\frac{t-k}{16}\right) \cdot s_{4,k},$$

тоді на j -му кроці маємо:

$$f^j(t) = \sum_{j=1}^N \sum_k \psi\left(\frac{t-k}{2^j}\right) \cdot d_{j,k} + \sum_k \varphi\left(\frac{t-k}{2^j}\right) \cdot s_{N,k}, \quad (7.76)$$

$$\psi(t-k) = (-1)^n \varphi(t-N+k) \quad (7.77)$$

є сімейством базисних функцій; j, k – відповідно глибина розкладання і крок підсумовування, а $d_{j,k}$ і $s_{j,k}$ – коефіцієнти розкладання:

$$d_{j,k} = \sum_k \psi\left(\frac{t-k}{2^j}\right) f(t); \quad s_{j,k} = \sum_k \varphi\left(\frac{t-k}{2^j}\right) f(t). \quad (7.78)$$

Коефіцієнти розкладання $d_{j,k}$ і $s_{j,k}$ для кожного значення j і k відмінні від нуля також протягом лише декількох кроків підсумовування, тобто короткої області сигналу. Для виконання ДВП був розроблений швидкий алгоритм, у якому перехід до представлення сигналу $f(t)$ зі змінною роздільною здатністю за частотою і часом виконується за допомогою банку фільтрів.

Всі фільтри розкладання $H_0^j(z)$, $F_0^j(z)$ і відновлення $H_1^j(z)$, $F_1^j(z)$ мають коефіцієнти, однакові із фільтрами на першому ступені розкладання /

відновлення. Таким чином, розрахунок фільтрів виконується одноразово. Дана процедура відповідає визначенню базисної вейвлет-функції.

Шкалування виконується за рахунок зміни частоти дискретизації. Кількість ступенів банку відповідає числу етапів шкалування вейвлет-функції. В силу простоти реалізації та мінімальної складності обчислень (яка складає близько N – довжину сигналу) даний алгоритм називається *швидким вейвлет-перетворенням* (ШВП).

Створення алгоритму ШВП уможливило застосування на практиці переваг ВП:

- 1) локалізованого частотно-часового представлення довільного сигналу з одночасним інформативним розподілом за частотою і часом;
- 2) компактного запису сигналу у вигляді комбінації коефіцієнтів розкладання, кінцевої для більшості сигналів, зумовленої подібною формою сигналу і базисних функцій;
- 3) ефективного стиску (компресії) сигналу;
- 4) нелінійної фільтрації.

Нарощування банку виконується зазвичай по низькочастотній гілці (рис. 7.19). Однак, у ряді випадків, наприклад, для сигналу електрокардіограми, доцільно змінити структуру дерева (рис. 7.20).

Теоретично вибір оптимально інформативного шляху розкладання може бути визначений шляхом оцінки ентропійної характеристики кожного вузла (рівня) розкладання:

$$\Xi^j = \sum_k \psi(2^{-j}x - k) \cdot f(x) \cdot \log \left[\psi(2^{-j}x - k) \cdot f(x) \right] = \sum_k d_{jk} \cdot \log d_{jk}, \quad (7.79)$$

де d_{jk} – вейвлет-коефіцієнти j -го вузла. Формування дерева йде по шляху зменшення ентропійної оцінки.

Зворотне перетворення може бути отримано за допомогою дуальної структури (рис. 7.21 відповідає прямому ДВП на рис. 7.19).

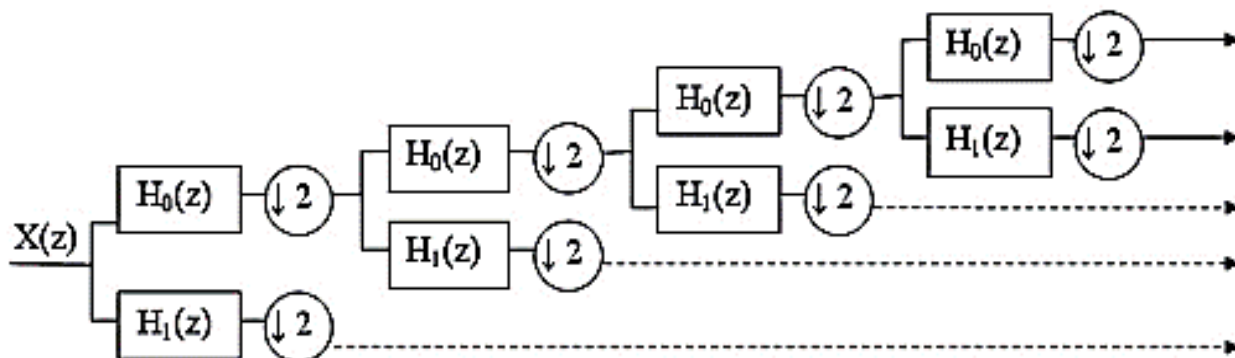


Рис. 7.19. Каскадування банку фільтрів по низькочастотній гілці

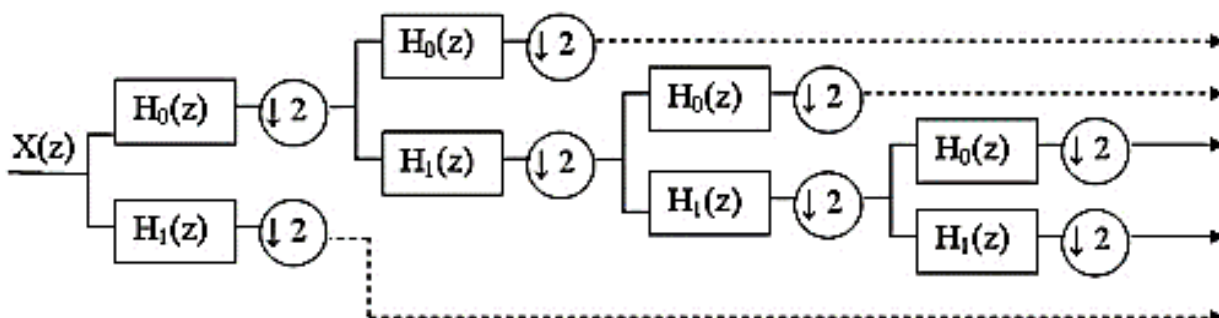


Рис. 7.20. Приклад оптимізованого банку фільтрів

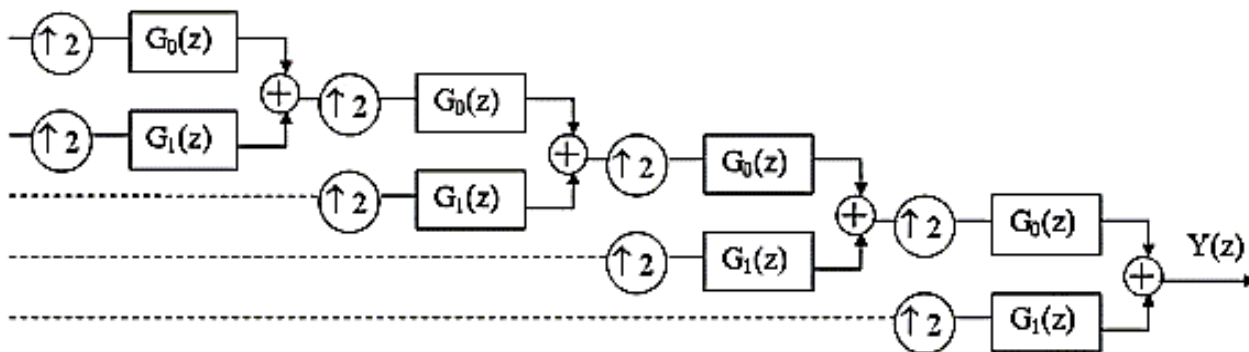


Рис. 7.21. Банк фільтрів для зворотного ДВП сигналу

Число рівнів розкладання j може бути довільним. В (7.79) $\varphi\left(\frac{t-k}{2^j}\right)$ і $\psi\left(\frac{t-k}{2^j}\right)$ – імпульсні характеристики (ІХ) пари фільтрів; множник 2^j відповідає розтягуванню ІХ фільтрів, тобто звуженню їх смуги пропускання, крок підсумовування k визначається інтервалом дискретизації вхідного сигналу.

Послідовності $\{d_{j,k}\}$ і $\{s_{j,k}\}$ отримуються при згортці $f(t)$ з ІХ фільтрів. Розробник також може отримати готові коефіцієнти вейвлет-фільтрів, користуючись засобами пакетів Matlab, Matematika та інших.

При представленні ВП у вигляді структури фільтрів видно, що в процесі ВП з сигналу виділяється його основна низькочастотна огинаюча – $s_{N,k}$ і ряд його високочастотних особливостей – $d_{j,k}$. Описана техніка ВП дозволяє варіювати тривалість часового вікна. Для визначення низькочастотних складових сигналу потрібні більш тривалі інтервали часу спостереження сигналу, ніж при виявленні його високочастотних компонент, де потрібні відносно короткі часові вікна.

7.7. Узагальнений алгоритм частотно-часового вейвлет-аналіза

Узагальнений алгоритм частотно-часового ВА містить наступні етапи:

1. Вибрати базис ВА – базисну вейвлет-функцію $\psi(t)$ і парну їй функцію, яка масштабується $\varphi(t)$: $\{\varphi(t), \psi(t)\}$, з огляду на передбачувану природу сигналу $f(t)$ (табл. 7.1).

2. Обчислити коефіцієнти на всьому протязі $f(t)$ за допомогою зсуву $\varphi(t-k)$, $\psi(t-k)$.

3. Обнулити зсув і виконати шкалування.

4. Виконати черговий прохід сигналу за допомогою зсуву і повторити для всіх значень шкального множника a .

Після закінчення ВП маємо набір вейвлет-коефіцієнтів $CWT_f(a,b)$, кожен з яких є результатом розкладання $f(t)$ за шкалами і зсувів базисних функцій.

Розглянемо докладніше співвідношення між поняттями шкала і частота. Зі зменшенням шкального множника a , отже, при стисненні вейвлет-функції

виявляються його складові високої частоти. При значеннях шкального множника $a \cong 1.0$, тобто при розтягнутій функції, що має велику протяжність, в процесі обчислення коефіцієнтів базисна функція порівнюється з більшою областю сигналу $f(t)$, а значення коефіцієнтів описують входження в сигнал його низькочастотних складових. Значення параметра k служать при цьому мітками в часовій області.

Властивості базисних функцій вейвлет-аналізу

Таблиця 7.1

Група вейвлет-функцій	Вейвлет-функції грубої апроксимації		Нескінченно диференційовані вейвлети	Ортогональні функції із компактним носієм				Біортогональні
	Morlet	Mexican Hat		Meyer	Haar	Daubechies	Coiflet	
Тип вейвлет-функції	Morlet	Mexican Hat	Meyer	Haar	Daubechies	Coiflet	Symlet	B-spline
Ортогональність	-	-	+	+	+	+	+	-
Компактний носій	-	-	-	+	+	+	+	+
Наявність зворотного перетворення	-	-	+	+	+	+	+	+
Апаратна реалізація	-	-	за допомогою НІХ-фільтрів	реалізується за допомогою КІХ-фільтрів				
Інші властивості	симетрія, явний математичний опис		симетрія, нескінченна	симетрія, явний математичний опис				

Процес отримання вейвлет-коефіцієнтів досліджуваного сигналу $f(t)$ і їх подальший аналіз дозволяє не тільки дослідити реальні сигнали довільної форми, але й виконувати більш ефективно стиснення (компресію) сигналу або зображення, а також видалення шуму (фільтрацію).

Перехід від частотно-часового перетворення до час-шкального обумовлюється також і тим, що усереднення всплесків сигналу в процесі підсумовування за різними шкалами робить ВП менш чутливим до шуму.

7.8. Перетворення Гільберта-Гуанга

Під *перетворенням Гільберта-Гуанга* (Hilbert-Huang transform – ННТ) розуміється емпіричний метод декомпозиції (EMD) нелінійних і нестационарних процесів у поєднанні з гільбертовим спектральним аналізом (HSA). ННТ є частотно-часовим аналізом даних (сигналів) і *не вимагає апріорного* функціонального базису. Функції базису виходять адаптивно безпосередньо з даних процедурами відсіювання EMD. Миттєві частоти обчислюються від похідних фазових функцій гільбертовим перетворенням функцій базису. Заключний результат представляється в частотно-часовому просторі.

EMD-HSA був запропонований Норденом Гуангом у 1995 р. в США (NASA) для вивчення поверхневих хвиль тайфунів і узагальнений на аналіз довільних часових рядів колективом співавторів у 1998 р. У наступні роки, у міру розширення застосування EMD-HSA для інших галузей науки і техніки, замість терміна EMD-HSA був прийнятий більш короткий термін – перетворення ННТ.

Традиційні методи аналізу даних призначені, як правило, для лінійних і стаціонарних сигналів і систем, і тільки останнім часом почали активно розвиватися методи аналізу нелінійних, але стаціонарних і детермінованих систем, і лінійних, але нестационарних даних (ВА, розподіл Wagner-Ville тощо). Тим часом, більшість природних матеріальних процесів, реальних фізичних систем і відповідних цим процесам і системам даних у тій чи іншій мірі є нелінійними і нестационарними, та при аналізі даних використовуються певні спрощення, особливо щодо апріорно встановлюваного базису аналізу даних.

Необхідна умова коректного подання нелінійних і нестационарних даних полягає у тому, щоб мати можливість формування адаптивного базису, який функціонально залежить від змісту самих даних. Такий підхід й реалізується в методі ННТ.

7.8.1. Емпіричний метод декомпозиції.

Функції внутрішніх мод сигналів. Декомпозиція сигналів заснована на припущенні, що будь-які дані складаються з різних режимів коливань. У будь-який момент часу дані можуть мати багато різноманітних існуючих режимів коливань, нанесених одне на одне. Кожен режим, лінійний або нелінійний, являє собою просте коливання, яке має екстремуми і нульові перетини. Крім того, коливання буде у певній мірі «симетричне» щодо локального середнього значення. Результат – кінцеві складні дані.

Кожен з цих коливальних режимів може бути представлений функцією внутрішньої моди (Intrinsic Mode Function – IMF) із наступним визначенням:

1. Число екстремумів і число нульових перетинів функції повинні бути рівними або відрізнятись не більше ніж на 1.

2. У будь-якій точці функції середнє значення огинаючих, визначених локальними максимумами і локальними мінімумами, має бути нульовим.

IMF є коливальним режимом, як частина простої гармонійної функції, але замість постійної амплітуди і частоти (як в простій гармоніці), у IMF можуть бути змінна амплітуда і частота, як функції незалежної змінної (часу, координати тощо).

Будь-яку функцію і будь-який довільний сигнал можна розділити на сімейство функцій IMF, дотримуючись викладеної нижче методики. Для наочності методику реалізації EMD розглянемо на прикладі розкладання цифрового масиву модельного сигналу $y(k)$, представленого на рис. 7.22.

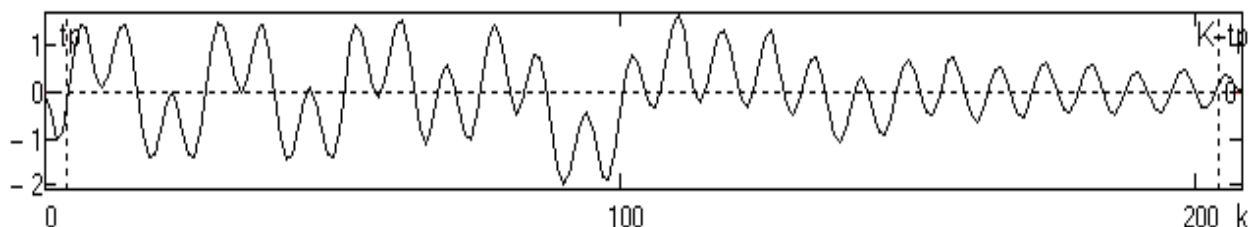


Рис. 7.22. Модельний сигнал $y(k)$

Сигнал змодельований сумою трьох нестационарних за амплітудою гармонік різної частоти на інтервалі відліків по k від 0 до 200 і продовжений на початковій і кінцевій ділянках на інтервали $tp=4$ для завдання початкових і кінцевих умов перетворення і усунення похибок перетворення на кінцевих інтервалах оброблюваного масиву даних.

Процес відсіву функцій IMF. Алгоритм емпіричної декомпозиції сигналу складається із наступних операцій його перетворення.

Операція 1. Ідентифікуємо за координатами і амплітудами всі локальні екстремуми (максимуми і мінімуми) сигналу (рис. 7.23). Групуємо окремо масиви векторів координат (номерів відліків) $x_{\max}(k)$ і відповідних амплітудних значень $y_{\max}(k)$ максимумів, і аналогічні масиви векторів $x_{\min}(k)$ і $y_{\min}(k)$ мінімумів всіх виділених екстремумів.

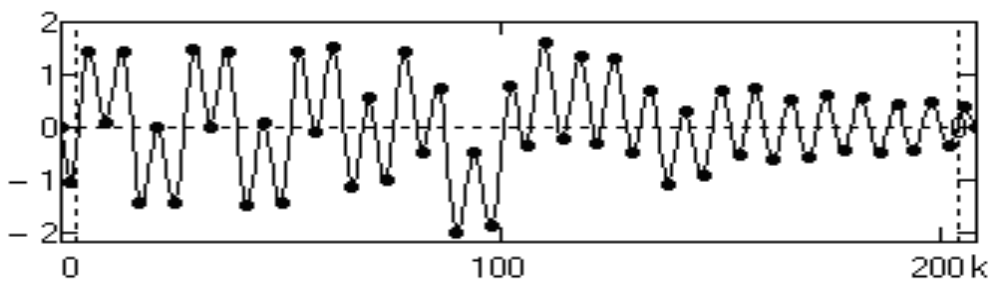


Рис. 7.23. Локалізація екстремумів

Операція 2. Кубічним (або будь-яким іншим) сплайном обчислюємо верхню і нижню огинаючі сигналу по виділеним максимумам і мінімумам, як це показано на рис. 7.24 (криві 1 і 2 відповідно).

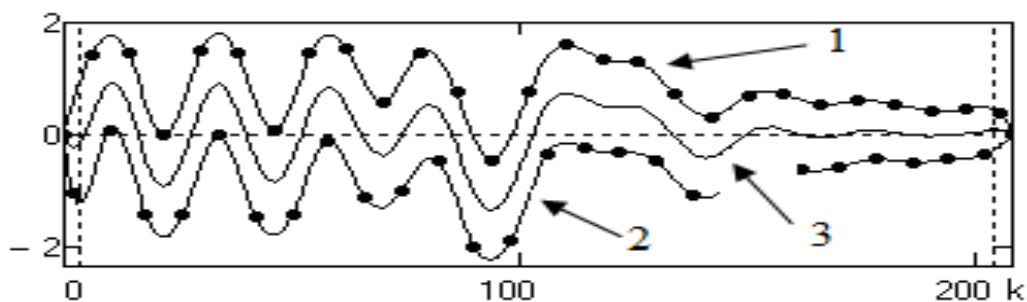


Рис. 7.24. Інтерполяція екстремумів

Визначаємо функцію середніх значень $m_1(k)$ між огинаючими (крива 3) і знаходимо перше наближення до першої функції IMF:

$$h_1(k) = y(k) - m_1(k). \quad (7.80)$$

Операція 3. Повторюємо операції 1 і 2, приймаючи замість $y(k)$ функцію $h_1(k)$, і знаходимо друге наближення до першої функції IMF – функцію $h_2(k)$

$$h_2(k) = h_1(k) - m_2(k). \quad (7.81)$$

Аналогічно знаходимо третє і наступні наближення до першої функції IMF. У міру збільшення кількості ітерацій функція $m_i(k)$, так само як і функція $h_i(k)$, прагне до незмінної форми. З огляду на це, природним критерієм зупинки ітерацій є завдання певної межі по нормалізованій квадратичній різниці між двома послідовними операціями наближення, яка визначається як

$$\delta = \sum_k |h_{i-1}(k) - h_i(k)|^2 / \sum_k h_{i-1}^2(k). \quad (7.82)$$

Приклад зміни значень δ в процесі ітерацій наведено на рис. 7.25. При порозі $\delta = 10^{-4}$ кількість ітерацій, як правило, не перевищує 6-8.



Рис. 7.25. Залежність порогу зупинки δ від кількості ітерацій

Зупинка ітерацій за нормалізованою квадратичною різницею (7.82) історично був першим із застосування. Однак для складних і об'ємних (за кількістю відліків) сигналів в процесі ітерацій може змінюватися кількість виділених екстремумів, при цьому, як правило, спостерігаються сплески значення δ в більшу сторону, починаючи з яких знову починається процес

зменшення δ і загальна кількість ітерацій може збільшуватися до 20-30 без істотного підвищення якості відсіву. Досвід показує, що для оптимальних відсіювань число ітерацій близько 6-8 є цілком достатнім, і все більше застосувань використовують зупинку за максимальною ітерацією, при якій число екстремумів і нульових перетинань функції $h(k)$ не змінюється у порівнянні з попередніми ітераціями.

Останнє значення $h_i(k)$ ітерацій приймається за найбільш високочастотну функцію $c_1(k)=h_i(k)$ сімейства IMF, яка безпосередньо входить до складу вихідного сигналу $y(k)$. Це дозволяє вирахувати $c_1(k)$ зі складу сигналу і залишити в ньому більш низькочастотні складові (рис. 7.26):

$$r_1(k) = y(k) - c_1(k). \quad (7.83)$$

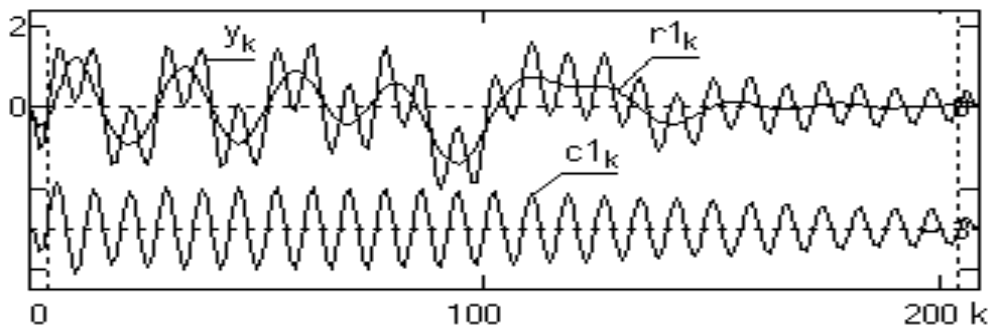


Рис. 7.26. Процес відсіювання найбільш високочастотної складової $c_1(k)$

Функція $r_1(k)$ обробляється як нові дані за аналогічною методикою зі знаходженням другої функції IMF – $c_2(k)$:

$$r_2(k) = r_1(k) - c_2(k), \quad \text{і так далі.} \quad (7.84)$$

Таким чином, досягається декомпозиція сигналу в n – емпіричному наближенні:

$$y(t) = \sum_n c_n(t) + r_n(t). \quad (7.85)$$

Критерії зупинки процесу декомпозиції сигналу можуть бути наступними:

1. Залишок $r_n(k)$ у всьому інтервалі завдання сигналу стає несуттєвим за своїми значеннями у порівнянні з сигналом.

2. Залишок $r_n(k)$ стає монотонною функцією, з якої не може бути вилучено функцій IMF.

3. Оскільки в кінцевому підсумку підсумовування всіх функцій IMF (реконструкція сигналу) має давати початковий сигнал, то можна зупиняти розкладання завданням відносної похибки середньоквадратичної реконструкції (без урахування залишку $r_n(k)$).

4. В міру збільшення кількості функцій IMF відносна середньоквадратична похибка реконструкції досить складних і протяжних сигналів зменшується, але, зазвичай, має певний мінімум. Мабуть, це визначається спробами алгоритму розкласти залишок на функції, які частково компенсують одна одну. Відповідно, зупинка програми може виконуватися, якщо наступна виділена функція IMF збільшує похибку реконструкції.

Іншими словами, зупинка декомпозиції сигналу повинна відбуватися при максимальному «випрямленні» залишку, тобто перетворення його в тренд сигналу по інтервалу завдання з числом екстремумів не більше 3. Навіть для даних з нульовим середнім значенням кінцевий залишок може відрізнятись від нуля. Щоб застосовувати метод EMD, центрування даних не потрібно, метод потребує тільки локалізації екстремумів. Нульова лінія для кожної компоненти декомпозиції буде сформована процесом відсіювання.

На рис. 7.27 наведено приклад повної декомпозиції сигналу із зупинкою за критерієм 2. На верхньому графіку рисунка наведено вхідний сигнал перетворення і сигнал зворотної реконструкції підсумовуванням функцій розкладання c_i (c_1 - c_5).

Компоненти EMD зазвичай фізично значимі, оскільки характеристичні параметри функцій IMF визначаються матеріальними даними.

Ортогональність базису декомпозиції. Таким чином, вхідний сигнал $y(k)$ відповідно до виразу (7.85) розкладається по базису, який не визначений аналітично, але задовольняє усім традиційним вимогам базису. На підставі перевірки на модельних і експериментальних даних він є:

- закінченим і таким, що сходиться (сума усіх функцій IMF і залишку дорівнює початковому сигналу й не залежить від критеріїв зупинки ітерацій);
- ортогональним (усі IMF і залишок ортогональні один одному);
- єдиним.

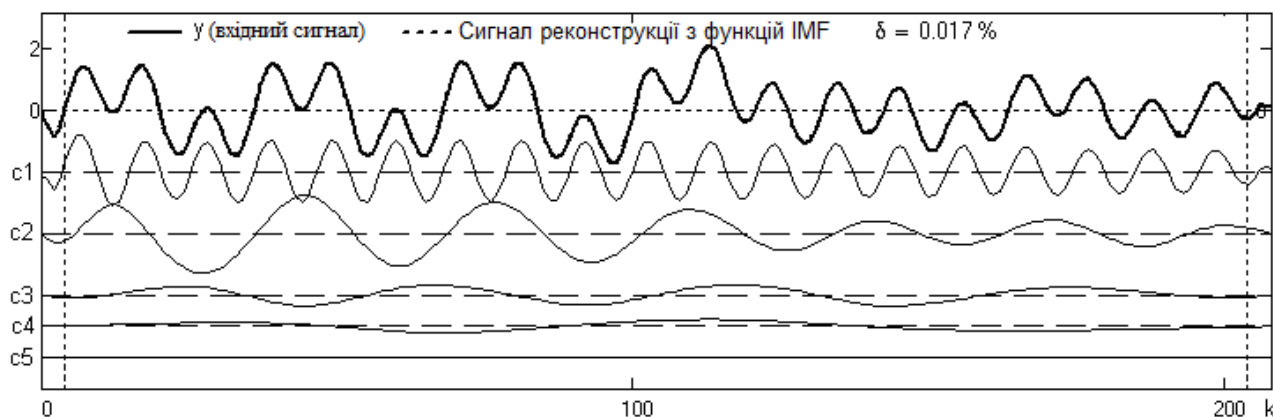


Рис. 7.27. Повна декомпозиція сигналу

І, що найголовніше – він є адаптивним, оскільки отриманий безпосередньо з аналізованих даних емпіричним методом.

Ортогональність базису легко може бути перевірена скалярним добутком будь-яких пар компонентів IMF. Сума (7.85) усіх компонентів IMF, включаючи залишок, повинна реконструювати вхідний сигнал і може використовуватися для визначення похибки декомпозиції. Зазвичай, найбільші локальні похибки декомпозиції спостерігаються на кінцевих ділянках вхідного масиву даних. Щоб виключити похибки рекомендується ставити інтервали початкових і кінцевих умов, а сигнал на цих інтервалах формувати будь-якою функцією прогнозування, або продовжувати (парно або непарно) функцією самого сигналу.

Н. Гуанг стверджує також, що базис розкладання є єдиним. Але це твердження можна вважати спірним. Емпіричний процес розкладання сигналу в силу своєї адаптивності некерований, принаймні, в цій формі. Навіть монотональна локальна складова сигналу при певному впливі дестабілізуючих факторів (шумів, імпульсних завад тощо) може при декомпозиції розділитися

на дві або три функції IMF. Звичайно, при підсумовуванні цих функцій така локальна складова може бути виділена, але це потребує від користувача певних апріорних знань про склад сигналів.

7.8.2. Спектральний аналіз Гільберта.

IMF, визначені вищенаведеним способом, допускають обчислення фізично значущих миттєвих частот, що дає можливість створити частотно-часове представлення сигналу на основі *перетворення Гільберта*.

Для дійсного сигналу $s(t)$ перетворення Гільберта визначається головним значенням (PV) інтеграла

$$H_s(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau. \quad (7.86)$$

Це призводить до аналітичного сигналу

$$z(t) = s(t) + jH_s(t) = a(t) \exp(j\theta(t)), \quad (7.87)$$

$$a(t) = \sqrt{s(t)^2 + H_s(t)^2}, \quad \theta(t) = \arctg \left[\frac{H_s(t)}{s(t)} \right], \quad (7.88)$$

де $a(t)$ і $\theta(t)$ – відповідно, *амплітуда* і *фаза* сигналу під час t .

Миттєва частота може бути обчислена за формулою

$$\omega(t) = d\theta(t) / dt. \quad (7.89)$$

Зауважимо, що вищенаведені як амплітуда, так і миттєва частота – функції часу. Після виконання перетворення Гільберта на кожній компоненті IMF початкові дані можуть бути представлені як дійсні у такій формі:

$$x(t) = \Re \left\{ \sum_{j=1}^n a_j(t) \exp \left[i \int \omega_j(t) dt \right] \right\}. \quad (7.90)$$

Тут, залишок r_n був не врахований, оскільки це або монотонна функція, або постійна. Ті ж самі дані в представленні Фур'є мають вигляд:

$$x(t) = \Re \left[\sum_{j=1}^n a_j e^{i\omega_j(t)t} \right]. \quad (7.91)$$

На рис. 7.28 наведено зіставлення частотно-часового представлення модельного сигналу у трьох формах.

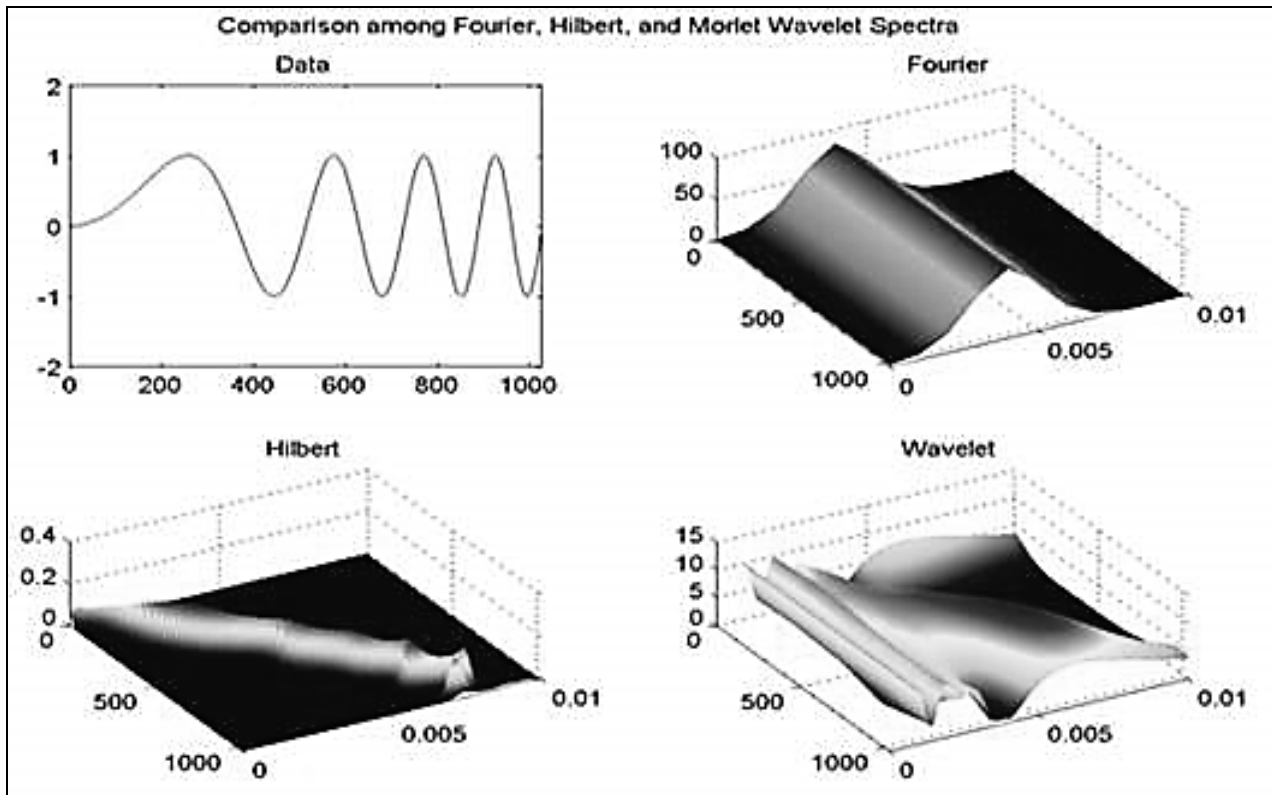


Рис. 7.28. Частотно-часове представлення модельного сигналу в трьох формах

7.8.3. Емпірична декомпозиція шумових сигналів.

Шуми, які супроводжують корисну інформацію в сигналі, в принципі, не належать до типу коливальних в прямому сенсі цього поняття. Але вони повністю задовольняють наведеним вище визначенням функцій IMF. При розподілі в усьому частотному діапазоні вхідного сигналу і виконанні EMD, вони розподіляються по всіх функціях IMF. Оскільки інформація в головному частотному діапазоні дискретних сигналів зазвичай є низькочастотною, то шуми «відсіваються», в основному, в високочастотні функції IMF. Але в ці функції можуть «просочуватись» і високочастотні спектральні складові

інформаційної частини сигналу в залежності від їх положення в головному частотному діапазоні. Відповідно, на перший план висувається задача формування певних критеріїв відбору тільки шумових функцій IMF (для виключення їх при подальшій реконструкції сигналу) і впливу на цей відбір як статистичних та спектральних характеристик самих шумів, так і спектрального складу корисної інформації в сигналі.

EMD «білого шуму». Білий шум є стаціонарним випадковим процесом $q(t)$, автокореляційна функція якого описується δ -функцією Дірака і, відповідно, спектральна щільність потужності шумів не залежить від частоти і має постійне значення $W_q(f) = \sigma^2$, рівне дисперсії значень $q(t)$. По суті, це ідеалізований випадковий процес з нескінченною енергією. Але в разі постійності спектральної щільності потужності випадкового процесу в кінцевому діапазоні частот це істотно спрощує аналіз сигналів. Багато завод в радіотехніці, в телекомунікаціях та в інших галузях техніки розглядають як білий шум, якщо ефективна ширина спектру сигналів набагато менша ефективної ширини спектра шумів, і спектральна щільність потужності шумів слабо змінюється в інтервалі спектра сигналу.

Поняття «білий шум» визначає тільки спектральну характеристику випадкового процесу, а, отже, під це поняття підпадають будь-які випадкові процеси, що мають рівномірний енергетичний спектр і різні закони розподілу. На рис. 7.29 наведена гістограма одиничної реалізації модельного «білого шуму» $y(k)$ в системі Mathcad (5000 відліків) з рівномірним розподілом відліків від -0.55 до +0.55 і дисперсією 0.1.



Рис. 7.29. Гістограма шумів

Спектральну щільність потужності модельної реалізації шуму можна спостерігати на рис. 7.30. Тут наведено спектри щільності потужності сигналу $y(k)$ і перших семи функцій IMF в головному частотному діапазоні сигналу $0-\pi$ (5000 відліків шуму з $\Delta k=1$, відліки по спектру $0-2500$ з кроком $\Delta\omega=\pi/2500$).

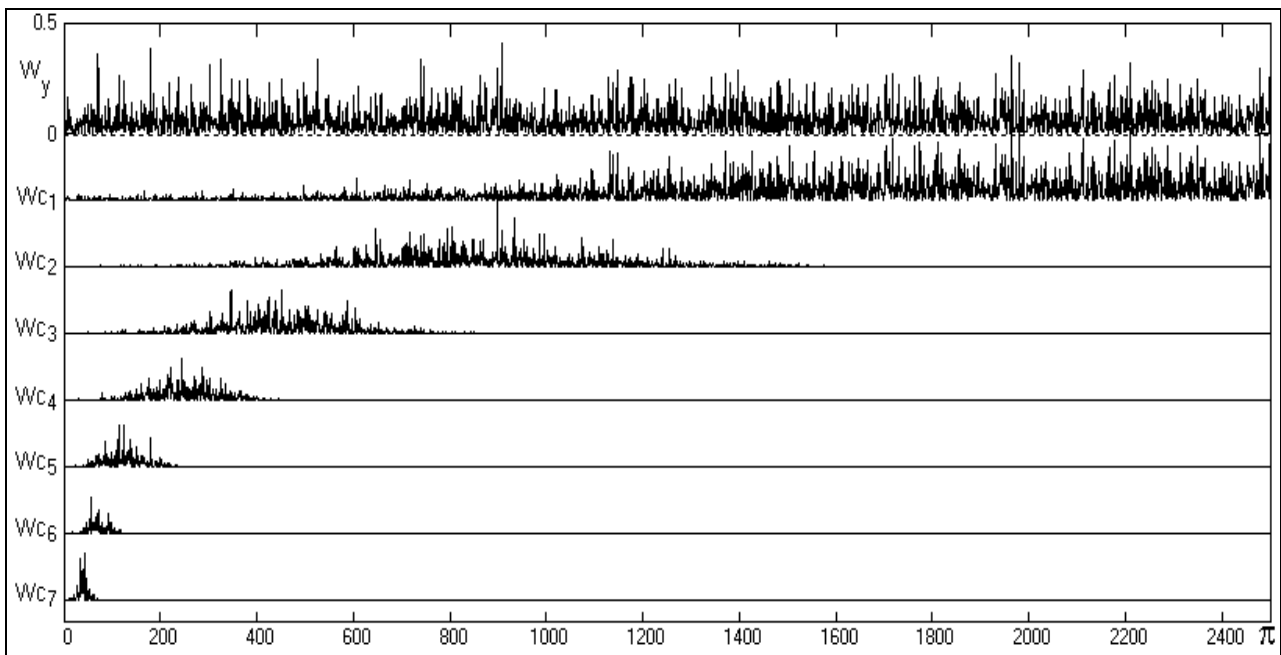


Рис. 7.30. Спектральна щільність потужності модельної реалізації шуму

На рис. 7.31 наведено результат EMD моделі $y(k)$ шуму на перших 750 відліках. Шумовий сигнал даної реалізації був розкладений на 12 функцій IMF, перші дев'ять з яких наведені на рисунку (з яких 4 останніх функції у збільшеному масштабі).

Зупинка процесу декомпозиції виконана за мінімумом похибки реконструкції без урахування залишку (рис. 7.32).

Кількість функцій IMF в різних реалізаціях випадкового сигналу змінюється від 8 до 14. Зупинка ітерацій при обчисленні кожної функції IMF була встановлена за відносною розбіжністю між послідовними ітераціями з порогом 0.01%, при цьому кількість ітерацій для перших функцій IMF досягає 30 і, як правило, поступово знижується. Похибка реконструкції з урахуванням

залишку практично дорівнює нулю. Обчисленням скалярного добутку будь-яких двох функцій IMF можна переконатися в їх взаємній ортогональності.

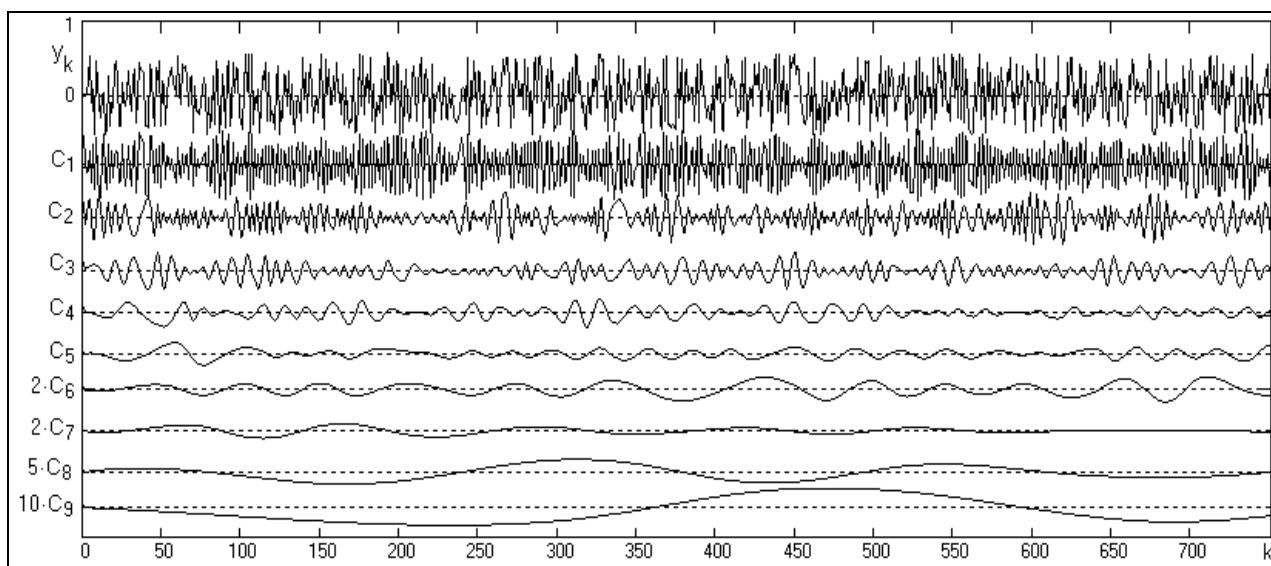


Рис. 7.31. Процес декомпозиції модельного «білого шуму» $y(k)$

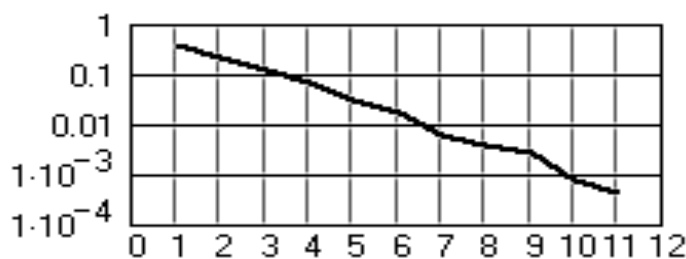


Рис. 7.32. Залежність похибки від кількості функцій IMF

На рис. 7.33,*a* наведено гістограми перших чотирьох IMF в зіставленні із гістограмою вхідного сигналу $y(k)$. Як впливає з цих графіків, EMD істотно змінює щільності розподілу вихідних функцій. Розподіл першої IMF стає двомодальним з прогином униз на малих (близьких до нульових) амплітудах. Це пояснюється тим, що для поруч розташованих однополярних імпульсів при EMD виділяються екстремуми імпульсів більшої амплітуди, які і відсіваються в першу функцію IMF. При відніманні цієї функції з вхідного сигналу розподіл решти шумів стає близьким до гаусового із нульовим середнім значенням і різким скороченням поруч розташованих однополярних імпульсів. На відборі

всіх наступних IMF цей фактор вже не позначається, і вони мають розподіл, близький до гаусового, а поруч розташовані однополярні імпульси сприймаються, як більш низькочастотні складові шуму. Статистика послідовної реконструкції шумового сигналу частково показана на рис. 7.33,б.

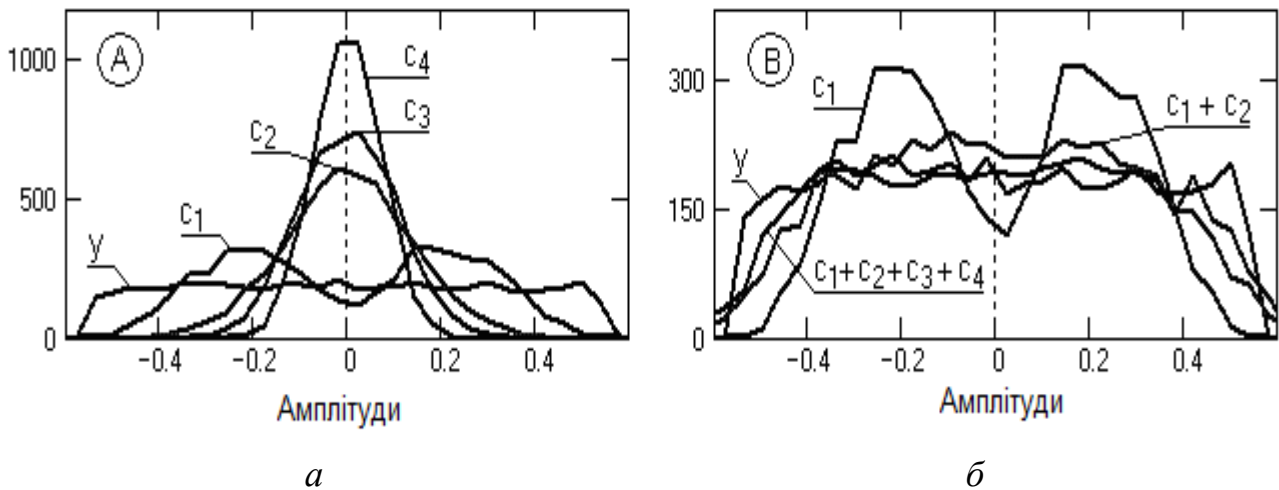


Рис. 7.33. Гістограми перших чотирьох IMF (а)

і статистика часткової реконструкції шумового сигналу (б)

Перевірка процесу EMD на шумових сигналах з іншими законами розподілу (Гауса, Пуассона та інших) показала, що якісний характер процесу залишається практично незмінним.

Частотна вибірковість EMD. З рис. 7.30 видно, що процес EMD має цілком певну частотну вибірковість на кожному рівні EMD. Але говорити про якісь частотних передавальних функцій EMD, мабуть, буде некоректним, оскільки будь-яка частотна складова ω_i початкового сигналу в процесі EMD може бути розщеплена за амплітудою і фазою на складові різних рівнів функцій IMF. Це можна бачити на рис. 7.34, де наведені графіки модулів «еквівалентних» частотних передавальних характеристик розкладання для перших п'яти функцій IMF, отримані усередненням відношення спектрів функцій до спектру початкового сигналу. Для отримання досить гладких «еквівалентних» передавальних функцій вхідний білий шум реалізувався

масивом з 30000 відліків, а згладжування відношення спектрів виконувалось у ковзному часовому вікні довжиною 2000 відліків.

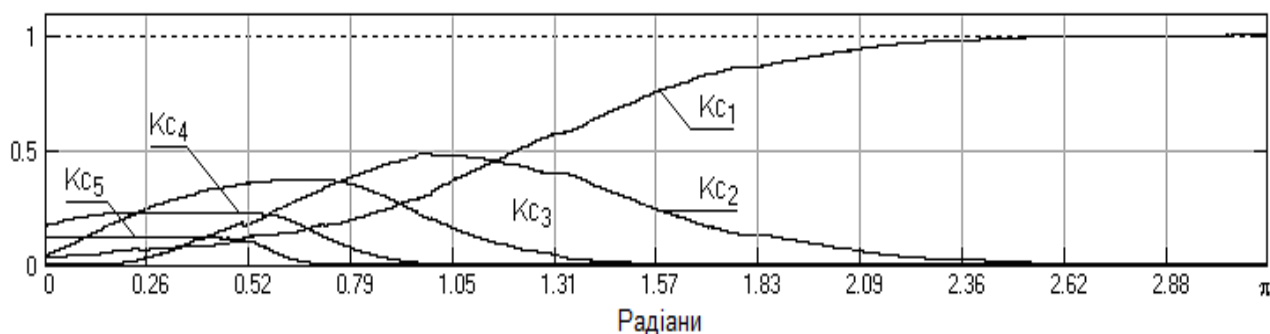


Рис. 7.34. Модулі «еквівалентних» частотних передавальних характеристик розкладання для перших п'яти функцій IMF

На рис. 7.35 наведені графіки послідовного підсумовування коефіцієнтів «еквівалентних» передавальних функцій, які показують процес послідовного перекриття усього частотного діапазону вхідного сигналу.

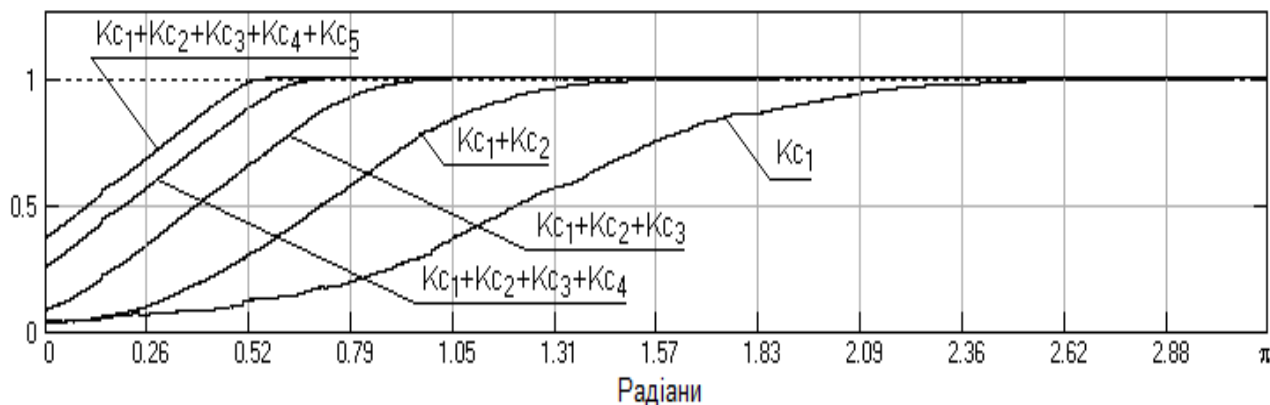


Рис 7.35. Процес послідовного перекриття усього частотного діапазону вхідного сигналу

Зворотним перетворенням Фур'є за спектрами потужності можуть бути обчислені нормовані автокореляційні функції сімейства IMF, перші 5 з яких наведені на рис. 7.36.

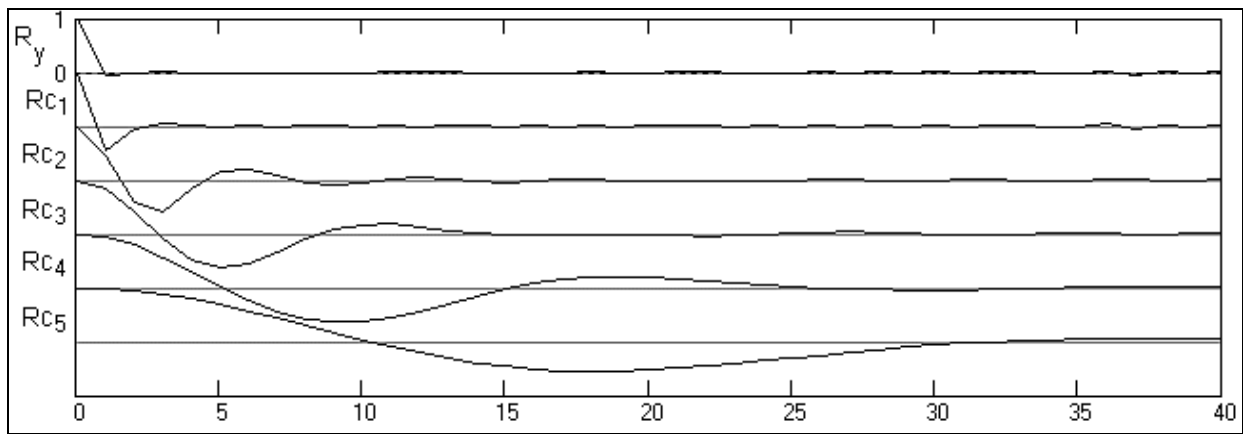


Рис 7.36. Нормовані автокореляційні функції сімейства IMF

Як випливає з графіків, статистична незалежність відліків в якійсь мірі зберігається тільки для першої IMF. Але навіть в ній з'являється негативна (знакозмінна) кореляція між послідовними відліками. У всіх інших функціях чітко простежується поява загасаючої косинусоїдальної залежності між відліками з послідовним збільшенням інтервалу кореляції зі збільшенням номеру IMF.

Контрольні запитання

1. Площина частота-час.
2. Обмежене в часі Фур'є-перетворення.
3. Принцип невизначеності перетворень частота-час і час-частота.
4. Базисні функції частотно-часового аналізу.
5. Властивості безперервного вейвлет-перетворення.
6. Дискретизація масштабу і зсуву. Фрейми.
7. Швидкий алгоритм дискретного вейвлет перетворення.
8. Узагальнений алгоритм частотно-часового вейвлет-аналізу.
9. Перетворення Гільберта-Гуанга.
10. Емпіричний метод декомпозиції.
11. Спектральний аналіз Гільберта.
12. Емпірична декомпозиція шумових сигналів.

ЧАСТИНА ІІІ. МОДЕЛЮВАННЯ СКЛАДНИХ ПРОЦЕСІВ В КЕРУВАННІ, КІБЕРБЕЗПЕЦІ ТА ТЕЛЕКОМУНІКАЦІЯХ

Глава 8. ІДЕНТИФІКАЦІЯ ТА ПРОГНОЗУВАННЯ СКЛАДНИХ НЕЛІНІЙНИХ ПРОЦЕСІВ ЯК ОБ'ЄКТІВ КЕРУВАННЯ

8.1. Адаптивне прогнозування сигналів і стану об'єктів керування

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 фільтра і корегування його параметрів за значенням вихідної похибки.

Структурні схеми використання АФА як прогнозуючий фільтр сигналів і апроксиматор (еталонна модель) ОК наведені на рис. 8.1.

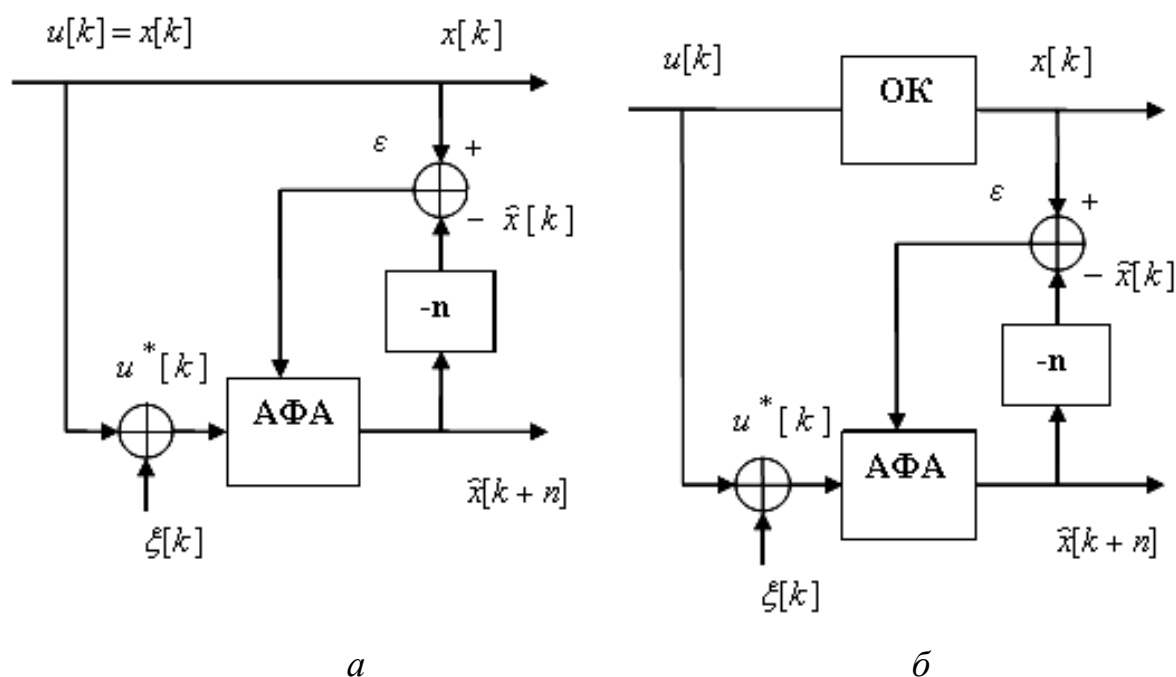


Рис. 8.1. Прогнозування сигналу (а) та ідентифікація ОК (б)
за допомогою АФА

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

$$u^*[k] = u[k] + \xi[k], \quad (8.1)$$

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.

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. 8.2. Тут ЛДЛ є лініями затримки, величини яких (глибина пам'яті) визначаються розмірністю вхідних $d_i - 1$ 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.

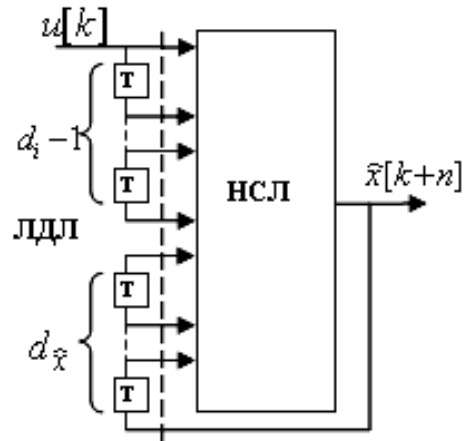


Рис. 8.2. Структура Вінера-Гаммерштейна прогнозуючої моделі нелінійного динамічного ОК

8.1.1. Лінійний адаптивний фільтр-апроксиматор.

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

$$x[k] = \sum_{m=-\infty}^{\infty} u[m] \cdot h[k - m] = \sum_{m=-\infty}^{\infty} h[m] \cdot u[k - m], \quad (8.2)$$

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:

$$\hat{x}[k + n] = \sum_{r=0}^R b_r \cdot u[k - r], \quad (8.3)$$

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

Параметрами настройки (навчання) цієї НМ 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:

$$\hat{x}[k+n] = \sum_{m=1}^M a_m \cdot x[k-m] + \sum_{r=0}^R b_r \cdot u[k-r], \quad (8.4)$$

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

Параметрами настройки (навчання) цієї НМ 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.

8.1.2. 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 розповсюдження з прихованим шаром представляється у вигляді рівняння згортки:

$$\hat{x}[k+n] = \sum_{\tau \in P} F_{\hat{x}} \left\{ \sum_{l \in Q} v_l[\tau] \cdot F_l \left(\sum_{m \in Q} v_{l,m}[\tau] \cdot u_m[k-\tau] \right) \right\}, \quad (8.5)$$

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:

$$\hat{x}[k+n] = \sum_{\tau \in P} F_{\hat{x}} \left\{ \sum_{l,m \in Q} v_l \cdot F_l(\vartheta_l, \|u_m[k-\tau] - v_l\|) \right\}, \quad (8.6)$$

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

Параметрами настройки НМ (8.6) є $\{v_l, \vartheta_l, v_l\} \subset a$.

Ідентифікація параметрів (настройка – навчання) НМ прямого розповсюдження (8.5) 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:

$$\bar{\varepsilon} = E \left\{ \|x[k+n] - \hat{x}[k+n]\| \right\} \rightarrow \min, \quad (8.7)$$

де E – математичне очікування.

При навчанні НМ з РБФ (8.6) спочатку визначаються центри і відхилення для радіальних елементів, після цього оптимізуються параметри лінійного вихідного шару: $\{v_l, \vartheta_l, v_l\} \subset a$.

8.1.3. 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 відрізьку $[0, 1]$.

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

$$\hat{x}[k+n] = \sum_{\tau \in P} \sum_{m \in Q} \beta_m[\tau] \cdot \alpha_m[k-\tau], \quad (8.8)$$

де $\beta_m[\tau] = U_m^{-1}(\alpha_m[\tau] / \sum_m \alpha_m[\tau])$; $\alpha_m[k-\tau] = T_n \{L_{l,m}(u_m[k-\tau])\}$; $U = U(a_U)$;

$L = L(a_L)$.

Тут U_m^{-1} – функція, зворотна функції належності проміжного виходу m мережі з параметрами a_U ; α_m – значення проміжного виходу; T_n – довільна t -норма моделювання логічної операції «I»; $L_{l,m}$ – функція належності нечіткого правила l входу m з параметрами a_L .

Параметрами настройки НМ (8.8) є $\{a_U, a_L\} \subset a$, її навчання виконується аналогічно НМ (8.6).

Перевагою алгоритмів параметричного навчання НМ є простота і швидкодія, а недоліком – їх локальність (висока вірогідність застрявання в локальному екстремуму).

8.1.4. 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 (див. п. 2.5.2):

1. Розрахунок потенціалу кожної точки (міри просторової близькості між нею та іншими із M):

$$E(x_m) = \frac{1}{M} \sum_{i=1}^M e^{-\frac{\|x_m - x_i\|}{(R_c/2)^2}}. \quad (8.9)$$

2. Встановлення кількості кластерів $k_c = 0$.

3. Визначення точки з найвищим потенціалом $E(x_p)$, x_p :

$$p = \arg \max_{i=1}^M E(x_i). \quad (8.10)$$

4. Ініціалізація центра j -го кластера:

$$k_{c_j} = x_p. \quad (8.11)$$

5. Зниження потенціалу всіх точок:

$$E(x_i) = E(x_i) - E(k_{c_j}) e^{-\frac{\|k_{c_j} - x_i\|}{(r/2)^2}}, \quad r = [1, 1.5] R_c. \quad (8.12)$$

6. Перевірка значення потенціалу відносно встановленого порогу thr :

$$\max_{i=1}^M E(x_i) < thr. \quad (8.13)$$

Якщо умова (8.13) виконується, то «кінець», інакше повинен бути перехід до пункту 3.

Таким чином, рівняння АФА Genfis2 можна представити у вигляді:

$$\hat{x}[k + n] = FF_j(FV(a_{FV}, \sum_{\tau \in P} \sum_{h=1}^{k_c} PV_h(D_{mh}))), \quad (8.14)$$

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 $\{\alpha, R_c\} \subset a_{FV}$.

8.1.5. 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 (див. п. 2.5.3):

1. Встановлення нечітких кластерів у матрицю розділів

$$M_D = [\mu_{\theta i}]; \mu_{\theta i} \in [0, 1]; \theta = \overline{1, \Theta}; i = \overline{1, k_c}. \quad (8.15)$$

де

$$\sum_{i=1}^{k_c} \mu_{\theta i} = 1; 0 < \sum_{\theta=1}^{\Theta} \mu_{\theta i} < \Theta. \quad (8.16)$$

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 (8.16).

4. Розрахунок центрів кластерів Ω_i :

$$\Omega_i = \frac{\sum_{\theta=1}^{\Theta} \mu_{\theta i}^{\varpi} * |X_{\theta}|}{\sum_{\theta=1}^{\Theta} \mu_{\theta i}^{\varpi}}; i = \overline{1, k_c}. \quad (8.17)$$

5. Розрахунок відстані між об'єктами від матриці спостереження X і центрами кластерів:

$$D_{\theta i} = \sqrt{\|X_{\theta} - \Omega_i\|^2}. \quad (8.18)$$

6. Перерахунок елементів матриці розділів:

- якщо $D_{\theta i} > 0$, то

$$\mu_{\theta i} = 1 / \left(D_{j\theta}^2 * \sum_{j=1}^{k_c} \frac{1}{D_{j\theta}^2} \right)^{1/(\varpi-1)}, \quad (8.19)$$

- якщо $D_{\theta i} = 0$, то

$$\mu_{\theta i} = \begin{cases} 1, j = i \\ 0, j \neq i \end{cases}; j = \overline{1, k_c}. \quad (8.20)$$

7. Перевірка умови:

$$\|M_D - M_D^*\| < \varepsilon. \quad (8.21)$$

Якщо «ні», то перейдіть до кроку 4, якщо "так", то «закінчіть».

Тоді рівняння АФА Genfis3 можна представити як рівняння згортки:

$$\hat{Y}[k+n] = FF_j(FD(a_{FV}, \sum_{\tau \in P} \sum_{h=1}^Q V_h(D_{mh}(z_n, y_m[k-\tau]))) \quad (8.22)$$

де $D_{mh} = \|V_h - y_m[k-\tau]\|$; $V_h = \frac{\sum_{m=1}^M (\mu_{mh})^\alpha \cdot y_m[k-\tau]}{\sum_{k=1}^M (\mu_{mh})^\alpha}$; $h = \overline{1, Q}$; $m = \overline{1, M}$; $F = \{\mu_{mh}\}$;

$$\mu_{mh} \in [0, 1]; \sum_{h=1}^Q \mu_{mh} = 1; 0 < \sum_{m=1}^M \mu_{mh} < M.$$

Налаштуванням цього АФА $\epsilon \{Q, M, L, \epsilon\} \subset a_{FV}$.

8.1.6. 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 фільтру є нейронний вейвлет (НВ) АФА (рис. 8.3):

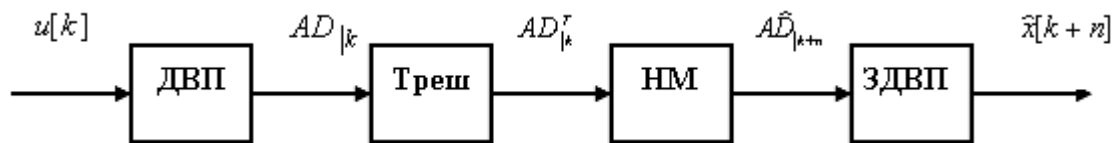


Рис. 8.3. Структура прогнозуючого нейронного вейвлет АФА

У ньому в основу фільтрації покладена процедура прямого дискретного вейвлет перетворення (ДВП). Для зменшення впливу шуму у фільтрі здійсню-

ється 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

$$AD|_k = \{A_L, D_L, D_{L-1}, \dots, D_1\} \quad (8.23)$$

з коефіцієнтами апроксимації по рівнях розкладання

$$A_L = \{a_{L,p}\} = \sum_{p=1}^{P_L} u[k] \cdot \varphi_{L,p}[k] \quad (8.24)$$

і коефіцієнтами деталізації

$$D_l = \{d_{l,p}\} = \sum_{l=1}^L \sum_{p=1}^{P_l} u[k] \psi_{l,p}[k]; \quad l = \overline{1, L}, \quad (8.25)$$

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 $[k + n]$:

$$A\hat{D}|_{k+n} = F_{NN}(AD|_k^T), \quad (8.26)$$

де F_{NN} – узагальнена функція перетворення НМ.

Для забезпечення асимптотичної стійкості АФА використовуються НМ без зворотних зв'язків і з нелінійними функціями активації (із-за нелінійності ОК і сигналів).

Тоді прогнозування сигналу $\hat{x}[k+n]$ глибиною n виконується за прогнозованими за допомогою НМ коефіцієнтами розкладання $AD|_{k+n} = \{\hat{a}_{L,p}, \hat{d}_{l,p}\}$ відповідно до виразу ЗДВП:

$$\hat{x}[k+n] = \sum_{p=1}^{P_L} \hat{a}_{L,p} \cdot \varphi_{L,p}[k+n] + \sum_{l=1}^L \sum_{p=1}^{P_l} \hat{d}_{l,p} \psi_{l,p}[k+n]. \quad (8.27)$$

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.

8.2. Ідентифікація нелінійних процесів за часовими реалізаціями

Ідентифікація процесу, як динамічної системи, полягає в одержанні чи уточненні за експериментальними даними математичної моделі цього процесу, висловленої за допомогою того чи іншого математичного апарату.

Нелінійний процес (динамічна система) зображується векторним рівнянням:

$$\dot{x} = \Phi(x, \lambda), \quad (8.28)$$

де Φ – нелінійна функція розмірності d ; x і λ – вектори координат і параметрів системи.

Процес у вигляді потоку (8.28) може бути також представлений дискретним відображенням Пуанкаре:

$$x[k+1] = \Phi\{x[k], \lambda\}; \quad x[k] = \{x_1[k], \dots, x_{d-1}[k]\}, \quad (8.29)$$

де k – такт часу $t = k \cdot T$; T – період дискретизації.

Рівняння виду (8.28) і (8.29) залежно від значень параметрів порядку λ мають чотири стійких рішення: стан рівноваги, коли після перехідного процесу система досягає стаціонарного стану; періодичне і квазіперіодичне рішення, а

також хаос. Цим типам рішень відповідають атрактори системи у вигляді стійкої рівноваги, граничного циклу, квазіперіодичного атрактора і хаотичного (дивного) атрактора.

Відмінною рисою останнього є його чутливість до початкових умов і дробова розмірність – фрактальність (властивість самоподоби на різних масштабах).

Зміна значень параметрів λ рівняння системи (8.28) і (8.29) викликає втрату стійкості одного стану (режиму функціонування) системи і перехід (біфуркацію) її в інший стан.

8.2.1. Характеристики нелінійних хаотичних процесів.

Найважливішою характеристикою хаотичного руху у фазовому просторі довільної розмірності є ентропія Колмогорова (сума старших показників Ляпунова), що описує динамічне поведіння на дивному атракторі.

Безладдя (хаос) є поняттям теорії інформації, тому ентропія Колмогорова K пропорційна швидкості втрати інформації про стан динамічної системи з часом і показує наскільки динамічна система хаотична:

$$K = - \lim_{T \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{NT} (K_{k+1} - K_k) = - \lim_{T \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{NT} \sum_{i_0 \dots i_k} P_{i_0 \dots i_k} \ln P_{i_0 \dots i_k}, \quad (8.30)$$

де $K_{k+1} - K_k$ – втрата інформації на інтервалі часу від k до $k+1$; $P_{i_0 \dots i_k}$ – спільна ймовірність перебування точки $x[0]$ в осередку фазового простору i_0 розміром ε , $x[1T]$ – в i_1 , ... і $x[kT]$ – в i_k ; N – тривалість часової реалізації.

K -ентропія дорівнює нулю для регулярного руху, нескінченна для випадкових систем, позитивна й обмежена для систем з детермінованим хаосом.

Колмогорівська ентропія дозволяє визначити середній час, на який можна передбачити стан системи з динамічним хаосом. Точне прогнозування стану цієї системи можливе тільки на інтервалі часу T_d , такому, що $\varepsilon \cdot e^{KT_d} = 1$, відкля

$$T_d = \frac{1}{K} \ln \left(\frac{1}{\varepsilon} \right). \quad (8.31)$$

Точність визначення початкового стану системи характеризується величиною дискретизації фазового простору ε і впливає на T_d логарифмічно.

Відстань між найближчими точками атрактора до і після біфуркацій знаходиться в універсальному відношенні. Самоподоба такого явища описується за допомогою фрактальної розмірності Хаусдорфа D_H , що характеризує швидкість росту числа осередків покриття $N(\varepsilon)$ даної множини при зменшенні розміру осередків ε :

$$D_H = - \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \varepsilon}, \quad (8.32)$$

де основа логарифма довільна.

8.2.2. Визначення характеристик стану нелінійних процесів.

Для рішення задачі ідентифікації необхідно визначити режим функціонування системи і її розмірність (порядок), а потім реконструювати модель системи (8.28) чи (8.29).

Експериментальний сигнал містить інформацію про режим роботи (атрактор) системи, що породжує. Якісними ознаками хаотичності руху системи є:

- нерегулярність часового сигналу;
- смугові складові на низьких частотах у його спектрі;
- швидкий спад автокореляційної функції сигналу;
- самоподібна структура його час-частотного (вейвлет) перетворення.

Для періодичних рухів спектр сигналу містить дискретні лінії, тоді як хаос через аперіодичність зображається смугою на низьких частотах.

Кореляційна функція для регулярних рухів постійна чи осцилює, а в хаотичному режимі експоненційно спадає.

Усі відомі дивні атрактори мають фрактальну розмірність (самоподібні), що легко видно за вейвлет перетворенням сигналу.

Доведено, що по одній часовій реалізації (що спостерігається) можна визначити наступні характеристики:

- фазовий портрет атрактора (режиму роботи);
- кореляційну розмірність D_C (нижню границю розмірності Хаусдорфа

$$(8.32) D_C < D_H);$$

- розмірність вкладення атрактора d (розмірність фазового простору) динамічної системи;

- кореляційну ентропію K_C (нижню границю ентропії Колмогорова (8.30) $K_C < K$), що характеризує наскільки хаотичний сигнал.

По отриманій часовій реалізації (сигналу $x = x(t)$, що спостерігається), задавши затримку τ і розмірність d фазового простору, будується його дискретне відображення:

$$x[k] = \{x[k], x[k-m], x[k-2m], \dots, x[k-(d-1)m]\}, \quad (8.33)$$

де m – ціле число ($m = \tau / T$).

При переборі по k виходить дискретний набір точок у d -мірному просторі, що при сталому режимі системи відповідно до *теорему Такенса* є фазовим портретом атрактора.

Чисельне визначення розмірності атрактора виконується за допомогою кореляційної розмірності:

$$D_C = \lim_{\varepsilon \rightarrow 0} \frac{\log(\sum_{i=0}^{N(\varepsilon)} p_i^2)}{\log \varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{\log C(\varepsilon)}{\log \varepsilon}, \quad (8.34)$$

яка обчислюється за кореляційним інтегралом:

$$C(\varepsilon) = \lim_{N \rightarrow \infty} N^{-2} \sum_{i,j} \chi[\varepsilon - \|x_i - x_j\|], \quad (8.35)$$

оцінюваному за відображенням (8.33). Величина D_C пов'язана з $C(\varepsilon)$ у такий

спосіб: $\sum_{i=0}^{N(\varepsilon)} p_i^2$ – імовірність того, що дві точки на аттракторі лежать усередині

осередку ε^D (імовірність того, що дві точки атрактора розділені відстанню, ме-

ншою ε); $\sum_{i,j} \chi[\varepsilon - \|x_i - x_j\|]$ – число пар i і j , для яких відстань $\|x_i - x_j\| < \varepsilon$; χ – східчаста функція Хевісайда; $x_i = x[iT]$.

Значення розмірності D_C визначається по нахилу лінійної ділянки залежності $\log C(\varepsilon)$ від $\log \varepsilon$. Крім того, будується залежність $D_C(d)$ зі збільшенням d у виразі (8.33) доти, поки D_C не досягне насичення. Розмірність d , починаючи з якої D_C перестає змінюватися, є мінімальна розмірність вкладення атрактора, тобто найменша ціла розмірність фазового простору, що містить весь атрactor.

Разом з тим, із теореми про вкладення випливає, що оцінка розмірності фазового простору d визначається через оцінку розмірності атрактора D_C реальної динамічної системи (формула Мане):

$$d \geq 2D_C + 1. \quad (8.36)$$

На практиці значення d за виразом (8.36) для відображень виявляється завищеним, тому часто обмежуються простором розмірності $d \geq D_C$.

Оцінку (знизу) колмогорівської ентропії K можна одержати з урахуванням (8.30) і (8.35) у виді:

$$K_C = - \lim_{\varepsilon \rightarrow 0} \lim_{k \rightarrow \infty} \ln[C_k(\varepsilon) / C_{k+1}(\varepsilon)] \leq K, \quad (8.37)$$

де узагальнений кореляційний інтеграл $C_k(\varepsilon)$ обчислюється по відображенню (8.33) відповідно до виразу:

$$C_k(\varepsilon) = \lim_{N \rightarrow \infty} N^{-2} \sum_{i,j} \chi[\varepsilon - \|x_i - x_j\|_k] = \sum_{i_1 \dots i_k} P_{i_1 \dots i_k}^2. \quad (8.38)$$

Тут $\|x_i - x_j\|_k = \sqrt{\sum_{n=0}^{k-1} (x_{i+n} - x_{j+n})^2}$.

Значення $K_C = \text{const} > 0$ є достатньою умовою існування хаотичного режиму.

Оцінка (зверху) інтервалу передбачуваності виконується як:

$$T_C = \frac{1}{K_C} \ln\left(\frac{1}{\varepsilon}\right) \geq T_d. \quad (8.39)$$

За часи, більші T_C , можливе тільки статистичне прогнозування, інтервал (глибина) якого залежить від кореляційної функції процесу.

Реконструкція моделі динамічної системи на основі аналізу часових реалізацій полягає у виборі базисних функцій (структури) і їх коефіцієнтів (параметрів) моделі, а також визначенні значень параметрів моделі, оптимальним чином відповідних часовій реалізації.

Для рішення задачі реконструкції моделі (8.28), (8.29) формується d -мірне відображення виду:

$$\begin{aligned} x_{1,i+1} &= \Phi_1 \{x_{1,i}, x_{2,i}, \dots, x_{d,i} | \lambda\}; \\ &\dots\dots\dots \\ x_{d,i+1} &= \Phi_d \{x_{1,i}, x_{2,i}, \dots, x_{d,i} | \lambda\}, \end{aligned} \quad (8.40)$$

де $x_{j,i}$ – координати вектора стану в моменти часу $i \cdot T$; $j = \overline{1, d}$; $i = \overline{d, N-1}$.

Далі еволюційні функції $\Phi_j \{x_i\}$ відображаються у вигляді розкладання за деяким базисом із шуканими коефіцієнтами (параметрами) λ .

Традиційно для апроксимації функцій використовуються поліноми Лежандра, Колмогорова-Габора чи інші. Коефіцієнти цих поліномів утворюють невідомі параметри λ , значення яких обираються так, щоб якнайкраще відповідати часовим реалізаціям, що спостерігаються, наприклад, за критерієм мінімуму похибки:

$$E_x^2 = \sum_{j=1}^d \sum_{i=d}^{N-1} [x_{j,i+1} - \Phi_j \{x_i\}]^2 \rightarrow \min. \quad (8.41)$$

Для розв'язання системи (8.40) за критерієм (8.41) звичайно застосовується метод найменших квадратів (за умови, що N більше кількості параметрів λ).

Перевагою традиційного підходу є його простота (пошук тільки однієї функції Φ), а істотними недоліками – громіздкість одержуваної моделі, а також

наявність процедури чисельного диференціювання, що призводить до великих похибок обчислень.

Більш продуктивним для апроксимації функцій $\Phi_j\{x_i\}$ є використання НМ чи гібридних НМ з нечіткою логікою, що є універсальними й ефективними апроксиматорами. Їх параметрами λ є ваги нейронів, а також коефіцієнти їх функцій активації і належності (для гібридних мереж). Навчання мереж здійснюється, наприклад, методом зворотного поширення похибки і полягає у визначенні значень параметрів λ , що оптимальним чином (за критерієм (8.41)) відповідають спостережуваним часовим реалізаціям.

Реконструйована модель, що правильно передає динаміку на аттракторі може бути неадекватною до опису траєкторії наближення до аттрактора і виходу в сталий режим. Для вирішення таких проблем необхідно додатково обробляти реалізації, що характеризують процеси встановлення. Це може бути здійснене шляхом реалізації адаптивної ідентифікації ОК у процесі функціонування системи керування, що відносно просто виконується за допомогою нейронних і гібридних мереж, але дуже трудомістко при традиційному підході.

8.3. Приклади ідентифікації породжувачих процесів та прогнозування їх стану

8.3.1. Ідентифікація процесу дроблення руди та прогнозування її кру- пності.

Розглянемо ідентифікації процесів, що породжують часову реалізацію відображення Ено та експериментальний сигнал вмісту класу +100мм у дробленій руді γ_{+100} .

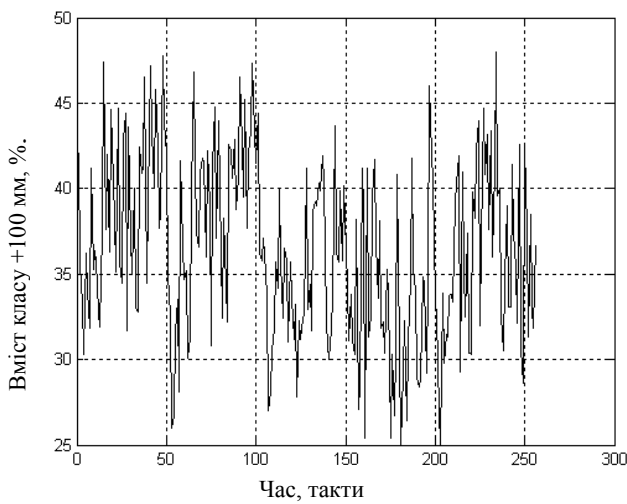
Нелінійне відображення Ено (Henon) описує поведінку ротора зі збудженням:

$$x_1[k+1]=1-ax_1^2[k]-bx_2[k]; \quad x_2[k+1]=x_1[k], \quad (8.42)$$

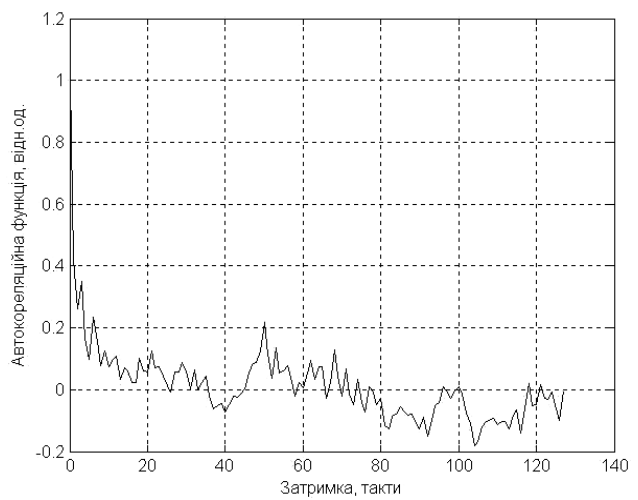
де a, b – параметри порядку $\lambda = \{a, b\}$; x_1, x_2 – координати; k – такт часу.

Параметр a характеризує просторовий розподіл взаємодії, а параметр b – загасання в системі. Він обернено пропорційний коефіцієнту тертя (реальними є значення $b > 0$). В залежності від значень параметрів порядку система Ено знаходиться в різних режимах: від рівноваги до хаосу.

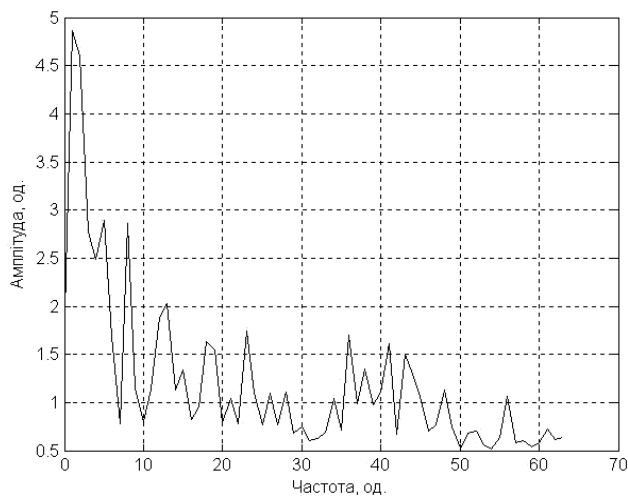
Часові і частотні характеристики експериментального сигналу γ_{+100} наведені на рис. 8.4.



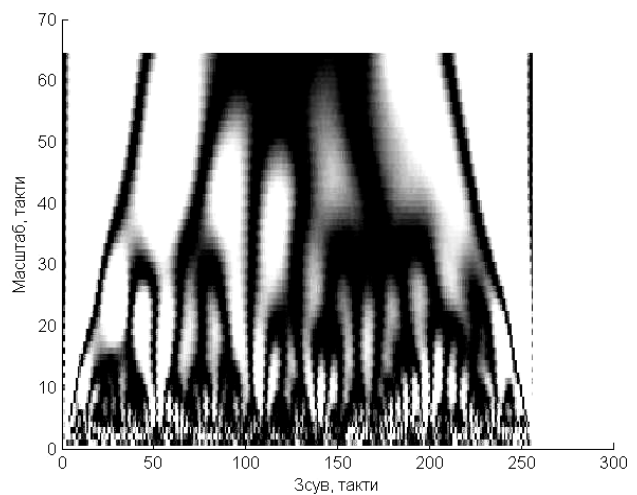
a



б



в



г

Рис. 8.4. Часова реалізація сигналу γ_{+100} (*a*), його кореляційна функція (*б*), спектральна щільність (*в*) і вейвлет перетворення (*г*)

За їх видом можна сказати про нерегулярність процесу, що його породжує. Це впливає з нерегулярного вигляду часової реалізації сигналу γ_{+100} (див. рис. 8.4,*a*), експоненційного спаду його кореляційної функції (рис. 8.4,*б*), наявності значної частини енергії спектра сигналу в низькочастотній області (рис. 8.4,*в*) і самоподібного (фрактального) характеру його вейвлет перетворення (гілляста структура на рис. 8.4,*г* зберігається на різних масштабах).

На рис. 8.5 наведені фазові портрети сигналів процесів, що ідентифікують. Видно, що внесення в систему білого шуму (з амплітудою 0,1 від рівня сигналу) порушує тонку структуру фазового портрета відображення Ено (див. рис. 8.5,*a* і 8.5,*б*).

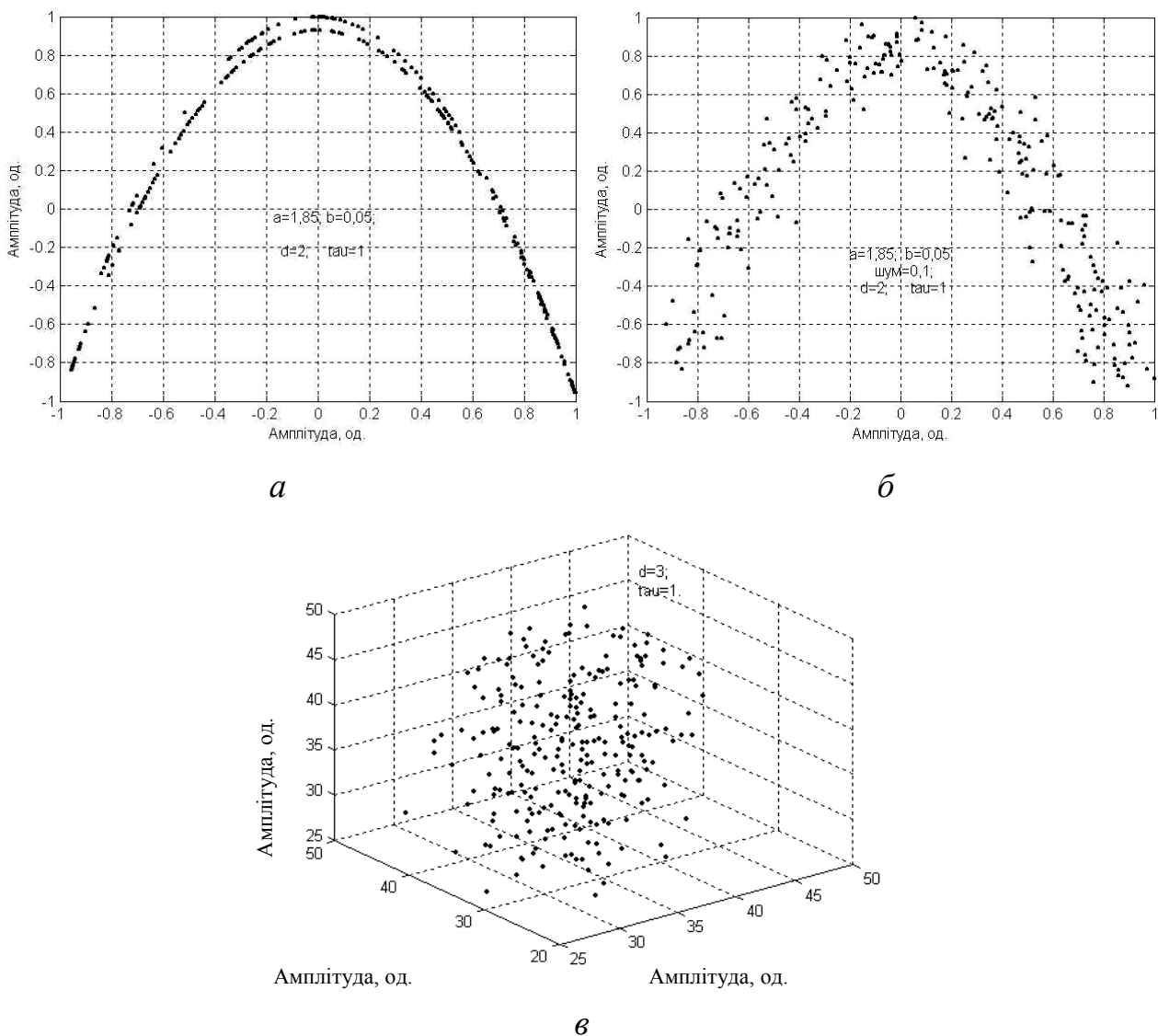


Рис. 8.5. Фазові портрети відображення Ено (*a*), відображення Ено із шумом (*б*) і сигналу γ_{+100} (*в*)

Побудова фазових портретів сигналу γ_{+100} при розмірності $d \leq 3$ і варіації затримки τ не дозволила знайти закономірність руху (див рис. 8.5,в). Це може бути викликано або великим рівнем шуму в сигналі, або перебуванням процесу, що породжує, у несталому режимі, або розмірність фазового простору процесу більше 3.

Розрахунки визначили значення кореляційних ентропії і розмірності атракторів відображення Ено $K_{CE} = 0,47$ і $D_{CE} = 1,14$ (при значеннях параметрів $a = 1,85$ і $b = 0,05$), а також сигналу γ_{+100} – $K_{C\gamma} = 0,39$ і $D_{C\gamma} = 3,04$.

При цьому інтервали точної передбачуваності складають $T_{CE} = 3,42$ такту і $T_{C\gamma} = 4,11$ такту (згідно (8.33) тривалість такту дорівнює τ , чи в дискретному часі – τ/T).

Для визначення розмірності фазового простору d_γ сигналу γ_{+100} за виразом (8.36) обчислювалася її оцінка зверху: $d_\gamma \leq 7,1$, а для оцінки значення d_γ знизу будувалась залежність $D_{C\gamma}(d_\gamma)$, що наведена на рис. 8.6.

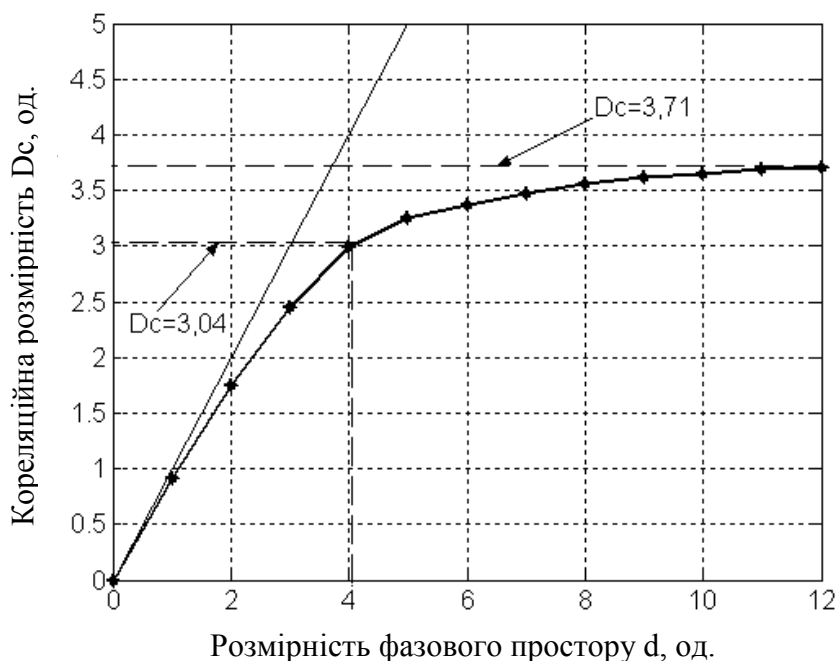


Рис. 8.6. Залежність розмірності атрактора від розмірності простору сигналу γ_{+100}

З неї випливає, що розмірність атрактора $D_{C\gamma}$ практично перестає зростати (входить у насичення) при розмірності фазового простору $d_\gamma \geq 4$.

З урахуванням розрахунків і графіка на рис. 8.6 одержимо $D_{C\gamma} \cong 3,37$ і $4 \leq d_\gamma \leq 7$, що багато більше 3 і, відповідно, пояснює відсутність закономірності в русі на фазовому портреті сигналу γ_{+100} (див. рис. 8.5,в).

Для реконструкції моделі відображення Ено й експериментального сигналу γ_{+100} використовували адаптивну нейронну систему нечіткого висновку Anfis з дзвіноподібною функцією належності. На її вхід подавалися відповідні часові реалізації зі своїми розмірностями d (глибиною пам'яті $d - 1$). Реалізації розбивалися на навчальну і перевіірочну послідовності нарівно, а прогноз виконувався глибиною до 10 тактів, що перевищує отримані вище значення інтервалів точної передбачуваності T_{CE} і $T_{C\gamma}$.

Як показник ефективності прогнозування використовувалася відносна величина середньоквадратичної похибки (8.41), нормованої за діапазонами зміни сигналів. Результати розрахунків цієї похибки для прогнозування відображення Ено, відображення Ено з білим шумом рівня 0,1, а також експериментального сигналу γ_{+100} наведені на рис. 8.7.

Для сигналу γ_{+100} розмірність приймалась $d_\gamma = 5$. З аналізу рис. 8.7 виходить, що відносна похибка прогнозування відображення Ено та сигналу γ_{+100} складає менше 0,10 при глибині прогнозу до 10 тактів та менше 0,03 для інтервалів точної передбачуваності цих сигналів. Для порівняння, відносна похибка прогнозування цих сигналів лінійними фільтрами (при лінійній апроксимації моделей породжуючих процесів) складає більше 0,20.

8.3.2. Прогнозування вмісту кремнію в чавуні при його виплавці.

До характерних особливостей доменного виробництва відносяться:

- випадковий характер змін в часі фізичних і хімічних властивостей шихтових матеріалів;

- велике число чинників (зокрема неконтрольованих), що впливають на кінцевий результат доменної плавки.

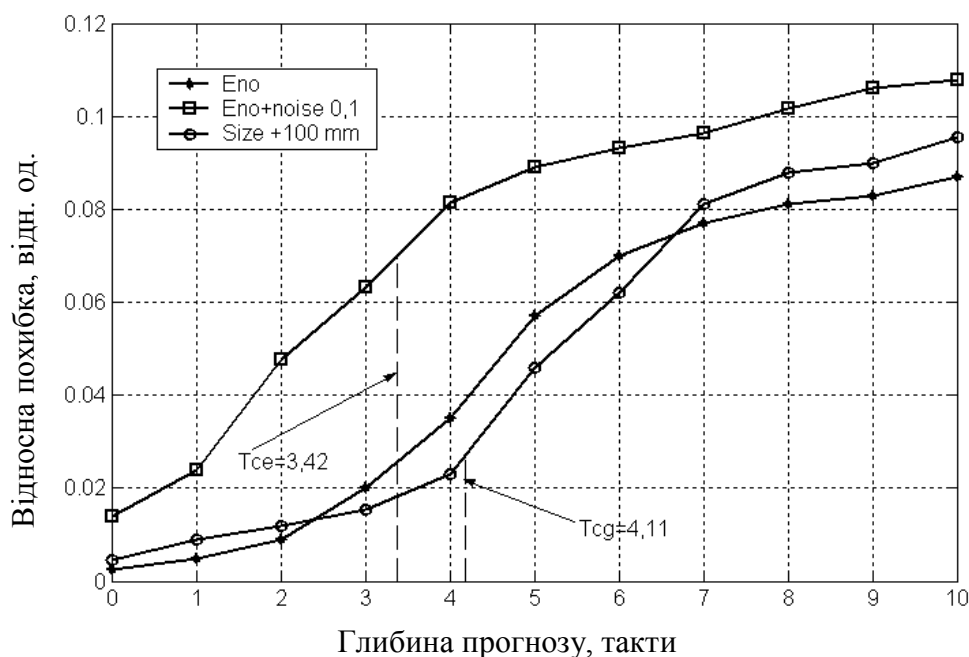


Рис. 8.7. Залежність відносної похибки від глибини прогнозу

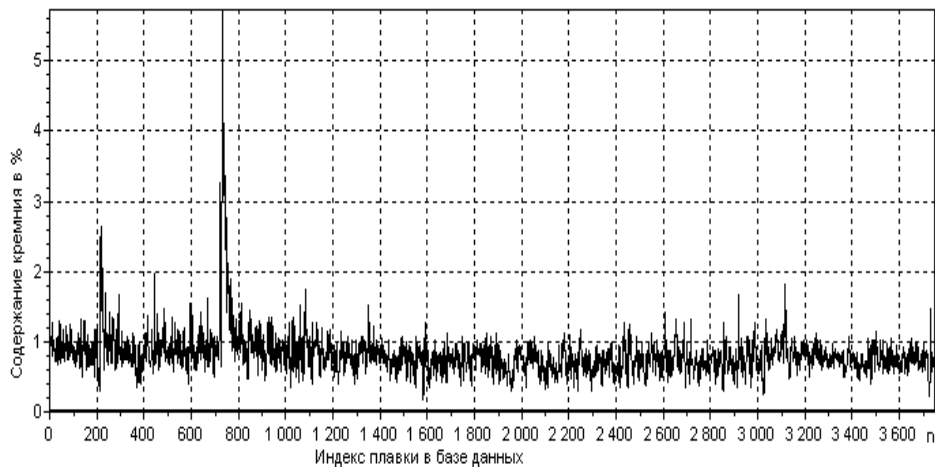
Відмічені особливості обумовлюють необхідність проведення досліджень властивостей часових рядів, якими представлені результати хімічного аналізу чавуну на випуску. Такі дослідження необхідні для розробки рекомендацій по створенню методик прогнозування хімічного складу чавуну в умовах діючого виробництва, адекватних характеру прогнозованого процесу.

Аналіз часових рядів на самоподібність виконувався на основі даних про вміст кремнію в чавуні (рис. 8.8), отриманих в різні часові періоди на доменній печі №3 (ДП-3) Маріупольського металургійного комбінату ім. Ілліча.

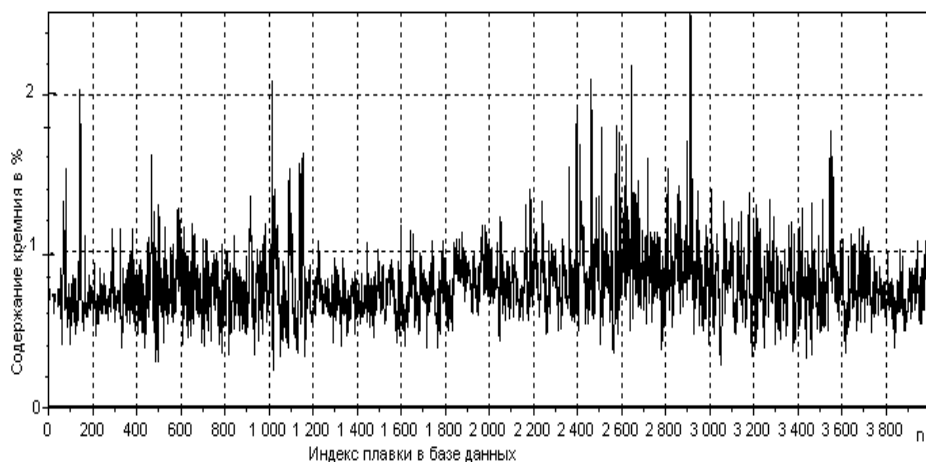
Аналіз фрактальних властивостей часових рядів, зокрема властивостей самоподобія, виконувався із застосуванням програми Fractan 4.4.

Однією з основних властивостей фрактального (самоподібного) процесу є повільне убавання АКФ (рис. 8.9), що характеризує тривалість пам'яті процесу.

З рис. 8.9 видно, що АКФ мають поволі спадаючий характер, а в їх «хвостах» відсутня тенденція прагнення до нуля, що характерно для самоподібних процесів.



a



б

Рис. 8.8. Вміст кремнію в чавуні на випуску:
часовий ряд 1 (за 2011 рік) (*a*) і ряд 2 (за 2012 рік) (*б*)

Показник Херста H інтерпретується таким чином:

- $H = 0.5$ відповідає часовому ряду у вигляді білого шуму (незалежний, випадковий процес);

- $0 \leq H < 0.5$ означає рожевий шум або антиперсистентність (часовий ряд міняє напрям частіше, ніж ряд випадкових незалежних величин);
- $0.5 < H \leq 1$ означає чорний шум або персистентність (часовий ряд характеризується ефектом довготривалої пам'яті і має схильність слідувати трендам).

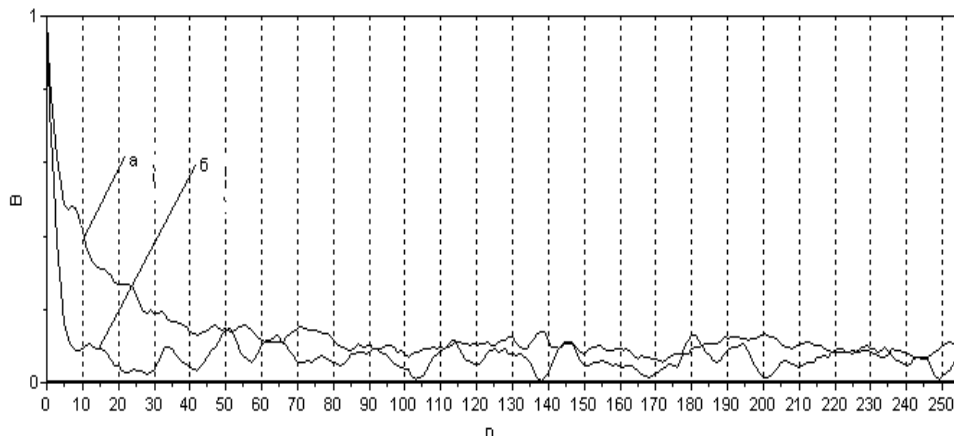


Рис. 8.9. АКФ досліджуваних часових рядів 1 (а) і 2 (б)

Трендостійкість поведінки процесу збільшується при наближенні H до одиниці.

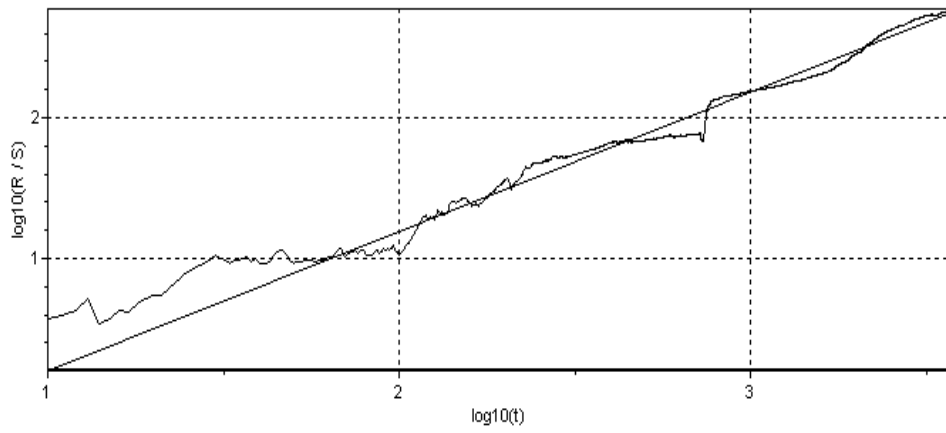
Відзначимо, що показник Херста пов'язаний з топологічною фрактальною розмірністю співвідношенням $D = 2 - H$.

Розрахункові значення показника Херста склали $H_1 = 0,9866 \pm 0,1329$ і $H_2 = 0,8963 \pm 0,2916$, що також підтверджує самоподібний характер процесів. Оцінка показника H виконувалася методом R/S -аналіза (рис. 8.10).

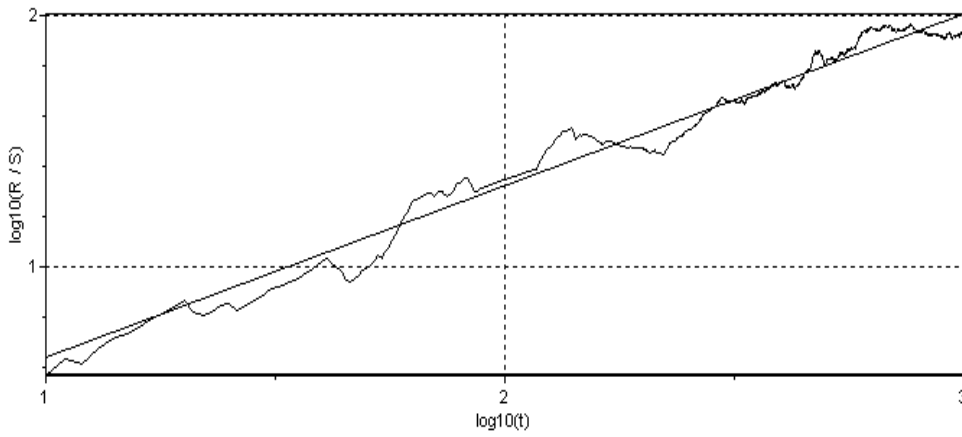
Таким чином встановлено, що досліджувані часові ряди носять фрактальний характер і мають властивість самоподоби. Отже, подальше дослідження необхідно проводити не з використанням класичних стохастичних методів, а з використанням фрактальних методів і методів стохастичної динаміки, які адекватні характеру досліджуваних процесів.

Вид дивних атракторів (двомірний фазовий простір) для досліджуваних часових рядів в нормованих координатах наведений на рис. 8.11. Тут видно області тяжіння, що є щільним «ядром». В той же час для випадкової послідовно-

сті, як вже наголошувалося вище, точки відновленого псевдоатрактора утворюють безструктурну хмару в фазовому просторі.

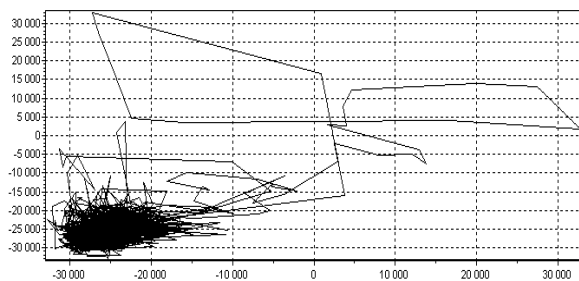


a

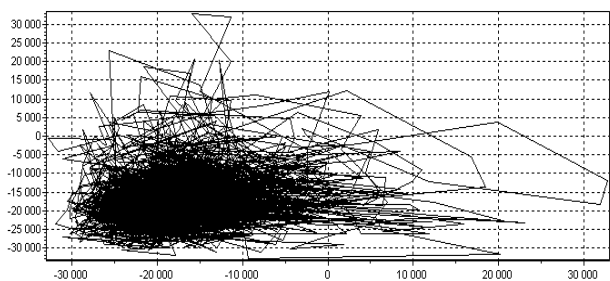


б

Рис. 8.10. Показник Херста для рядів 1 (*a*) і 2 (*б*)



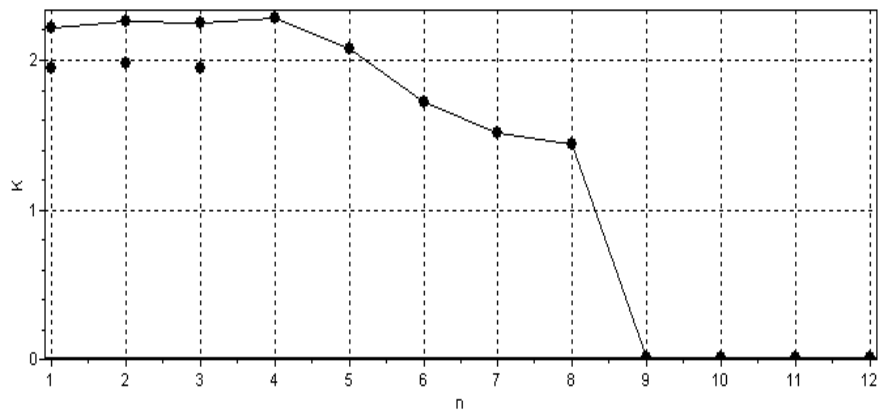
a



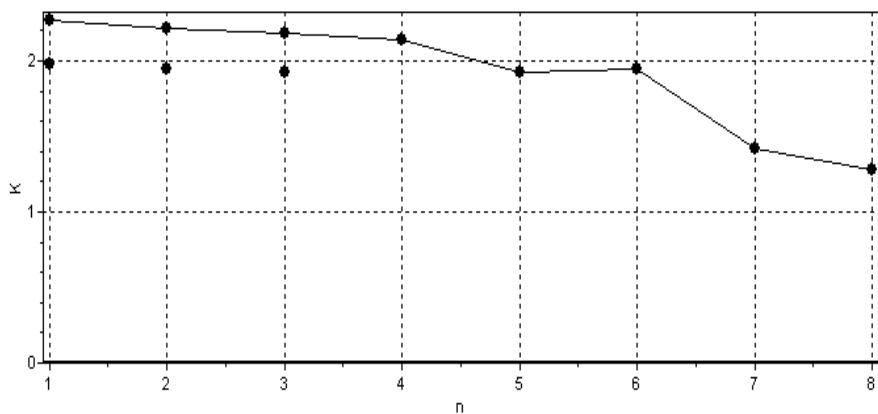
б

Рис.8.11. Фазові портрети дивних атракторів для рядів 1 (*a*) і 2 (*б*)

На рис. 8.12 наведені залежності кореляційних ентропій від розмірності простору вкладення n досліджуваних часових рядів. Кореляційні ентропії з моментів $n=4$ і $n=6$ не зростають, що свідчить про присутність хаотичної складової. Величина K в обох випадках достатньо мала ($K_1 = 2,221$, $K_2 = 1,276$).



a



б

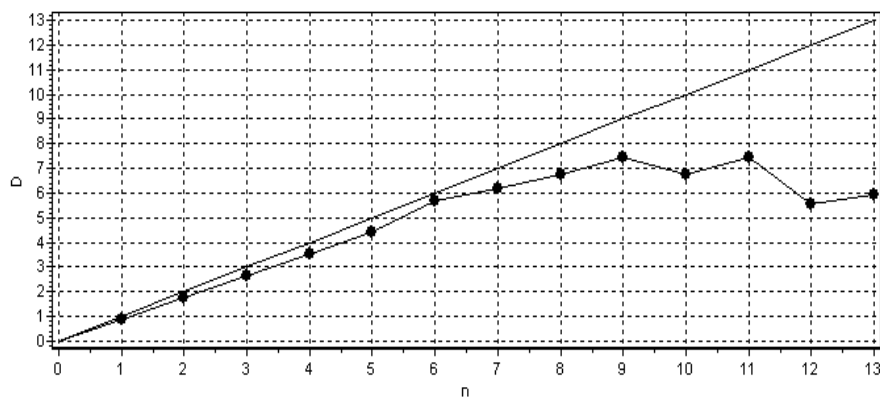
Рис. 8.12. Залежність кореляційної ентропії від розмірності вкладення для рядів 1 (*a*) і 2 (*б*)

Кореляційні розмірності часових рядів склали: $D_1 = 7,467$ і $D_2 = 6,353$ (рис 8.13).

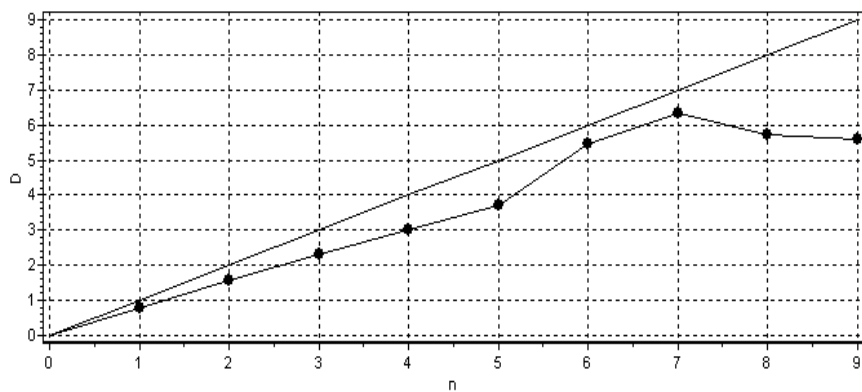
Прогнозування вмісту хімічного складу чавуну на виході (що є одним з основних показників плавки) може значно підвищити якість контролю процесу доменної печі.

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 .



a



б

Рис. 8.13. Залежності кореляційних розмірностей від розмірності вкладення для рядів 1 (*a*) і 2 (*б*)

Початкова послідовність 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):

$$C = \frac{\|Y_B^*[k+n] - \hat{Y}_B[k+n]\|}{\|Y_B^*[k+n]\|}. \quad (8.44)$$

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:

$$C_{pr} = \frac{\|Y^*[k] - \hat{Y}[k]\|}{\|Y^*[k]\|}. \quad (8.45)$$

Результати прогнозування вмісту кремнію в чавуні на виході наведені на рис. 8.14, де позначено: TREND – часовий сигнал хімічного складу чавуну на виході, PREDICT – його прогноз, URR – показник режиму прогнозу.

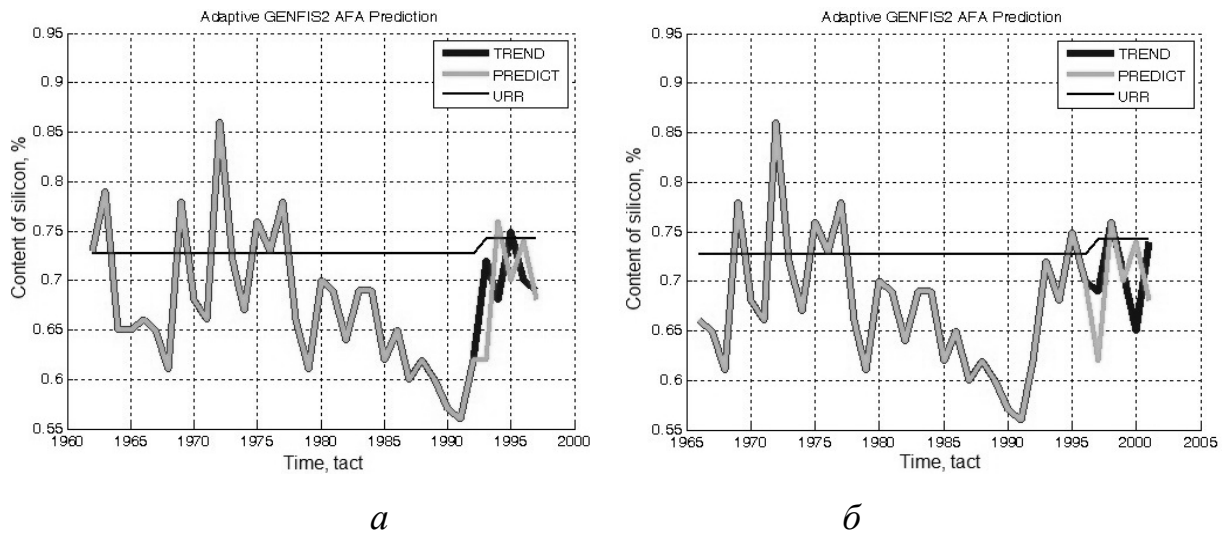


Рис. 8.14. Прогноз вмісту кремнію протягом 1 (а) та 5 (б) циклів наперед

Відносні похибки прогнозу склали: Anfis – 0,0653; Genfis2 – 0,0563; Genfis3 – 0,0625. Встановлено, що мінімуму критерію (8.44) відповідає АФА Genfis2 з впливом центру кластера $R_c = 0,4$. При цьому забезпечується і най-

менше значення відносної похибки прогнозу (8.41), що для Genfis2 дорівнює 0,0563.

Статистична перевірка за критерієм знаків показала значимість отриманих результатів.

8.3.3. Прогнозування теплового стану доменної печі та характеристик якості чавуну.

Експериментальні дослідження проводилися в умовах Маріупольського металургійного комбінату ім. Ілліча.

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 (PP), що характеризує тепловий стан доменної печі. Для визначення динамічних властивостей сигналу PP по його миттєвих значеннях був отриманий тренд PP, згладжений за допомогою вейвлет фільтру по середньому значенню часу плавки \bar{t}_{pl} , а також обчислена АКФ тренда PP і його спектральна щільність (рис. 8.15).

Визначимо характеристики стану процесу, що породжує тренд PP доменної печі. По вигляду часо-частотних характеристик тренда PP можна сказати про нерегулярність процесу, який його породжує. Це витікає з нерегулярного виду часової реалізації сигналу тренда (див. рис. 8.15,*a*), авторегресійного виду

його АКФ (рис. 8.15,б) і наявності значної частини енергії спектру сигналу в низькочастотній області (див. рис. 8.15,в).

Розрахунки виконані за допомогою стандартних програм дозволили визначити для тренда РР його 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).

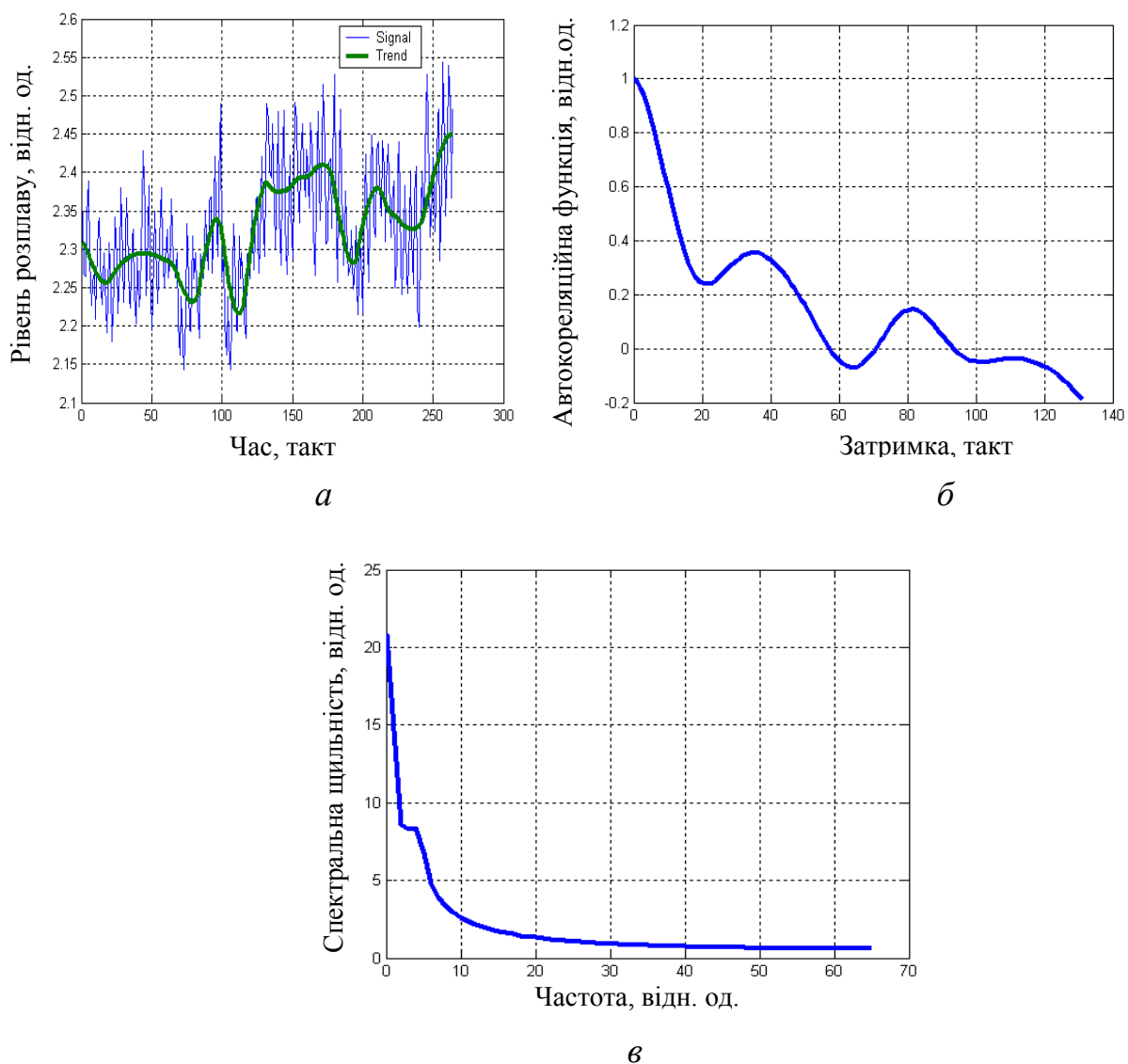


Рис. 8.15. Реалізація РР (Signal) і його тренда (Trend) (а), автокореляційна функція (б) і спектральна щільність (в) тренда

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), нейро-нечіткий (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 перевіро-чну послідовності нарівно, а як критерій оптимізації використовувалася віднос-

на середньоквадратична похибка між реальними і прогнозованими значеннями тренда РР в перевірочній 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. 8.16.

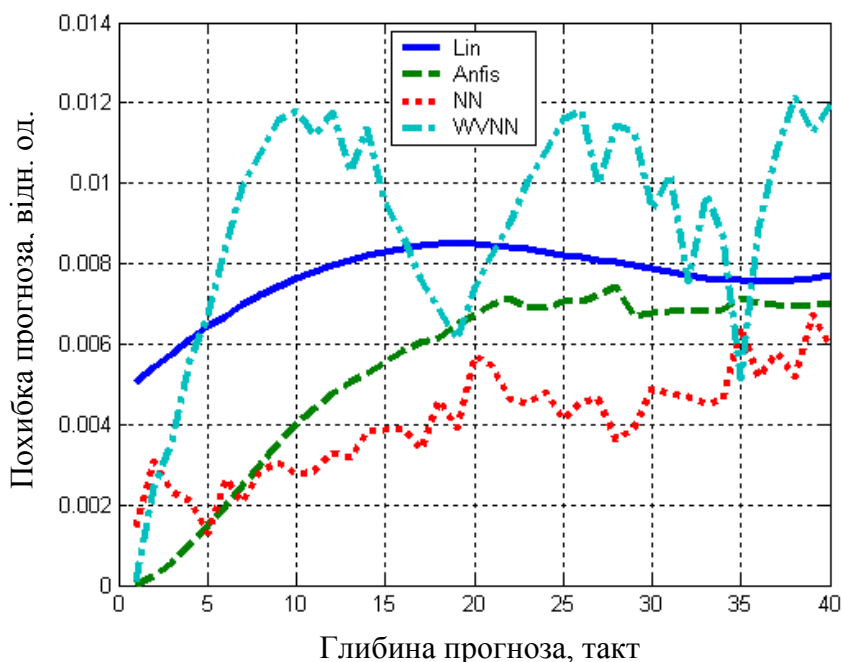


Рис. 8.16. Похибки прогнозу тренда УР для різних АФА: лінійного (Lin), нейронечіткого (Anfis), нейромережевого (NN) і нейровейвлетного (WVNN)

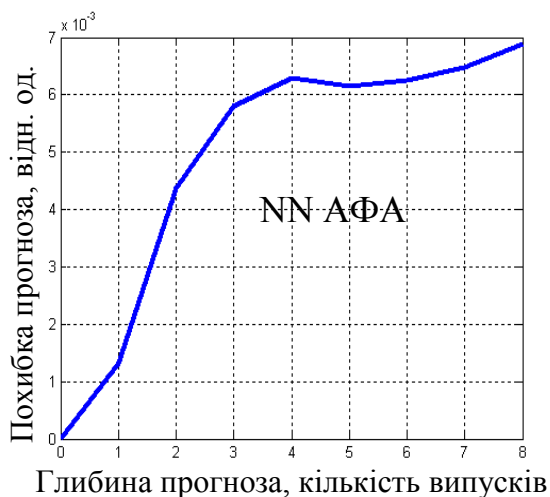
Аналіз показує, що похибка прогнозу не перевищує 0,012. При цьому, найменші похибки має NN АФА.

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. 8.17 наведені графіки похибок тих АФА, які дали кращі результати.

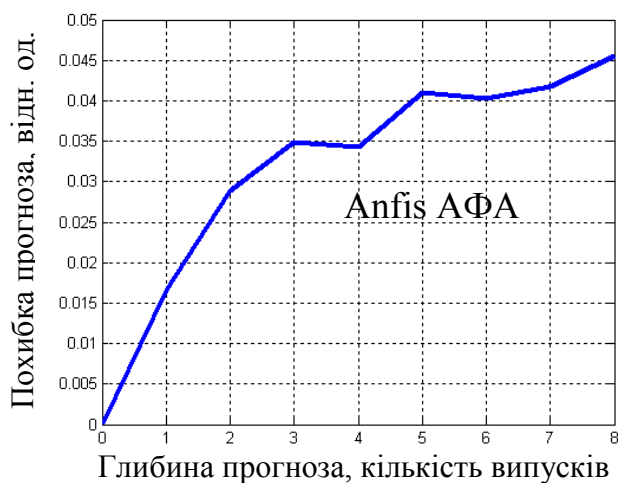
При моделюванні всі показники були занормовані в межі ± 1 , а їх середні значення склали для: температури чавуну – $1481 \text{ }^\circ\text{C}$, вмісту кремнію – $0,755 \%$ і вмісту сірки – $0,022 \%$.

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.

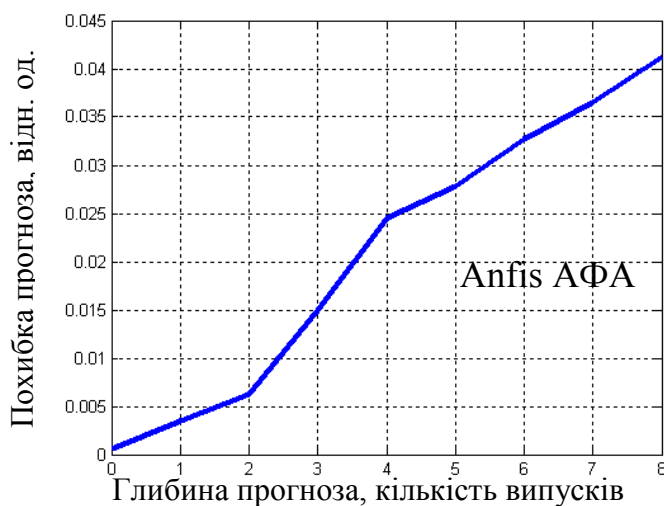
Статистична перевірка за непараметричним критерієм знаків підтвердила, що результати прогнозу з рівнем значущості 0,01 адекватні експериментальним реалізаціям.



a



б



в

Рис. 8.17. Похибки прогнозу температури чавуну (*a*) та вмісту кремнія (*б*) і сірки (*в*) в ньому

8.4. Структурно-параметрична ідентифікація об'єктів керування

8.4.1. Постановка задачі.

Сформулюємо задачу ідентифікації ОК таким чином: на підставі експериментальної множини функцій (часових рядів) збурень, керувань і виходів в умовах завод визначити структуру (узагальнену функцію Φ) і вектор параметрів a моделі виду:

$$\widehat{Y}[k+n] = \Phi\{Y[k], u[k], w[k], \xi[k], a[k], k\}, \quad (8.46)$$

що достатньо точно (у сенсі деякого критерію) апроксимують ОК відносно вхідних і вихідних величин у всьому функціональному просторі. Тут $Y[k], u[k], w[k], \xi[k]$ – відповідно, вектори (матриці) виходу процесу, його керувань, збурень і шуму до поточного часу k з відповідними глибинами пам'яті; n – глибина прогнозу (для компенсації чистого запізнювання і часу на синтез і реалізацію керування).

Вважаємо, що оцінка $\widehat{Z}[k]$ стану ОК (8.46) виконується за допомогою відповідних фільтрів спостереження:

$$\widehat{Z}[k] = \Phi_{\widehat{Z}}\{Z[k-n_Z+1], \alpha(Z[k-n_Z+1]), \xi_{\widehat{Z}}[k], a_{\widehat{Z}}[k], k\}, \quad (8.47)$$

де $\{Y[k], u[k], w[k]\} \subset \widehat{Z}[k]$ і $\widehat{Z}[k]$ – вектор (матриця) оцінки стану каналу системи; $\Phi_{\widehat{Z}}$ – узагальнена функція перетворення (метод, алгоритм); $Z[k-n_Z+1], \alpha(Z[k-n_Z+1]), \xi_{\widehat{Z}}[k], a_{\widehat{Z}}[k]$ – відповідно, вектор (матриця) відомих (вимірних) значень передісторії стану каналу (до моменту часу $k-n_Z+1$); лінійно незалежні функції, що характеризують властивості часового ряду (наприклад, автокореляційна функція, дисперсія тощо); шуми вимірювань та параметри прогнозуючого фільтра.

Як міру точності ідентифікації використаємо, наприклад, критерій мінімуму похибки між експериментальними $Y^*[k+n]$ і модельними значеннями виходу (8.46):

$$J = E \left\{ \left\| Y^* [k+n] - \hat{Y}[k+n] \right\|^2 \right\} \rightarrow \min \quad (8.48)$$

при дотриманні обмежень на функціональний простір. Тут E – математичне очікування.

Таким чином, формування вектора $I_s = \{\Phi, a\}$ оцінки структури Φ (структурна ідентифікація) і параметрів a (параметрична ідентифікація) моделі ОК (8.46) здійснюється на основі векторів сигналів спостереження $\hat{Z}[k]$ (8.47) шляхом мінімізації прийнятого функціонала:

$$J[I_s] \rightarrow \min_{I_s \in S} J \Rightarrow I_s^{opt} = \{\Phi_{opt}, a_{opt}\}, \quad (8.49)$$

де обмеження S , в загальному випадку, рівні:

$$S = \begin{cases} \{h(I_s) \geq 0\} \subseteq S_h; \\ \{g(I_s) = 0\} \subseteq S_g; \\ \{\phi_i, i = \overline{1, D}\} \subseteq S_D. \end{cases} \quad (8.50)$$

Тут h, g – неперервні функції; ϕ_i – елементи дискретного вектора D можливих значень структурної функції $\Phi = \{\phi_i\}$.

При розв'язанні задачі (8.49)-(8.50) мають бути визначені (обрані):

- ефективні методи оптимізації;
- зміст критерію (функціонала) J ;
- засоби врахування обмежень S ;
- тип структури моделі;
- базисні функції.

8.4.2. Структурна оптимізація.

Вирази (8.49)-(8.50) є комбінацією безперервної задачі математичного програмування і задачі дискретного програмування. Остання при малих D вирішується шляхом повного перебору, а при великих D – або структурними методами, або зведенням дискретної задачі до безперервної і подальшим її вирішенням методами параметричного пошуку.

Крім того, задача (8.49)-(8.50), зважаючи на нелінійність ОК і довільний вигляд функціонала, є багатоекстремальною (полімодальною), що вимагає використання методів глобальної оптимізації.

Найбільш ефективними при глобальній оптимізації нелінійних систем є пошукові методи, в яких алгоритм пошуку оптимального рішення зв'язує наступні один за одним рішення: $I_s(K+1) = F[I_s(K)]$, де F – алгоритм пошуку, що вказує які операції слід зробити на кроці K при $I_s(K)$, щоб отримати рішення $I_s(K+1) \succ I_s(K)$. Тут знак переваги \succ при мінімізації функціонала має сенс:

$$J[I_s(K+1)] < J[I_s(K)]. \quad (8.51)$$

Алгоритм рішення задачі оптимізації F виконує функції збору інформації і прийняття рішення. Наприклад, в алгоритмах прямого випадкового пошуку (ПВП) задаються напрямки пошуку і визначаються значення функціонала J в точках $I_s(K) \pm \gamma\sigma$. Рішення полягає у виборі кроку в напрямку зменшення цього функціонала:

$$I_s(K+1) = I_s(K) - \omega\sigma\{J[I_s(K) + \gamma\sigma] - J[I_s(K) - \gamma\sigma]\}, \quad (8.52)$$

де ω, σ, γ – параметри, що визначають сфери прийняття рішення (ω) і збору інформації (γ) та одиничний випадковий напрямок (σ). У загальному випадку ці параметри можуть змінюватися (адаптуватися) до процедури пошуку і виду гіперповерхні прийнятого функціонала.

Розвитком методів випадкового пошуку, насамперед, у підвищенні ефективності процедур спрямованого пошуку, є еволюційні алгоритми, що реалізують біокібернетичний підхід для пошуку оптимальних рішень. Вони моделюють процес біологічної еволюції: мутації структури і параметрів δI_s , їх схрещування (розмноження) $I_s(K+1) = I_s(K) + \delta I_s(K)$ і правило добору. Це дозволяє виявляти їх сприятливі варіації, за допомогою яких будується послідовність поліпшуваних рішень із властивістю (8.51).

Найбільше поширення серед цих алгоритмів одержали генетичні алгоритми (ГА), засновані на моделюванні розвитку біологічної популяції на рівні ге-

номів. Нехай у задачі оптимізації за допомогою ГА потрібно знайти мінімум функціонала (функції пристосованості) $J[I_s]$. Популяція (геном) являє собою набір векторів $R_t = \{I_{si}(t)\}, i = \overline{0, M}$, де M – розмір популяції, а t – час її життя (шаги K). Елементи $I_{si}(t)$ – особи (гени-рішення), а поточний набір $\{I_{si}(t)\}$ складає генофонд популяції.

Елементи множини R_t здатні еволюціонувати за наступними правилами:

1) якщо $J[I_{si}(t)]$ малий, то особа $I_{si}(t)$ вважається вдалою й одержує пріоритет при розмноженні. Імовірність загибелі цієї особи знижується;

2) якщо $J[I_{si}(t)]$ великий, то особа $I_{si}(t)$ вважається невдалою, імовірність розмноження для цієї особи знижується і підвищується імовірність загибелі;

3) мутації: будь-яка особа має рівну імовірність мутації, тобто зсуву на невелику величину $I_{si}(t) \rightarrow I_{si}(t) + \delta I_s$. Закон визначення δI_s залежить від реалізації алгоритму. Типовий вибір – багатомірний нормальний розподіл з нульовим математичним очікуванням;

4) розмноження (схрещування): відповідно до ймовірностей, визначених на кроках 1 і 2, кожна особа має імовірність розмноження тим більшу, чим менше відповідний їй функціонал. При схрещуванні дві особи утворюють нові:

$$I_{si}(t+1) = c_t[I_{sj}(t) - I_{si}(t)] + I_{si} + \chi_t; \quad I_{sj}(t+1) = c_t[I_{si}(t) - I_{sj}(t)] + I_{sj} + \chi_t, \quad (8.53)$$

де c_t – скаляр і χ_t – вектор, що залежать від часу $t; i, j \in M$.

Схрещування передбачає далекі (випадкові) стрибки в просторі пошуку, забезпечуючи можливість виходу з локальних екстремумів. Крім того, схрещування "лагодить" деякі ушкодження, що роблять мутації, і забезпечує постійну мінливість;

5) загибель (добір): відповідно до імовірності, визначеної на кроках 1 і 2, особа може загинути, тобто бути вилученою з множини R_t .

Тут у наявності спадковість, мінливість і добір, – рушійні сили біологічної еволюції за Ч. Дарвіним.

До переваг ГА відносять здатність знаходження глобального екстремуму і продуктивність, а до їх недоліків – потребу у великому обсязі пам'яті (пропорційному розміру популяції) і відносно низьку швидкодію на фон-нейманівських ЕОМ.

8.4.3. Критерії оптимізації та обмеження.

При розв'язанні задачі (8.49)-(8.50) актуальним є також вибір ефективних для конкретного випадку критеріїв якості моделі (функціонала J). Для структурної ідентифікації ефективними вважаються зовнішні критерії, що адекватні задачі побудови моделей із мінімальною дисперсією похибки прогнозу, і поділяються на критерії регулярності і критерії незміщеності (мінімуму зсуву).

До критеріїв регулярності відноситься критерій мінімуму відносної похибки покрокового інтегрування:

$$J_{\varepsilon} = \frac{\|Y^*[k+n] - \hat{Y}[k+n]\|}{\|Y^*[k+n]\|}, \quad (8.54)$$

який обчислюється на всій вибірці експериментальних даних N , а також критерій мінімуму вибіркової відносної похибки:

$$J_{\varepsilon B} = \frac{\|Y^*[k+n] - \hat{Y}_A[k+n]\|_B}{\|Y^*[k+n]\|_B}. \quad (8.55)$$

Тут оптимізація моделі здійснюється на навчальній вибірці A (вихід моделі $\hat{Y}_A[k+n]$), а перевірка її ефективності (величини похибки) на перевірочній послідовності B , що відмічене в (8.55) індексом знизу. Вся вибірка $N = A + B$.

Критерії (8.54), (8.55) чутливі до рівня шуму в початкових даних і при збільшенні завад їх мінімум зміщується в область простіших моделей.

Стійкіші до завад критерії незміщеності. Наприклад, критерій мінімуму зсуву, заснований на аналізі рішень, має вигляд:

$$J_{см} = \frac{\|\widehat{Y}_A[k+n] - \widehat{Y}_B[k+n]\|}{\|Y^*[k+n]\|}, \quad (8.56)$$

де $\widehat{Y}_A[k+n]$ і $\widehat{Y}_B[k+n]$ – виходи моделей, навчених на вибірках A і B , відповідно. Тут обчислення зсуву здійснюється на всій вибірці N .

Для узгодження вимог критеріїв (8.54) і (8.56) використовується комбінований критерій з ваговим коефіцієнтом μ , наприклад, у вигляді:

$$J_{комб} = \mu J_{\varepsilon} + (1 - \mu) J_{см}; \quad 0 \leq \mu \leq 1. \quad (8.57)$$

Для критеріїв регулярності навчальна і перевірна послідовності (вибірки), зазвичай, рівні $A = 0,7N$ і $B = 0,3N$, а для критеріїв мінімуму зсуву – $A = 0,5N$ і $B = 0,5N$.

Обмеження, що накладаються при розв'язанні задачі (18.49)-(8.50), $I_s(K) + \delta I_s(K) \in S$ припускають $\Phi + \delta\Phi \in S$ і $a + \delta a \in S$. Обмеження $S_1 = S_h \cap S_g$ формують безперервну задачу математичного програмування, а $S_2 = S_h \cap S_D$ – задачу дискретного програмування. При цьому обмеження $S = \{S_1, S_2\}$ можна врахувати наступним чином:

- при обмеженнях типу нерівності $S = S_h$ вдалим вважається крок, коли виконується умова (8.51) і $I_s(K) \in S$, інакше крок невдалий і треба повернутися в попередній стан з подальшим новим кроком із нього;

- при обмеженнях типу рівності $S = S_g$ вводять коридор $\|g(I_s)\| < \varepsilon_{дон}$, обмежений допустимою похибкою $\varepsilon_{дон}$;

- обмеження $S = S_h \cap S_g$ є комбінацією перших двох випадків;

- при обмеженнях на дискретність структурних функцій $S = S_D$ здійснюється випадковий вибір нової точки біля початкової точки: $\|I_{si}(K+1) - I_{si}(K)\| < \varepsilon$, де I_{si} – вектор структурно-параметричних факторів зі значенням структурної функції ϕ_i . Якщо $\Theta_{\varepsilon}(K)$ – множина векторів I_{si} , що задовольняє вищенаведеній умові, то на $K+1$ кроці із $\Theta_{\varepsilon}(K)$ обирають точку

(значення I_{si}), яке задовольняє умовам $\mathcal{J}[I_{si}(K+1)] < \mathcal{J}[I_{si}(K)]$ і $I_{sz}(K+1) \in \Theta_\varepsilon(K)$.
Очевидно, що при $S = S_h \cap S_D$ додатково потрібно, щоб $I_{si}(K+1) \in S_h$.

8.4.4. Структура і базисні функції моделі.

Однією з основних проблем при рішенні задачі ідентифікації ОК є вибір типу структури моделі. Зазвичай структуру визначають, виходячи з фізичних законів, що реалізуються в системі. Проте, така модель, часто, має високу розмірність, що утрудняє її практичне використання.

Продуктивнішим представляється визначення структури моделі ОК на основі спостережень його вхідних і вихідних змінних із використанням апроксимацій у вигляді функціональних рядів Вольтерра, поліномів Колмогорова-Габора, блочно-орієнтованих моделей тощо.

У блочно-орієнтованих моделях нелінійні динамічні ОК представляються шляхом композиції лінійного динамічного (ЛДБ) і нелінійного статичного (НСБ) блоків, наприклад, у вигляді моделей Вінера, Гаммерштейна або їх комбінацій (дивись, наприклад, рис. 8.2).

Модель Вінера містить послідовно сполучені лінійний динамічний і нелінійний статичний блоки (рис. 8.18,*а*), а в моделі Гаммерштейна, навпаки, лінійний блок приведений до виходу (рис. 8.18,*б*).

Модель Вінера-Гаммерштейна утворюється шляхом об'єднання блоків моделей Вінера і Гаммерштейна, для чого вводиться зворотний зв'язок (рис. 8.19,*а*), а модель Гаммерштейна-Вінера – послідовним з'єднанням моделей Гаммерштейна і Вінера, при цьому в центрі з'єднання виявляються два лінійні динамічні блоки, які об'єднуються в один (рис. 8.19,*б*).

На рис. 8.18 і 8.19 як ЛДБ використовуються лінії затримки T , величини яких (глибина пам'яті) визначаються розмірністю вхідних d_i і вихідних $d_{\hat{y}}$ змінних, а НСБ – як традиційні засоби: поліноми Лежандра або Колмогорова-Габора, так і інтелектуальні – НМ, гібридні НМ з нечіткою логікою тощо. Вхо-

дом моделей є вектори спостережень $\{z_i[k]\} \in \hat{Z}[k]$ для моментів часу k , а виходом – його прогноз на n тактів ($\hat{Y}[k+n]$).

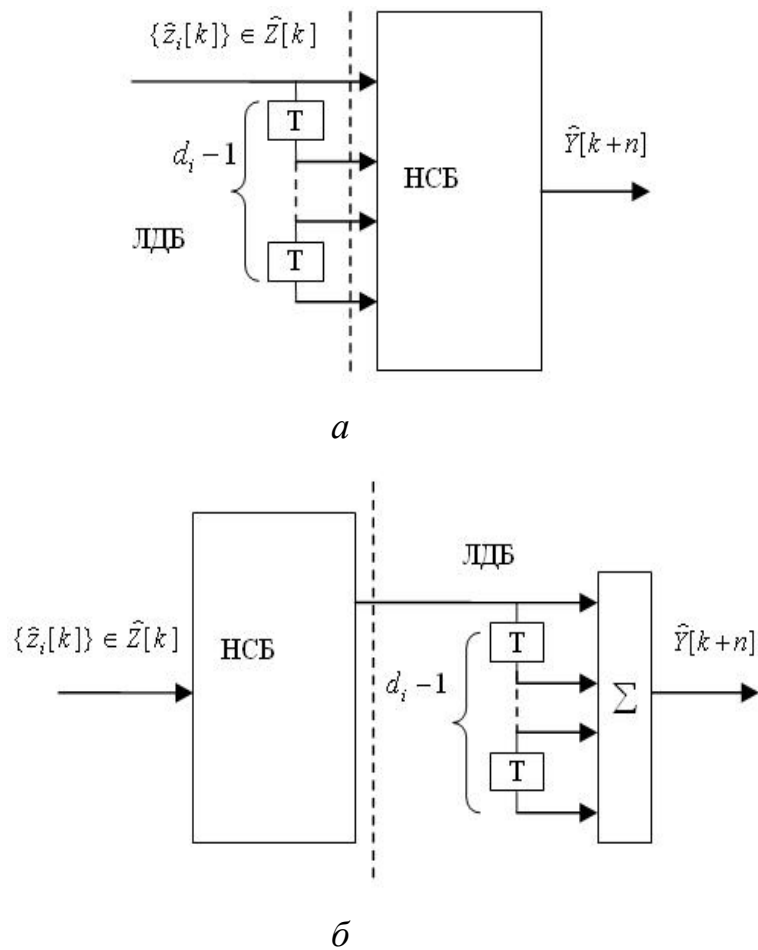


Рис. 8.18. Структура Вінера (а) і Гаммерштейна (б) прогнозуючої моделі нелінійного динамічного ОК

Окрім моделей Вінера, Гаммерштейна і їх комбінацій до блочно-орієнтованих моделей відносяться також моделі авторегресії з додатковим входним сигналом – ARX-моделі, які описують нелінійні структури, використовуючи паралельну комбінацію нелінійних і лінійних блоків.

Розглянемо, для прикладу, формування базисних функцій на основі НМ та Anfis, які використовувались при побудові відповідних прогнозуючих фільтрів (див. п.п. 8.1.2, 8.1.3).

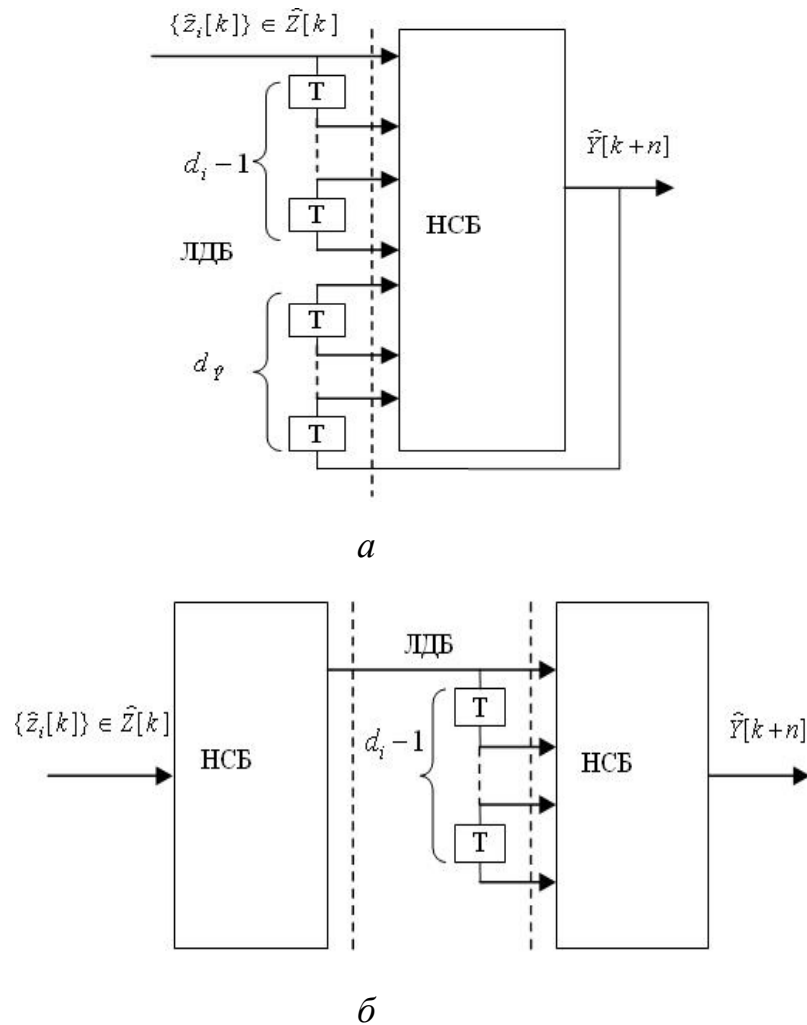


Рис. 8.19. Структура Вінера-Гаммерштейна (а) і Гаммерштейна-Вінера (б) прогнозуючої моделі нелінійного динамічного ОК

Рівняння ОК (8.46) при використанні базових функцій у вигляді НМ прямого розповсюдження (НМПР) із прихованим шаром являє собою рівняння згортки:

$$\hat{Y}[k+n] = \sum_{\tau \in P} \Phi_{\hat{Y}} \left\{ \sum_{l \in Q} v_l[\tau] \cdot \Phi_l \left(\sum_{m \in Q} v_{l,m}[\tau] \cdot y_m[k-\tau] \right) \right\}, \quad (8.58)$$

де P – множина глибини пам'яті відповідних входів; $\Phi_{\hat{Y}}$ – активаційна функція вихідного шару НМ; Q – множина входів нейронів; l – порядковий номер входу вихідного шару НМ; v_l – вагові коефіцієнти вихідного шару; Φ_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; y_m – вхід НМ.

У загальному випадку входами НМ (8.58) згідно (8.46), (8.47) є $\{\hat{Y}[k], \hat{Z}[k]\} \subset \{y_m[k-\tau]\}$, а її структурними характеристиками – $\{T_s, P, \Phi_{\hat{Y}}, \Phi_l, r_s\} \subset \Phi$, де T_s – тип структури і розмір прихованого шару – $r_s \subset Q$. При цьому параметрами НМ є $\{v_l, v_{l,m}\} \subset a$.

Рівняння ОК на основі НМ із РБФ має вигляд:

$$\hat{Y}[k+n] = \sum_{\tau \in P} \Phi_{\hat{Y}} \left\{ \sum_{l,m \in Q} v_l \cdot \Phi_l(\vartheta_l, \|y_m[k-\tau] - v_l\|) \right\}, \quad (8.59)$$

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

Структурними характеристиками НМ (8.59) є $\{T_s, P, \Phi_{\hat{Y}}, \Phi_l, r_s\} \subset \Phi$, а її параметрами – $\{v_l, \vartheta_l, v_l\} \subset a$.

Рівняння ОК на основі гібридної НМ з нечіткою логікою Anfis являє собою рівняння згортки виду:

$$\hat{Y}[k+n] = \sum_{\tau \in P} \sum_{m \in Q} \beta_m[\tau] \cdot \alpha_m[k-\tau], \quad (8.60)$$

де $\beta_m[\tau] = U_m^{-1}(\alpha_m[\tau] / \sum_m \alpha_m[\tau])$; $\alpha_m[k-\tau] = T_n \{L_{l,m}(y_m[k-\tau])\}$; $U = U(a_U)$;

$L = L(a_L)$.

Тут U_m^{-1} – функція, зворотна функції належності проміжного виходу m мережі із параметрами a_U ; α_m – значення проміжного виходу; T_n – довільна t -норма моделювання логічної операції «І»; $L_{l,m}$ – функція належності нечіткого правила l входу m із параметрами a_L .

Структурними характеристиками НМ (8.60) є $\{T_s, P, U_m, L_{l,m}, r_p\} \subset \Phi$, де кількість правил розкладання по входам $r_p \subset Q$, а її параметрами – $\{a_U, a_L\} \subset a$.

Ідентифікація параметрів (навчання) НМПР (8.58) здійснюється, зазвичай, за допомогою градієнтних алгоритмів, наприклад, алгоритму зворотного поширення похибки в просторі параметрів $\{v_l, v_{l,m}\} \subset a$ при заданих структурі моделі ОК і структурних характеристиках Φ з метою мінімізації похибки прогнозу (8.49) чи (8.54).

При навчанні НМ із РБФ (8.59) спочатку визначаються центри і відхилення для радіальних елементів, після цього оптимізуються параметри лінійного вихідного шару: $\{v_l, \theta_l, v_l\} \subset a$.

Навчання гібридної НМ Anfis (8.60) виконується аналогічно НМ (8.58) шляхом оптимізації параметрів функцій належності мережі $\{a_U, a_L\} \subset a$.

Перевагою цих алгоритмів параметричного навчання НМ є простота і швидкодія, а недоліком – їх локальність (висока ймовірність застрягання у локальному екстремумі).

Таким чином до структурних характеристик необхідно віднести і множини методів навчання НМ.

Критерії (8.55)-(8.56) використовуються для структурної оптимізації, а критерій (8.54) – для параметричної.

Відповідно вищевикладеному, задача (8.49)-(8.50) має велику розмірність (кількість структурних характеристик і параметрів у реальності становлять сотні, тисячі й десятки тисяч). Внаслідок цього самостійне використання пошукових алгоритмів глобальної оптимізації для структурно-параметричної ідентифікації має низьку ефективність (малу швидкість збіжності) через їх відносно низьку швидкодію. Тому доцільним є реалізація спільного використання (композиція) алгоритмів глобальної оптимізації (для структурної ідентифікації моделі) і алгоритмів локальної оптимізації (для параметричного навчання та ідентифікації моделі).

8.4.5. Приклад структурно-параметричної ідентифікації нелінійного об'єкту.

Як приклад виконано ідентифікацію нелінійного ОК, що описується послідовно з'єднаними аперіодичною ланкою із запізненням та безінерційною квадратичною ланкою (ПТ2).

Для моделювання ОК (моделі ПТ2) в якості вхідного сигналу використовувався смуговий кореляційний нормований сигнал із гаусовим шумом амплітудою 10% від рівня сигналу (відношення сигнал/шум приблизно 20 дБ). При

цьому, у відповідності із динамічними властивостями ОК глибина прогнозу склала $n = 3$ такти, глибина пам'яті – 4 такти, а розмір реалізації – $N = 1024$.

У якості критерію структурної оптимізації обрано комбінований критерій, поверхні якого для ОК у просторі «кількість нейронів прихованого шару НМ (8.58) – рівень шуму», а також «кількість нейронів – глибина прогнозу» наведені на рис. 8.20.

Легко бачити, що цей критерій має малу чутливість до варіації шуму і глибини прогнозу (малий зсув глобального мінімуму у просторі ознак).

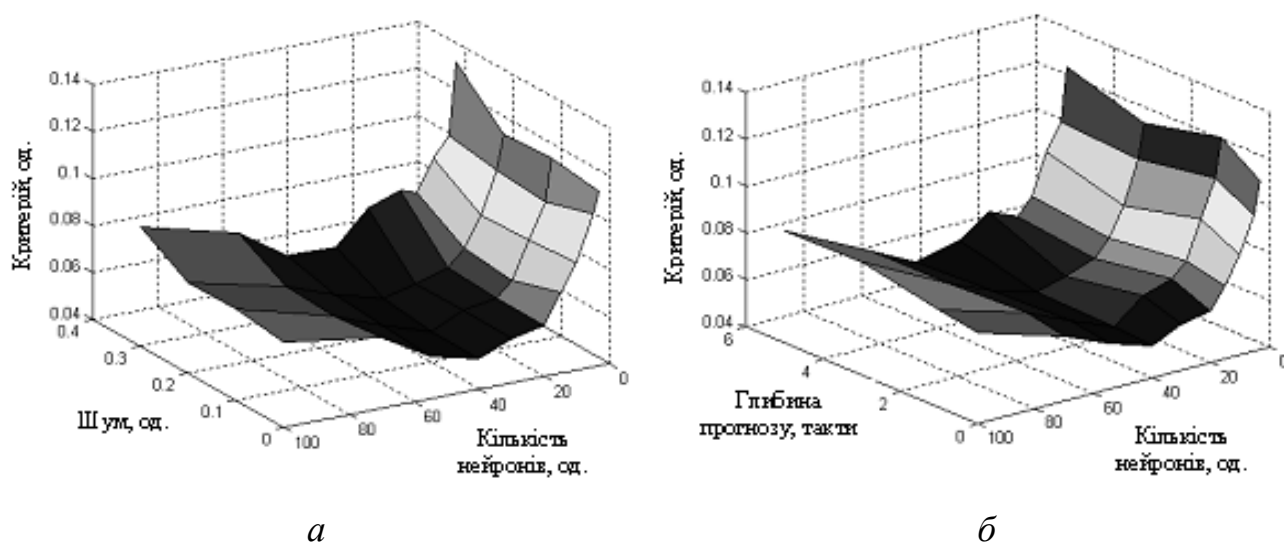


Рис. 8.20. Вигляд поверхні комбінованого критерію для ОК при варіаціях шуму (*а*) та глибини прогнозу (*б*)

Із рис. 8.20 можна також визначити обмеження на кількість нейронів прихованого шару в НМ (8.58) для ідентифікації ОК – від 20 до 80 нейронів.

При ідентифікації використовувалася структура моделей Вінера-Гаммерштейна із базисними функціями НМ (8.58)-(8.60), а у якості глобальних методів оптимізації застосовувалися алгоритм прямого випадкового пошуку (ПВП) і генетичний алгоритм (ГА). Результати глобальної оптимізації моделі ОК наведені на рис. 8.21,*а* і *б* (для ПВП) і на рис. 8.21,*в* і *г* (для ГА).

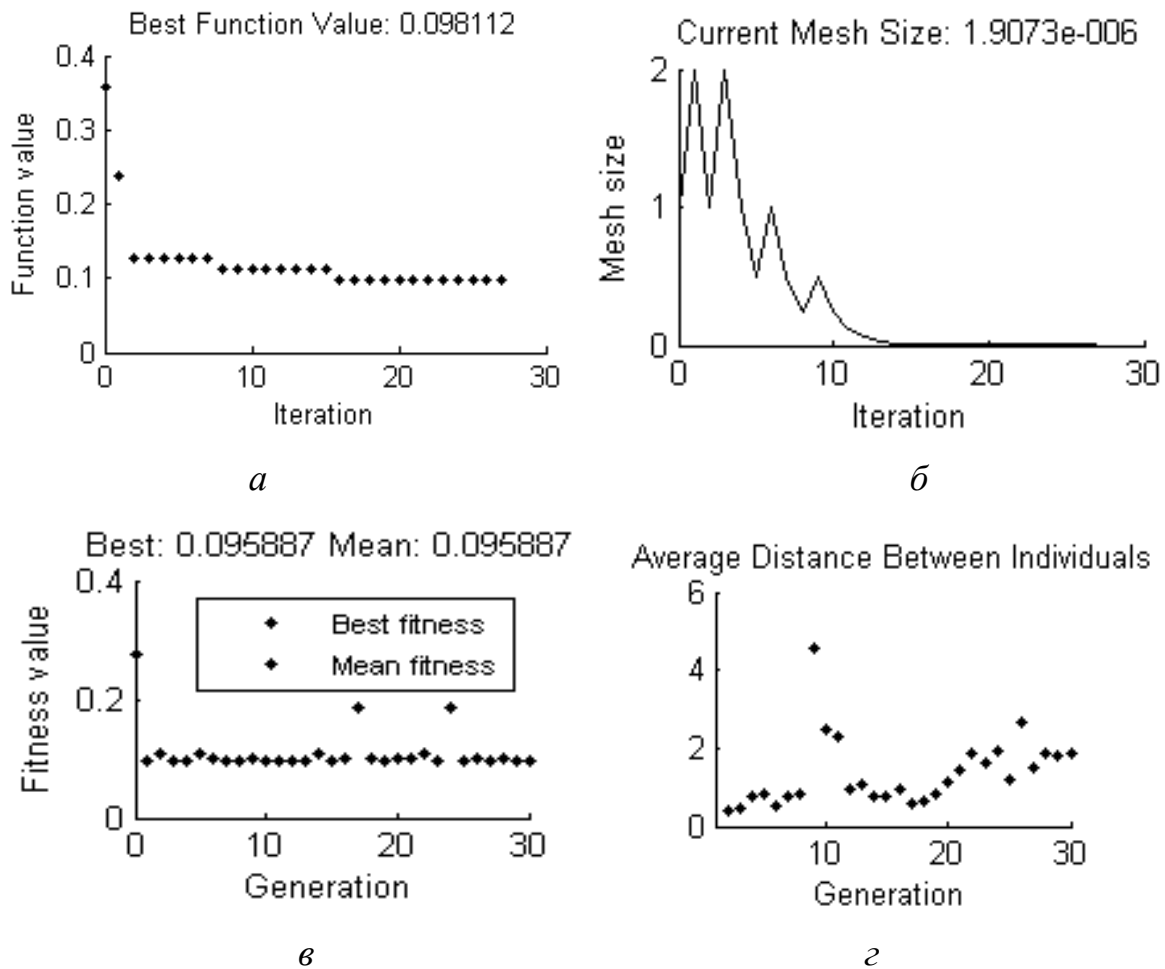


Рис. 8.21. Результати глобальної оптимізації структури моделі за ПВП (а, б) і за ГА (в, з)

ГА мав одноточкове схрещування, селективний вибір батьків та формування нової популяції із витисненням, а алгоритм ПВП – прискорювач (адаптивний механізм) кроку пошуку. Кількість ітерацій (для ГА – поколінь) обмежувалась 150, а розмір простору пошуку (популяції) – 30.

В цілому ці алгоритми по точності дали подібні результати. При цьому ГА виявив вищу швидкість збіжності (ГА виходить в область оптимальних рішень на перших поколіннях, а ПВП – після 15 ітерацій), а алгоритм ПВП – вищу швидкодію (приблизно 6 с на ітерацію у ПВП і 9,5 с на покоління у ГА) при обчисленнях на комп'ютері із процесором Pentium IV.

В результаті моделювання встановлено, що мінімуму комбінованого критерію для розглянутих процесів відповідають каскадні НМ моделі прямого поширення (8.58) із логістичною функцією активації прихованого шару і лінійною функцією у вихідному шарі. При цьому кількість нейронів у прихованому шарі

для моделі ОК складає 47 нейронів. Тип методу параметричного навчання НМ виявився несуттєвим фактором.

Моделі у вигляді НМ із РБФ (8.59) потребують значно менше обчислень, але їх значення комбінованого критерію суттєво вище, що можна пояснити поганою прогностичною здатністю цих НМ.

Моделі у вигляді НМ із нечіткою логікою Anfis (8.60) мають достатню точність, але у них дуже низька швидкодія при розмірностях вектора входів більше 5.

Як міру точності ідентифікованої моделі використовувався критерій мінімуму відносної похибки покрокового інтегрування (8.54), значення якого склало – 0,0348.

Контрольні запитання

1. Сформулюйте задачу адаптивного прогнозування сигналів і стану об'єктів керування.
2. Розкрийте сутність нейромережевого та нейронечіткого адаптивних фільтрів-апроксиматорів.
3. Розкрийте сутність нечітких адаптивних фільтрів-апроксиматорів з субтрактивною кластеризацією та кластеризацією С-середніх.
4. Сформулюйте задачу ідентифікації нелінійних процесів за часовими реалізаціями.
5. Які характеристики мають нелінійні хаотичні процеси?
6. Сформулюйте задачу структурно-параметричної ідентифікації об'єктів керування.
7. Розкрийте сутність структурної оптимізації.
8. Які критерії використовуються при структурно-параметричній ідентифікації об'єктів керування.
9. Які базисні функції використовуються при структурно-параметричній ідентифікації об'єктів керування.
10. Які особливості ідентифікації та прогнозування процесу дроблення руди за часовими реалізаціями?
11. Які особливості прогнозування теплового стану доменної печі та показників якості чавуну?

Глава 9. КОМПЛЕКСНА ОЦІНКА, ІДЕНТИФІКАЦІЯ ТА ПРОГНОЗУВАННЯ ТРАФІКУ В ІНФОРМАЦІЙНИХ ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖАХ

Постійне збільшення обсягу трафіку в інформаційних телекомунікаційних мережах (ІТМ) породжує необхідність його ідентифікації і прогнозування для попередження перевантажень в мережах і уникнення погіршення якості послуг, що надаються.

Трафік в ІТМ є нелінійним стохастичним процесом з властивостями самоподоби та з хаотичною і фрактальною динамікою. Крім того, встановлено, що агрегований трафік від різних джерел на малих часових масштабах проявляє мультифрактальний характер.

Розвиток телекомунікаційних мереж пов'язаний з їх плануванням, що дозволяє визначати необхідні мережеві ресурси і забезпечити задану комутаційну ємність. Проектування засобів телекомунікацій, заснованих на статистичних даних і необхідному математичному апараті з використанням прогнозування, дозволяє збалансувати навантаження на маршрутизатори, комутатори та канали зв'язку і, таким чином, підвищити надійність і ефективність мереж.

Оцінка характеристик мережевого трафіку необхідна не тільки для класифікації породжуюючих процесів, але і для побудови його адекватної імітаційної моделі, що дозволяє забезпечити якість обслуговування в ІТМ.

Таким чином актуальним є побудова і використання для керування передачею інформації моделей мережевого трафіку на основі оцінки його характеристик.

Дослідження властивості самоподібності мережевого трафіку призвело до виникнення моделей на основі фрактальних випадкових процесів. У главі 8 розглянуто комплексний метод ідентифікації та оцінки характеристик складних нелінійних процесів, який включає процедури:

- 1) оцінки характеристик процесу;
- 2) вибору типу (структури) моделі процесу;
- 3) визначення параметрів моделі (параметричну ідентифікацію процесу).

9.1. Комплексна оцінка характеристик трафіку

Дослідження сучасного трафіку в ІТМ показали, що він є нелінійним стохастичним процесом з хаотичною і фрактальною динамікою, властивості якого значно відрізняються від пуассонівських моделей.

Фрактальним (самоподібним) є стаціонарний випадковий процес $x = \{x_t : t = 0, 1, \dots\}$, де t – такт часу, із постійним середнім $\mu = E[x_t]$, нескінченною дисперсією $\sigma^2 = E[(x_t - \mu)^2]$ і автокореляційною функцією (АКФ) $r(k) = E[(x_t - \mu)(x_{t+k} - \mu)] / E[(x_t - \mu)^2]$; ($k = 0, 1, 2, \dots$), яка залежить від зсуву такту часу k і дорівнює АКФ $r^{(m)}$ агрегованого процесу $x^{(m)} = (x_k^{(m)} : k = 1, 2, 3, \dots)$, усередненого за блоками довжини $m=1, 2, 3, \dots$. Компоненти цього процесу визначаються як $x_k^{(m)} = (1/m) \cdot (x_{km-m+1} + \dots + x_{km})$, ($k \geq 1$), тобто $r^{(m)}(k) = r(k)$. Крім цього, АКФ можна записати як:

$$r(k) \sim L_1(t)k^{-\beta} \text{ при } k \rightarrow \infty, \quad (9.1)$$

де L_1 – поволі змінювана на нескінченності функція, тобто для $x > 0$ вона має $\lim_{t \rightarrow \infty} L_1(xt) / L_1(t) = 1$; $\beta = 2 - 2H$; $0 < \beta < 1$; H – параметр Херста.

На основі моделювання властивостей трафіку можливо розробити засоби забезпечення якості обслуговування в ІТМ. Проте, в даний час немає єдиних загальновизнаних моделей самоподібного трафіку і не існують достовірні і визнані методики оцінки характеристик і показників якості ІТМ при самоподібному трафіку.

Одні дослідники підходять до рішення задачі ідентифікації мережевого трафіку із використанням фрактального аналізу, інші – статистичного, а треті – спектрального. Разом з тим, сумісне використання цих трьох підходів

дозволить, на нашу думку, підвищити достовірність визначення властивостей, оцінки характеристик і, зрештою, отримати адекватні моделі трафіку в ІТМ.

9.1.1. Часо-частотний аналіз.

Аналіз АКФ процесу дозволяє визначити, яка із залежностей характерна для нього: довготривала або короткострокова.

Для самоподібних процесів з довготривалою залежністю (з поволі убуючою залежністю – ПУЗ) АКФ визначається по формулі (9.1) і є неінтегрованою ($\sum_k r(k) = \infty$), при цьому вона гіперболічно убואє зі зростанням часової затримки.

Навпаки, процеси з короткостроковою залежністю (з швидко убуючою залежністю – ШУЗ) характеризуються експоненційно спадаючою інтегрованою АКФ:

$$r(k) \sim c\rho^k, \quad k \rightarrow \infty, \quad 0 < \rho < 1, \quad 0 < \sum_k r(k) < \infty, \quad (9.2)$$

де c – константа.

Для визначення типу залежності в експериментальному часовому ряді необхідно його АКФ апроксимувати за допомогою формул (9.1) і (9.2), визначивши параметри моделей $L_1(t)$, β , c , ρ , наприклад, методом найменших квадратів. Далі по величині дисперсії різниць (похибки) між експериментальною АКФ і її апроксимаціями можна встановити тип процесу (ПУЗ або ШУЗ), реалізований в досліджуваному часовому ряді. Спектральна щільність самоподібного процесу підкорюється ступеневому закону:

$$S(f) \sim L_2(t)f^{-\gamma} \quad \text{при } f \rightarrow 0, \quad (9.3)$$

де $L_2(t)$ – поволі змінювана в нулі функція; $0 < \gamma < 1$; $\gamma = 2H + 1$.

При частоті $f \rightarrow 0$ спектральна щільність $S(f) \rightarrow \infty$. Процес даного типу називають « $1/f$ - процесом» або «флікер-шумом».

Самоподібні процеси також можуть бути класифіковані залежно від значення показника γ . Так, до фрактального броунівського руху (ФБР)

відносяться нестационарні випадкові процеси, що мають нескінченну низькочастотну потужність і $1 < \gamma < 3$. При $\gamma = 2$ процес є класичним броунівським рухом. Похідними від ФБР є стаціонарні гаусові процеси, які мають $-1 < \gamma < 1$. Окремим випадком процесів даного типу є стаціонарний білий гаусовий шум з $\gamma = 0$.

Одним з методів якісного і кількісного визначення саподібності процесу є вейвлет аналіз. Його скелетон (потужність коефіцієнтів вейвлет перетворення) показує наявність самоподоби у вигляді розвиненої деревоподібної структури з розгалуженнями (гілками), залежність від масштабу яких описується по ступеневому закону.

Підраховуючи число точок максимумів коефіцієнтів вейвлет-перетворення $N(\eta)$ уздовж параметра зсуву в області масштабу η можна оцінити значення параметра Херста:

$$H = \log[N(\eta)] / \log(\eta). \quad (9.4)$$

9.1.2. Статистичний аналіз.

Якщо дисперсія процесу поволі убуває, тобто дисперсія вибіркового середнього має повільніший спад, ніж величина, зворотна довжині вибірки:

$$\sigma^2(x^{(m)}) \sim m^{-\beta} \text{ при } m \rightarrow \infty, \quad (9.5)$$

то даний процес є самоподібним. Якщо ж дисперсія швидко убуває, тобто вона зменшується зі зростанням обсягу вибірки:

$$\sigma^2(x^{(m)}) = \frac{1}{m} \sigma^2(x), \quad (9.6)$$

то даний процес описується традиційною пуассонівською моделлю пакетного трафіку.

З виразу (9.5) отримаємо оцінку коефіцієнта β :

$$\beta \sim -\log[\sigma^2(x^{(m)})] / \log(m) \text{ при } m \rightarrow \infty. \quad (9.7)$$

При $\beta = [0, 1]$ даний процес є самоподібним.

Інструментом статистичного аналізу є дослідження функції розподілу. Для самоподібних процесів зі ступеневим (гіперболічним) убуттям АКФ характерним є розподіл з «важким хвостом».

Випадкова величина Z має розподіл з «важким хвостом», якщо

$$P[Z > x] \sim cx^\alpha \text{ при } x \rightarrow \infty, \quad (9.8)$$

де $0 < \alpha < 2$ – параметр форми (показник «тяжкості хвоста»), c – позитивна константа.

Найчастіше для апроксимації гістограм експериментальних даних самоподібних процесів застосовуються функції субекспоненціальних законів розподілу:

- Парето:

$$F(x) = P[Z \leq x] = 1 - \left(\frac{k}{x}\right)^a, \quad f(x) = \frac{ak^a}{x^{a+1}}; \quad (9.9)$$

- Вейбулла:

$$F(x) = 1 - e^{-(x/a)^b}, \quad f(x) = \frac{b}{a} \left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b}; \quad (9.10)$$

- логнормальний:

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}} \int_0^x \frac{1}{t} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dt, \quad f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad (9.11)$$

де a, b, k, μ, σ – параметри розподілів.

Для перевірки адекватності теоретичних розподілів (9.9-9.11) експериментальним даним використовуються критерії згоди Колмогорова і Пірсона.

Показник «тяжкості хвоста» (параметр форми) α визначають по методу Хілла або шляхом побудови графіка додаткового розподілу в подвійному логарифмічному масштабі (LLCD).

Оцінка Хілла визначає показник форми як функцію найбільших елементів набору даних N :

$$\hat{\alpha} = \left(\frac{1}{N} \sum_{i=1}^N \log \frac{x_{(i)}}{x_{(N+1)}} \right)^{-1}, \quad (9.12)$$

де $x_{(1)} > x_{(2)} > \dots$ – порядкові статистики.

По графіку залежності $\hat{\alpha}$ від N значення показника форми, починаючи з якого відбувається стабілізація лінії графіка, є оцінкою показника «тяжкості хвоста».

Другий метод полягає в побудові додаткової функції розподілу $\bar{F}(x) = 1 - F(x)$. При цьому графік функції на ділянці, що представляє хвіст розподілу, є прямою лінією, для якої при великих значеннях x справедливий вираз

$$\alpha \sim - \frac{d \log \bar{F}(x)}{d \log x}. \quad (9.13)$$

Таким чином, присутність розподілу з «важким хвостом» в даних підтверджується наявністю на графіку лінійної ділянки. Побудувавши до неї лінію регресії, можна визначити шуканий показник α , який рівний тангенсу кута нахилу цієї лінії.

Варто відзначити, що для самоподібних процесів показник α пов'язаний з показником Херста співвідношенням

$$H = \frac{3 - \alpha}{2}. \quad (9.14)$$

Для перевірки нульової гіпотези про незалежність і тотожність розподілу значень часового ряду запропоновано BDS-тест, який також дозволяє виявити нелінійність породжуючої системи, відрізнити випадкові системи від детермінованого хаосу або від нелінійних стохастичних систем.

Даний тест базується на обчисленні BDS-статистики:

$$w_{d,N}(\varepsilon) = \sqrt{N - d + 1} \frac{C_{d,N}(\varepsilon) - (C_{1,N-d}(\varepsilon))^d}{\sigma_{d,N}(\varepsilon)}, \quad (9.15)$$

де $C_{d,N}(\varepsilon)$, $(C_{1,N-d}(\varepsilon))^d$ – кореляційні інтеграли $\sigma_{m,N}(\varepsilon)$ – середньоквадратичне відхилення чисельника.

Кореляційний інтеграл $C_{m,N}(\varepsilon)$ показує відносне число пар точок атрактора, що знаходяться на відстані не більше ε і визначається як:

$$C_{d,N}(\varepsilon) = \frac{2}{(N-d+1)(N-d)} \sum_{s=d}^N \sum_{t=s+1}^N \prod_{j=0}^{d-1} I_r(x_{s-j}^d, x_{t-j}^d), \quad I_r(x_i^d, x_j^d) = \begin{cases} 1, & \|x_i^d - x_j^d\| \leq \varepsilon; \\ 0, & \|x_i^d - x_j^d\| > \varepsilon, \end{cases} \quad (9.16)$$

де $I_r(x_i^d, x_j^d)$ – функція Хевісайда для всіх пар значень i й j ($0 \leq j \leq N$); N – число елементів часового ряду $\{x_i\}_{i=1}^N$; d – розмірність вкладення.

Доведено, що при $N \rightarrow \infty$ інтеграл $C_{d,N}(\varepsilon) \Rightarrow (C_{1,N}(\varepsilon))^d$, а $C_{d,N}(\varepsilon) - (C_{1,N-d}(\varepsilon))^d$ є нормально розподіленою величиною з нульовим середнім і середньоквадратичним відхиленням:

$$\sigma_{d,N}^2 = 4 \left(p^d + 2 \sum_{j=1}^{d-1} p^{d-j} (C_{1,N}(\varepsilon))^{2j} + (d-1)^2 (C_{1,N}(\varepsilon))^{2d} - d^2 p (C_{1,N}(\varepsilon))^{2d-2} \right), \quad (9.17)$$

$$\text{де } p = \frac{1}{(N-1)(N-2)N} \times \left\{ \sum_{t=1}^N \left[\sum_{s=1}^N I_r(x_t, x_s) \right]^2 - 3 \sum_{s=1}^N \sum_{t=s+1}^N I_r(x_t, x_s) + 2N \right\}.$$

Звідси випливає, що BDS-статистика $w_{d,N}(\varepsilon)$ є нормально розподіленою.

Якщо $w_{d,N}(\varepsilon)$ приймає значення $|w_{d,N}(\varepsilon)| \leq 1,96$, то нульову гіпотезу з вірогідністю 95% можна прийняти (спостерігається стохастичний процес, відліки якого незалежні, однаково розподілені випадкові величини). Інакше $|w_{d,N}(\varepsilon)| > 1,96$ – нульову гіпотезу необхідно відхилити (спостерігається хаотичний процес).

Якщо в результаті виконання BDS-теста для залишків (похибки) лінійної моделі виявиться, що нульову гіпотезу потрібно відхилити, то даний процес є нелінійним.

9.1.3. Фрактальний аналіз.

Вище (див. п. 8.2) розглянуті властивості та характеристики нелінійних динамічних процесів, які мають хаотичні режими, що суттєво впливає на методи їх ідентифікації. Розглянемо ці динамічні процеси в ІТМ. При цьому, для цілісності викладення матеріалу та спрощення його сприйняття, продублюємо тут деякі визначення та вирази, що наведені в п. 8.2.

Значна кількість експериментальних даних, отриманих при спостереженні природних явищ, у тому рахунку і мережевого трафіку, мають фрактальну (дробову) розмірність в часі.

Згідно теореми Такенса за часовою реалізацією спостережуваного сигналу $x = x(t)$, задаючи затримку τ і розмірність d фазового простору, можна отримати його дискретне відображення:

$$x[k] = \{x[k], x[k - m], x[k - 2m], \dots, x[k - (d - 1)m]\}, \quad (9.18)$$

де $m = \tau / T$; T – інтервал дискретизації за часом.

При переборі за тактом часу k маємо дискретний набір точок в d -мірному просторі, який при сталому режимі системи є фазовим портретом аттрактора.

При аналізі фазового портрета можна виявити напрями рухів за різних початкових умов, а також визначити якісні властивості динамічної системи, що породжує процес (9.18).

При відомому значенні кореляційного інтеграла, використовуючи залежність:

$$C(\varepsilon) \sim \varepsilon^{D_c}, \quad (9.19)$$

можна визначити кореляційну розмірність D_c :

$$D_c = \lim_{\varepsilon \rightarrow 0} \frac{\log C(\varepsilon)}{\log(\varepsilon)}. \quad (9.20)$$

Тоді, побудувавши графік кореляційного інтеграла в подвійному логарифмічному масштабі, можна визначити кореляційну розмірність як тангенс кута нахилу прямої, що апроксимує даний графік.

Розмірність фазового простору d , починаючи з якої D_C перестає змінюватися, є мінімальною розмірністю вкладення атрактора, тобто найменшою цілою розмірністю фазового простору, що вміщує весь атрактор. Оцінку розмірності визначають також по формулі Мане:

$$d \geq 2D_C + 1, \quad (9.21)$$

яка на практиці часто дає завищені значення d , тому обмежуються простором розмірності $d \geq D_C$.

Кореляційна розмірність D_C характеризує нижню межу фрактальної розмірності:

$$D_C = 2 - H. \quad (9.22)$$

Показник Херста H характеризує ступінь самоподоби процесу. Для його визначення використовують, наприклад, такі методи, як метод періодограм, вейвлет-аналіз, метод агрегованих дисперсій і R/S-аналіз. Згідно з останнім показник Херста визначається по формулі

$$H = \frac{\log(R / S)}{\log(cN)}, \quad (9.23)$$

де S – середньоквадратичне відхилення спостережуваного часового ряду, R – розмах накопиченого відхилення, N – число періодів спостережень, c – позитивна константа.

Даний показник свідчить про наявність тренда або про випадковість процесу, а також характеризує еволюцію досліджуваного процесу. Якщо $0,5 < H < 1$, то даний процес характеризується довготривалою пам'яттю і є персистентним, якщо ж $0 < H < 0,5$, то це говорить про антиперсистентність процесу. Значення $H = 0,5$ характерне для броунівського руху з незалежними добутками і відповідає випадковим відхиленням процесу від середнього.

Для визначення режиму породжуючого процесу, оцінюють його ентропію Колмогорова K , яка дорівнює сумі старших показників Ляпунова і характеризує швидкість втрати інформації про стан динамічної системи в часі:

$$K = -\lim_{T \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{NT} (K_{k+1} - K_k) = -\lim_{T \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{NT} \sum_{i_0 \dots i_k} P_{i_0 \dots i_k} \ln P_{i_0 \dots i_k}, \quad (9.24)$$

де $K_{k+1} - K_k$ – втрата інформації на інтервалі часу від k до $k + 1$; $P_{i_0 \dots i_k}$ – спільна імовірність перебування точки $x[0]$ в осередку фазового простору i_0 розміром ε , $x[1T]$ – в i_1 , ... і $x[kT]$ – в i_k ; N – тривалість часової реалізації.

K -ентропія дорівнює нулю при регулярному русі, нескінченна для випадкових систем, позитивна і обмежена для систем з динамічним хаосом.

Значення кореляційної ентропії оцінюють по кореляційному інтегралу $C(\varepsilon, d)$, залежному як від відстані ε , так і від розміру фазового простору d :

$$K_C(\varepsilon, d) = \frac{C(\varepsilon, d)}{C(\varepsilon, d+1)}. \quad (9.25)$$

Кореляційна ентропія також дозволяє визначити оцінку інтервалу точної передбачуваності процесу:

$$T_C = \frac{1}{K_C} \ln \left(\frac{1}{\varepsilon} \right). \quad (9.26)$$

Кореляційна ентропія K_C є нижньою межею K -ентропії. За час, більший T_C можливе тільки статистичне прогнозування.

9.1.4. Методика оцінки характеристик мережевого трафіку.

На основі вищевикладеного метод ідентифікації трафіку в ІТМ полягає в проведенні часо-частотного, статистичного і фрактального аналізів, котрі дозволяють комплексно класифікувати і оцінити характеристики породжуючої системи, і, таким чином, отримати адекватну динамічну модель трафіку.

При цьому методика оцінки характеристик мережевого трафіку влючає наступні етапи.

1. Часо-частотний аналіз:

- якісний аналіз виду часового сигналу і часо-частотних перетворень;
- визначення по АКФ процесу (вираз (9.2)) виду залежності (короткострокова або довготривала);

- побудова і аналіз графіка спектральної щільності потужності (по виразу (9.3));

аналіз виду вейвлет перетворення (оцінка показника Херста H по виразу (9.4)).

2. Статистичний аналіз:

- аналіз дисперсії вибіркового середнього (по виразу (9.5));

- побудова гістограми експериментального розподілу і визначення його адекватності теоретичним розподілам за допомогою критеріїв згоди Колмогорова і Пірсона;

- визначення показника «тяжкості хвоста» α розподілу методом Хілла (по виразу (9.12)) і методом LLCD (по виразу (9.13));

- застосування BDS-теста (згідно виразу (9.15)) для перевірки нульової гіпотези про незалежність і тотожність розподілу значень часового ряду та виявлення нелінійної залежності.

3. Фрактальний аналіз:

- побудова фазового портрета атрактора за дискретним відображенням часової реалізації (9.18);

- обчислення кореляційної ентропії K_C по виразу (9.25), яка є оцінкою знизу K -ентропії Колмогорова (9.24) і характеризує ступінь хаотичності режиму;

- обчислення кореляційного інтервалу прогнозованості (глибини прогнозу) процесу T_C по виразу (9.26), який є оцінкою зверху інтервалу точного прогнозування стану породжуючої системи;

- обчислення кореляційної розмірності атрактора D_C по виразах (9.20), (9.22);

- визначення показника Херста H по виразу (9.23);

- визначення розмірності вкладення атрактора d (розмірності фазового простору – глибини пам'яті) системи по виразу (9.21) і по графіку залежності $D_C(d)$.

Також слід зазначити, що для самоподібного процесу характерними є наступні ознаки:

- ступеневий закон виду графіка спектральної щільності;
- поволі убуваюча дисперсія;
- гіперболічне убування АКФ із зростанням часової затримки;
- розподіл з «важким хвостом»;
- фрактальна розмірність $1 < D < 1,5$;
- показник Херста $0,5 < H < 1$;
- деревоподібна структура скейлетона вейвлет перетворення.

9.2. Моделювання самоподібного трафіку

9.2.1. Моделі мережевого трафіку.

Розглянемо і проаналізуємо відомі моделі мережевого трафіку, що відображають його основні властивості.

Поширення отримали моделі фрактальних рухів. Фрактальний броунівський рух (ФБР) – це процес, що має приріст координати броунівської частинки для будь-якої пари моментів часу t і t_0 виду

$$x(t) - x(t_0) \sim \xi |t - t_0|^H, \quad (t \geq t_0), \quad 0 < H < 1, \quad (9.27)$$

де $x(t_0)$ – координата частинки в деякий початковий момент часу t_0 ; ξ – стандартний гаусовий випадковий процес з незалежними приростами; H – показник Херста. ФБР має нульове середнє, дисперсію приростів, яка збільшується з часом, а також нескінченно великий час кореляції.

Фрактальний гаусовий шум (ФГШ) – це строго самоподібний в широкому сенсі стохастичний процес з нульовим середнім та дисперсією виду

$$D[B_H(t) - B_H(t_0)] = D[\xi] \cdot |t - t_0|^{2H}, \quad (9.28)$$

де B_H – ФБР з показником Херста H і АКФ виду

$$r(k) = \left[|k+1|^{2H} - |2k|^{2H} + |k-1|^{2H} \right], \quad k \in \mathbb{N}, \quad 0 < H < 1. \quad (9.29)$$

Для побудови ФБР застосовують, наприклад, метод, заснований на використанні біортогональних вейвлетів, залежних від заданого базису і параметра Херста H . Основна ідея полягає у використанні дискретного вейвлет-перетворення (ДВП), деталізуючі коефіцієнти на кожному рівні якого є незалежними нормально розподіленими випадковими величинами, а апроксимуючі коефіцієнти – результат використання фрактальної авторегресійної інтегральної моделі з ковзним середнім (FARIMA) процесу:

$$B_H(t) = \sum_{k=-\infty}^{\infty} \Phi_H(t-k) S_k^{(H)} + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} 2^{-jH} \Psi_H(2^j t - k) \varepsilon_{j,k} - b_0, \quad (9.30)$$

де Ψ_H , Φ_H – біортогональний вейвлет та відповідна їй скейлінг функція; $S_k^{(H)}$ – стаціонарний гаусовий процес FARIMA; $\varepsilon_{j,k}$ – незалежні гаусові випадкові величини; b_0 – константа, така що $B_H(0) = 0$.

Перевагами моделей ФБР і ФГШ є властивості самоподоби і довготривалої залежності, що відповідають властивостям експериментальних даних, а також можливість їх аналітичного трактування. Проте, їх недоліки полягають в обмеженнях опису мережевого трафіку і складнощях підбору значень параметрів для генерації трафіку, подібного до експериментальних даних. Крім цього, єдиного параметра, що характеризує кореляційну структуру (асимптотичне убавання) ФБР і ФГШ, недостатньо для моделювання трафіку.

Фрактальний рух Леві є узагальненням ФБР, представником класу α -стійких процесів. Він має властивості самоподоби і нескінченний інтервал кореляції. Прирости даного процесу є самоподібними з $H = 1/\alpha$ і залежними один від одного, а також мають розподіли з «важкими хвостами» [6]. Відповідно, модель враховує два показники: Херста $H \in [1/2, 1]$ і Леві $\alpha \in [0, 1]$. Позитивами даної моделі є можливість її математичного опису, а також здатність враховувати розподіл з «важким хвостом». Серед недоліків варто відзначити необхідність враховувати декілька параметрів, що визначають стан

моделі (зокрема, для визначення показника α відсутній прямий метод оцінки).

Переваги моделей фрактальних рухів: концептуально прості; можуть описати складну динаміку багатьох реальних систем; відносно мале число параметрів і їх вибір наочний; дозволяють виявити стійкі стани системи.

Авторегресійні моделі припускають, що поточне значення процесу є сумою постійної, зваженої суми попередніх значень і похибки моделі. Прикладами таких моделей є: ARMA (авторегресійна модель ковзного середнього), ARIMA (авторегресійна інтегральна модель ковзного середнього) і FARIMA.

При оцінюванні та ідентифікації процесів для реалізації авторегресійних моделей використовують також АФА, процес адаптації яких включає оцінювання шуканого виходу фільтра та корегування його параметрів в залежності від значення вихідної похибки.

Вище (див. п. 8.1) докладно розглянуті лінійні, нейромережеві, нейронечіткі, нечіткі та нейровейвлетні АФА, які можуть бути використані при моделюванні процесів в ІТМ.

9.2.2. Приклад аналізу та моделювання мережевого трафіку.

Відповідно до методики, розглянутої в п. 9.1.4, 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 мережу 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 моделювання наведені на рис. 9.1-9.6 та в таблицях 9.1 і 9.2.

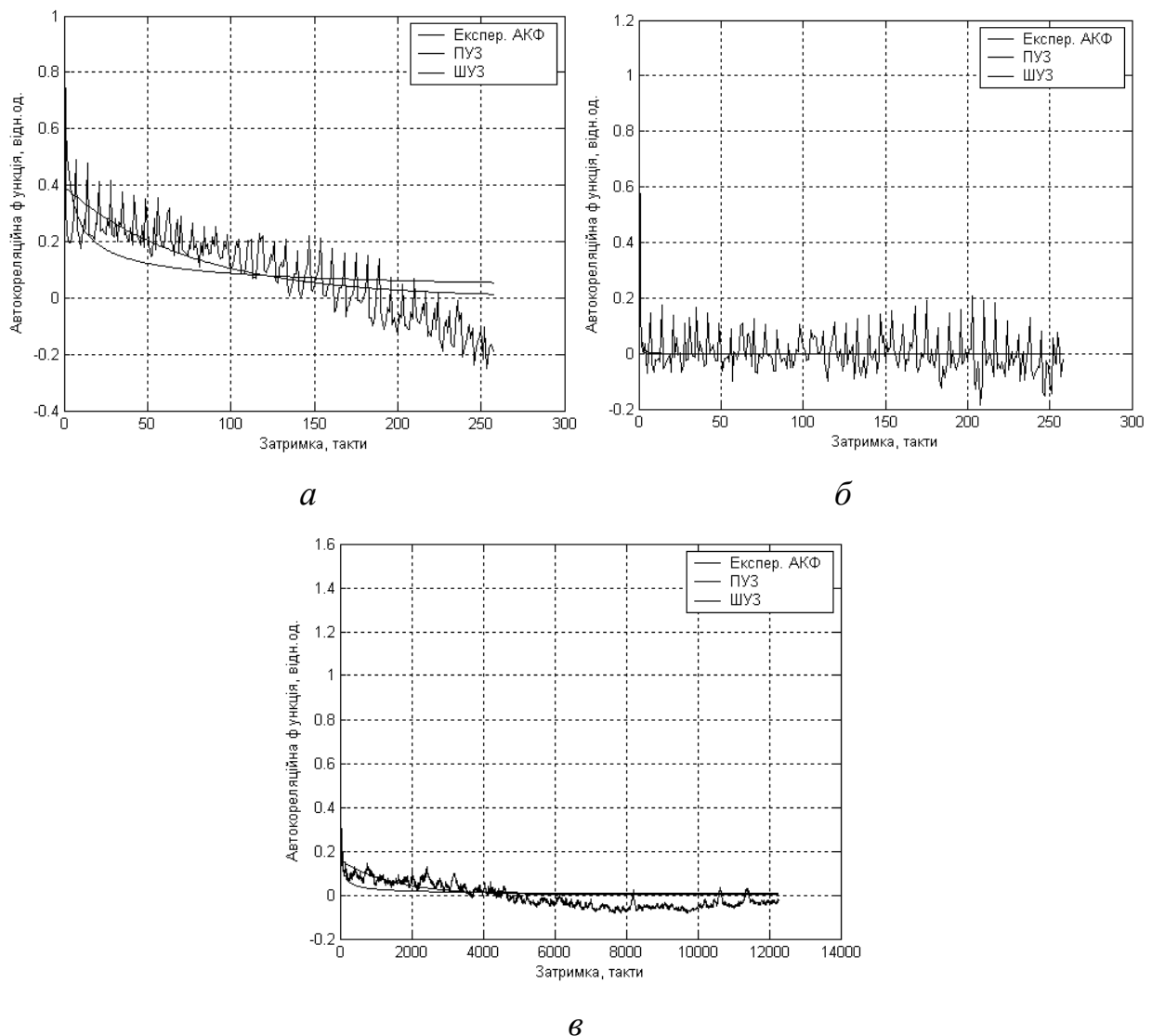


Рис. 9.1. АКФ експериментальних сигналів 1 (*a*), 2 (*б*) та 3 (*в*)

За видом АКФ (рис. 9.1,*a*) і спектру (рис. 9.3,*a*) експериментального сигналу 1 можна сказати про нерегулярність породжуючого процесу. Це

впливає з того, що його спектр має вид спектру фліккер-шуму, тобто спектральна потужність сигналу 1 приблизно зворотньопропорційна 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) убуваючих залежностей наведені на рис. 9.1,*a*).

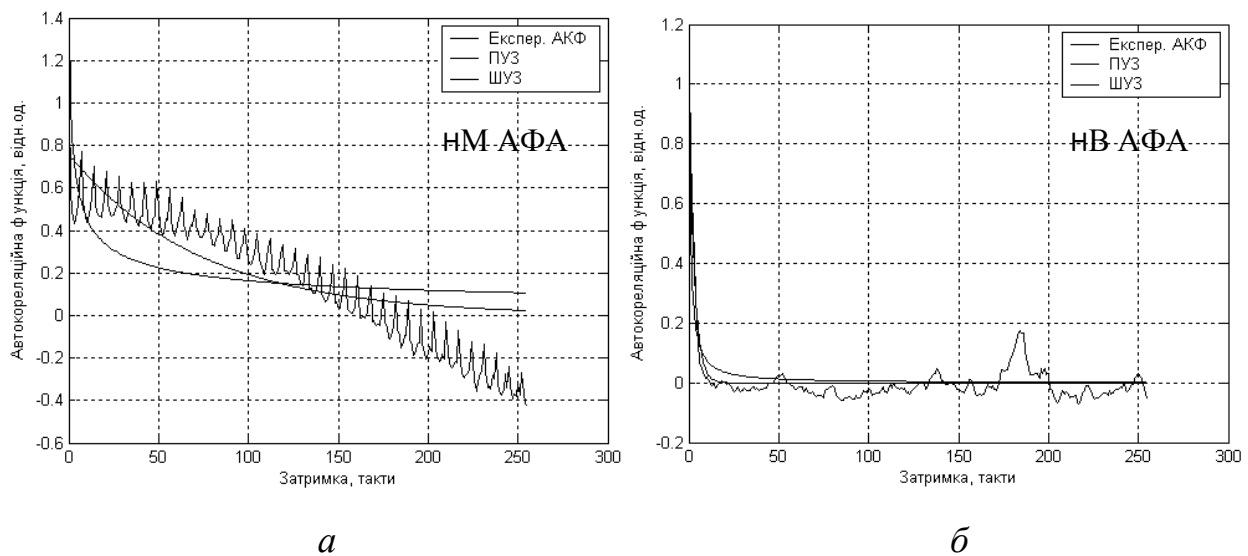


Рис. 9.2. АКФ модельних сигналів 1 (*a*) і 3 (*б*)

Значення показника Херста (табл. 9.1) сигналу 1 мають в залежності від методів, що використовувались, значення в діапазоні $[0,814; 0,923]$. Тобто це персистентні ряди: якщо вони зростають (убувають) 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. 9.5,*a*), а саме їх повільне убування, 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. 9.1).

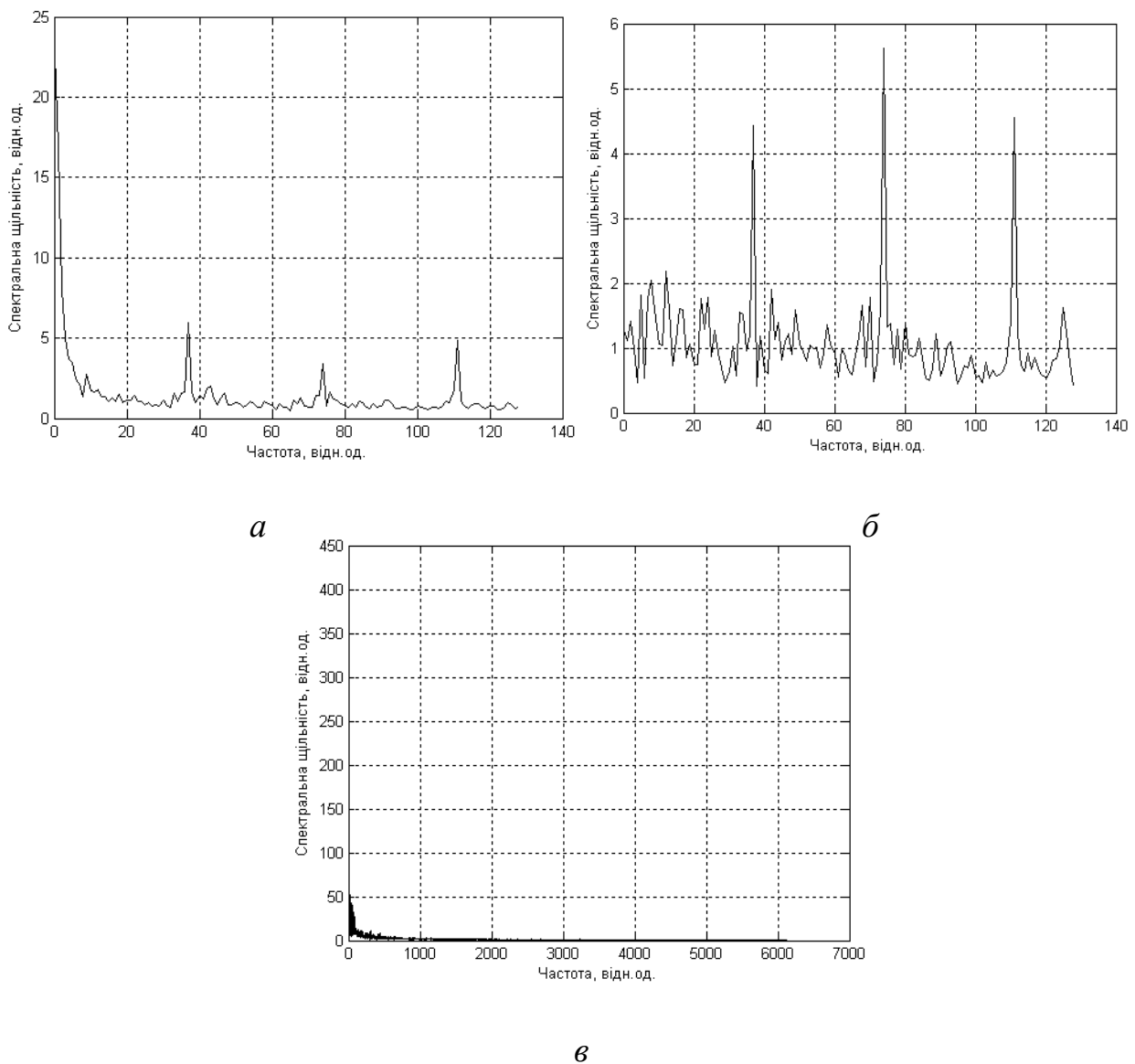
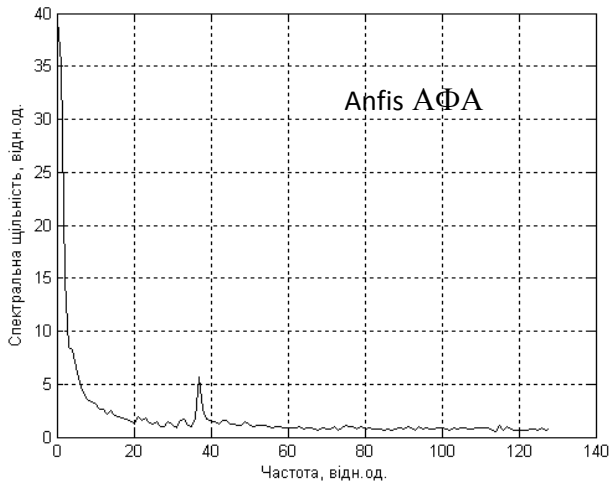
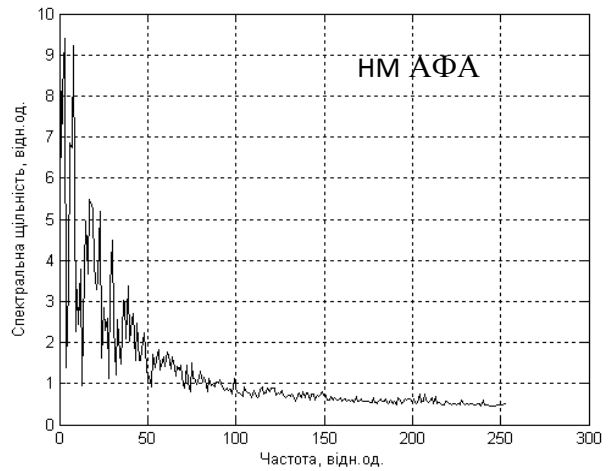


Рис. 9.3. Спектри експериментальних сигналів 1 (*a*), 2 (*б*) та 3 (*в*)

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.

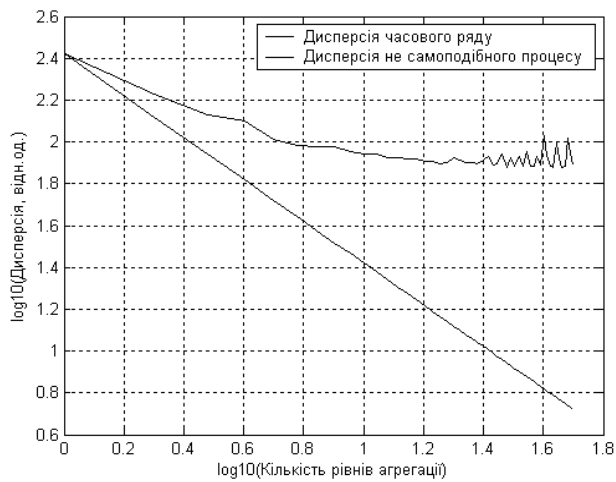


a

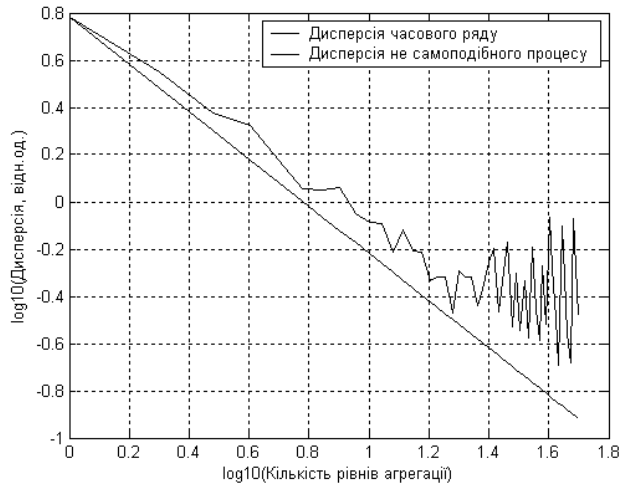


б

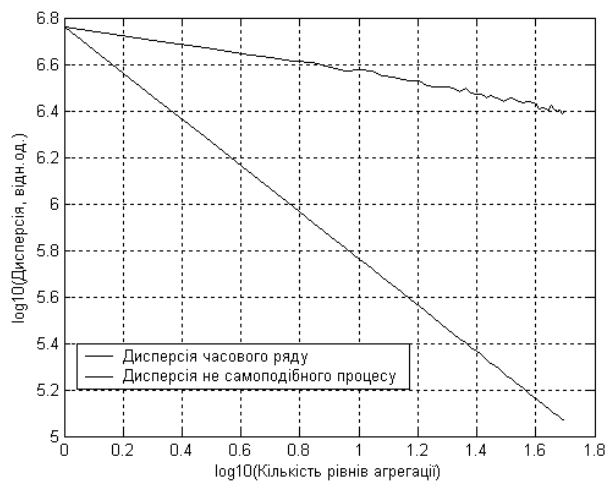
Рис. 9.4. Спектри модельних сигналів 1 (*a*) і 3 (*б*)



a



б



в

Рис. 9.5. Дисперсії експериментальних сигналів 1 (*a*), 2 (*б*) та 3 (*в*)

Характеристики експериментальних сигналів

Таблиця 9.1

Показник	Сигнал 1	Сигнал 2	Сигнал 3
АКФ	ШУЗ		ПУЗ
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 - метод агрегованої дисперсії - R/S аналіз	0,923 0,904 0,814	0,909 0,668 0,671	0,980 0,888 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	$4 \leq d \leq 6$	$5 \leq d \leq 8$	$2 \leq d \leq 4$

Швидкий спад АКФ експериментального сигналу 2 (рис. 9.1,б) свідчить про стохастичність процесу, що його породжує. Енергія спектру сигналу (рис.

9.3,б) представлена як в області низьких, так і високих частот. Значення показника Херста (табл. 9.1) змінюється в діапазоні $[0,814; 0,923]$. Графік дисперсії сигналу 2 (рис. 9.5,б) свідчить про відсутність властивості самоподоби. Підтвердженням стохастичності та нелінійності процесу є результат BDS-тесту.

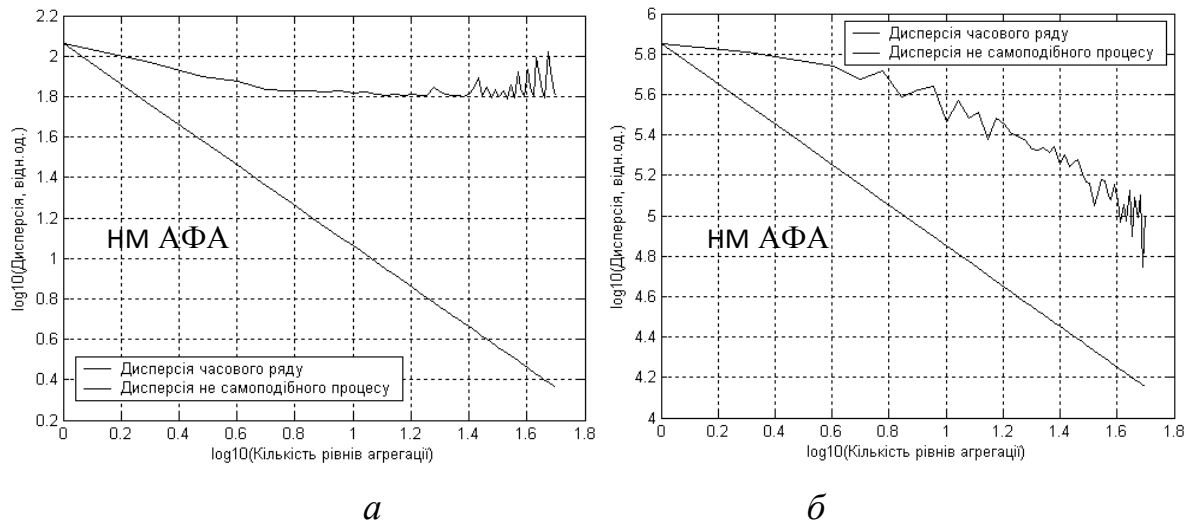


Рис. 9.6. Дисперсії модельних сигналів 1 (а) і 3 (б)

Породжуючий процес експериментального сигналу 3 має всі властивості, притаманні самоподібним процесам: гіперболічне убування АКФ із зростанням часової затримки (рис. 9.1,в), ступеневий закон графіка його спектральної щільності (рис. 9.3,в) та поволі убуваючу дисперсію (рис. 9.5,в). Сигнал 3 має (табл. 9.1) 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.

Показник	Сигнал 1			Сигнал 3	
	Anfis АФА	НМ АФА	НВ АФА	НМ АФА	НВ АФА
АКФ	ШУЗ		ПУЗ	ШУЗ	
Спектр	Ступеневий закон із сплесками			Ступеневий закон	
Вейвлет-перетворення	Деревоподібна структура скейлетона				
Дисперсія	Поволі убуваюча				
Закон розподілу за критерієм: - Колмогорова - Пірсона	Логнормальний Логнормальний		Логнормальний Гамма-розподіл	Логнормальний Логнормальний	Парето Експоненціальний
Показник «важкості хвоста»	~6,5	~6,1	~9,4	~1,2	~1,5
BDS-тест	Хаотичний нелінійний процес				
Показник Херста: - періодограмний метод - метод агрегованої дисперсії - R/S аналіз	0,999 0,967 0,837	0,997 0,979 0,992	1,013 1,039 0,849	0,936 0,687 0,788	0,974 1,454 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. 9.1, 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. 9.2), спектри (рис. 9.4) та дисперсії (рис. 9.6) модельних сигналів якісно відповідають графікам експериментальних сигналів (рис. 9.1, 9.3, 9.5).

Крім того, якісні та числові значення характеристик модельних сигналів в цілому відповідають характеристикам експериментальних сигналів (див. табл. 9.1 і 9.2).

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.

9.3. Мультифрактальний аналіз трафіку

Агрегований трафік від різних джерел на малих часових масштабах проявляє мультифрактальний характер, тобто має гнучкий закон масштабування.

Припустимо, що процес займає обмежену область розмірністю d у евклідовому просторі і є множиною точок $N \rightarrow \infty$, розташованих в цій області. Тоді розбивши її на кубічні осередки об'ємом ε^d і відібравши осередки, що

містять хоч одну точку, отримаємо загальну кількість зайнятих осередків $N(\varepsilon)$, номер яких змінюється на інтервалі $i \in [1, N(\varepsilon)]$. При цьому узагальнена статистична сума (часткова функція), визначається як

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon), \quad (9.31)$$

де p_i – вірогідність заселення осередків $p_i(\varepsilon) = \lim_{N \rightarrow \infty} \frac{n_i(\varepsilon)}{N}$, $n_i(\varepsilon)$ – кількість точок в i -ому осередку, q – коефіцієнт ($q \in \mathbb{R}$).

Залежність узагальненої статистичної функції від сторони осередку описується як

$$Z(q, \varepsilon) \approx \varepsilon^{\tau_q}, \quad (9.32)$$

де τ_q – скейлінгова експонента (масштабна функція), що визначається як

$$\tau_q = \lim_{\varepsilon \rightarrow 0} \frac{\ln Z(q, \varepsilon)}{\ln \varepsilon}, \quad (9.33)$$

$$\tau_q = (q - 1)D_q, \quad (9.34)$$

де D_q – спектр узагальнених фрактальних розмірностей Рені, який характеризує розподіл точок в деякій області і показує наскільки неоднорідною є досліджувана множина точок.

Якщо D_q залежить від q , то даною множиною точок є мультифрактал (неоднорідна фрактальна множина) з нелінійною функцією τ_q .

Для характеристики мультифракталів також використовують функцію мультифрактального спектру (спектр сингулярностей) $f(\chi)$.

Якщо $\chi = 1/t$, то початкові моменти процесу, якому притаманна властивість самоподоби, можна визначити як

$$M \left[|x(t)|^q \right] = M \left[|t^H x(t)|^q \right] = t^{qH} M \left[|x(1)|^q \right] = c(q) \cdot t^{qH}, \quad (9.35)$$

де $c(q)$ – коефіцієнт моменту, $c(q) = M \left[|x(1)|^q \right]$, M – математичне очікування; x – процес; t – час, H – показник Херста.

Мультифрактальний процес має стаціонарні прирости і задовольняє умові

$$M[|x(t)|^q] = c(q) \cdot t^{qh(q)} = c(q) \cdot t^{\tau(q)+1}, \quad (9.36)$$

де $h(q)$ – узагальнений показник Херста (показник Гельдера). У разі монофрактальних процесів $h(q) = H$.

Виявити мультифрактальну особливість мережевого трафіку можна шляхом розрахунку масштабної функції $\tau(q)$ і моментного коефіцієнта $c(q)$.

9.3.1. Мультифрактальні моделі.

Такі моделі ґрунтуються на використанні мультиплікативних каскадів, що застосовують ітеративну процедуру.

Спочатку одиничний інтервал ділиться на два рівні підінтервали, вагові коефіцієнти яких r_1 і $r_2 = 1 - r_1$ відповідно. Для кожного підінтервалу процедура повторюється. Вагові коефіцієнти є незалежними, випадковими і однаково розподіленими величинами $R \in [0,1]$ з функцією розподілу вірогідності $F_R(x)$, $M[R] = 1/2$. На k -ому кроці ітеративної процедури міра $\mu(\Delta t_k)$ підінтервалу довжини $\Delta t_k = 2^{-k}$ задовольняє співвідношенню масштабування:

$$M[\mu(\Delta t_k)^q] = (M[R^q])^k = \Delta t_k^{-\log M[R^q]}, \quad (9.37)$$

яке визначає мультифрактальний процес з функцією масштабування:

$$\tau_0(q) = -\log_2 M[R^q]. \quad (9.38)$$

Завдання каскадного моделювання полягає в знаходженні розподілу вірогідності для множників R такого, щоб виконувалась рівність (9.38).

Не менш поширеними є мультифрактальні вейвлет моделі (МВМ), що дозволяють моделювати позитивні часові ряди заданої довжини. В основі їх функціонування лежить метод вейвлет перетворення, а часовий ряд представляється у вигляді:

$$x(t) = \sum_k U_{J_0,k} \phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty} \sum_k W_{j,k} \psi_{j,k}(t), \quad (9.39)$$

де $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$, $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$ – вейвлет та відповідна йому масштабуюча функція; $W_{j,k} = \int x(t)\psi_{j,k}(t)dt$ – деталізуючі коефіцієнти; $U_{j,k} = \int x(t)\phi_{j,k}(t)dt$ – апроксимуючі (скейлінг) коефіцієнти; j – кількість рівнів вейвлет-розкладання; $j = J_0$ – найменша роздільна здатність; k показує просторове розташування аналізу.

Враховуючи, що на практиці часто розглядається дискретний сигнал $x^{(n)}(k)$ з кінцевою тривалістю, що апроксимує сигнал $x(t)$ з роздільною здатністю 2^{-n} , то напівнескінченну суму в (9.39) необхідно замінити на суму з кінцевим числом масштабів, тобто $0 \leq j < n$:

$$U_{j-1,k} = 2^{-1/2}(U_{j,2k} + U_{j,2k+1}), \quad W_{j-1,k} = 2^{-1/2}(U_{j,2k} - U_{j,2k+1}), \quad W_{j,k} = A_{j,k} U_{j,k}, \quad (9.40)$$

де $A_{j,k}$ – незалежні, випадкові, симетричні відносно нуля змінні $A_{j,k} \in [-1,1]$.

Для простоти і гнучкості генерування множників $A_{j,k}$ використовується симетричний бета-розподіл $\beta(p,p)$ та симетричний розподіл точкових мас.

Тоді отримаємо

$$U_{j,2k} = 2^{-1/2}(1 + A_{j+1,k})U_{j-1,k}, \quad U_{j,2k+1} = 2^{-1/2}(1 - A_{j+1,k})U_{j-1,k}. \quad (9.41)$$

Скейлінг-коефіцієнти, обчислені при найдетальнішій роздільній здатності, є вихідними значеннями процесу, тобто

$$x^{(n)}[k] = 2^{-n/2} U_{n,k}, \quad k = 0, \dots, 2^n - 1. \quad (9.42)$$

Мультифрактальні моделі забезпечують гнучкіший закон масштабної поведінки і вдало відтворюють трафік, агрегований від декількох джерел, що істотно відрізняються; аналітично прості. Проте вони оцінюють тільки верхню огинаючу мультифрактального спектру, що може призводити до помилкових результатів.

9.3.2. Мультифрактальний аналіз детрендових флуктуацій.

Для аналізу властивостей самоподоби і довготривалої залежності нестационарних часових рядів застосовують мультифрактальний аналіз детрендових флуктуацій (АДФ), який є різновидом дисперсійного аналізу. Цей метод застосовується для визначення масштабної поведінки часового ряду у присутності можливих трендів без знання їх походження і форми.

Інтегрований часовий ряд має вигляд

$$y(k) = \sum_{i=1}^k (x(i) - \bar{x}), \quad k = 1, \dots, N, \quad (9.43)$$

де $x(i)$ – часовий ряд, N – довжина ряду, \bar{x} – середнє значення ряду.

Ряд $y(k)$ розбивається на m підпоследовностей довжини n ($nm = N$), для кожного з яких методом найменших квадратів обчислюється локальний тренд $y_n(k)$ – рівняння прямої, апроксимуючої послідовність $y(k)$. Середньоквадратична похибка такої апроксимації визначається як

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (9.44)$$

Відповідні розрахунки повторюються для відрізків різної довжини. Прояв наявності самоподоби свідчить ступеневий характер залежності $F(n) \sim n^\alpha$ і лінійний характер в подвійному логарифмічному масштабі $\log F(\log n)$. При цьому показник скейлінга α визначається як кут нахилу прямої, що апроксимує графік $\log F(\log n)$. Залежно від значення даного показника можна класифікувати часовий ряд. Так, для випадкового процесу (некорельована поведінка) $\alpha = 0,5$; $0 < \alpha < 0,5$ відповідає антиперсистентному ряду; $0,5 < \alpha < 1$ визначає персистентні довготривалі кореляції; $\alpha = 1$ характерне для $1/f$ -шуму. При $\alpha > 1$ кореляції існують, проте ступеневу залежність вони не відображають; $\alpha = 1,5$ характерне для броунівського шуму.

При мультифрактальному АДФ обчислюється кумулятивний ряд $y(k)$ згідно (13), який потім розділяється на $N_s = N/s$ непересічних відрізків довжини s . Проте оскільки довжина ряду N не завжди кратна шкалі s , то для

обліку останнього відрізка, число елементів якого менше s , процедура розподілу повторюється, починаючи з протилежного кінця ряду. Таким чином, отримуємо $2N_s$ відрізків $v = 1, \dots, N_s, N_s + 1, \dots, 2N_s$ завдовжки s . При реалізації даного алгоритму прийнято обмежувати довжину сегментів: $10 < s < N / 4$.

Методом найменших квадратів для кожного з $2N_s$ сегментів визначається локальний тренд $y_v(k)$ та обчислюється дисперсія для кожного сегменту $v = 1, \dots, N_s$:

$$F^2(v, s) = \frac{1}{s} \sum_{k=1}^s \left\{ y[(v-1)s + k] - y_v(k) \right\}^2 \quad (9.45)$$

і для кожного $v = N_s + 1, \dots, 2N_s$:

$$F^2(v, s) = \frac{1}{s} \sum_{k=1}^s \left\{ y[N - (v - N_s)s + k] - y_v(k) \right\}^2. \quad (9.46)$$

Функції флуктуацій q -го порядку визначається усередненням дисперсій (9.45), (9.46):

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(v, s)]^{q/2} \right\}^{1/q}. \quad (9.47)$$

При $q = 0$ замість (9.17) використовується вираз

$$F_0(s) = \exp \left\{ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln [F^2(v, s)] \right\}. \quad (9.48)$$

Якщо залежність $F_q(s)$ від s у подвійному логарифмічному масштабі для фіксованому значенні q описується прямою лінією на графіку і має ступеневий характер

$$F_q(s) \sim s^{h(q)}, \quad (9.49)$$

то даний часовий ряд зводиться до самоподібної множини, що проявляє довготривалі кореляції.

З виразів (9.47), (9.49) випливає, що при $q = 2$ узагальнений показник Херста $h(q) = H$. Для монофрактальних часових рядів узагальнений показник

Херста не залежить від q , а флуктуаційна функція $F^2(v,s)$ однакова для всіх відрізків v .

Для мультифрактальних рядів $h(q)$ – нелінійна функція. При позитивних значеннях q основний внесок у функцію $F_q(s)$ дають сегменти з великими відхиленнями $F^2(v,s)$, а при негативних – з малими. Таким чином, при негативних значеннях q узагальнений показник Херста $h(q)$ описує поведінку сегментів, що проявляють малі флуктуації, а при позитивних – великі.

9.3.3. Метод максимумів модулів вейвлет-перетворення.

Даний метод знайшов широке застосування при аналізі мультифрактальних властивостей нестационарних процесів. Метод максимумів модулів вейвлет-перетворення (ММВП) зводиться до виконання двох етапів. На першому етапі обчислюється вейвлет перетворення часового ряду $x(i)$ довжини N :

$$W(n,s) = \frac{1}{s} \sum_{i=1}^N x(i) \psi \left[\frac{i-n}{s} \right], \quad (9.50)$$

де s – параметр масштабу, n – момент часу (параметр зсуву), $\psi(x)$ – вейвлет, за допомогою якого виключаються з часового ряду тренди, які можуть бути апроксимовані многочленами до $(m-1)$ -го порядку включно.

За наявності локальної сингулярності в даних вейвлет коефіцієнти визначаються відповідно до ступеневого закону:

$$W(n_0,s) \sim s^{h(n_0)}. \quad (9.51)$$

Узагальнений показник Херста $h(n_0)$ можна визначити, побудувавши графік залежності $W(n_0,s)$ від s у подвійному логарифмічному масштабі і обчисливши нахил прямої. Однією з особливостей даного методу є те, що $h(n_0)$ не залежить від вибору вейвлета ψ .

Другий етап полягає в обчисленні часткової функції:

$$Z(q,s) = \sum_{i=1}^{i_{\max}} |W(n_i, s)|^q. \quad (9.52)$$

При $s \rightarrow 0$ часткова функція описується ступеневим законом виду:

$$Z(q,s) \sim s^{\tau(q)}. \quad (9.53)$$

Оскільки в методі ММВП розглядаються нерегулярні сигнали, то замість спектру сингулярностей $f(\alpha)$ використовується аналогічна функція $D(h)$, для чого застосовується перетворення Лежандра:

$$\begin{cases} h = d\tau / dq; \\ D(h) = qh - \tau(q). \end{cases} \quad (9.54)$$

Для монофрактальних об'єктів залежність $\tau(q)$ є лінійною, а для мультифрактальних – нелінійною з великим числом узагальнених показників Херста $h(q) \neq \text{const}$.

9.3.4. Приклад аналізу експериментальних даних та моделювання.

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 2).

Система EWSD надає абонентам розширений набір послуг зв'язку: інтегральне обслуговування, обмін даними і обмін голосовою інформацією, крім того вона є зоною. Час агрегації (дискретизації) для сигналу 1 складає 1 год.

Дані реалізації трафіку мережі Інтернет представляють 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.

Для моделювання мережевого трафіку використовувались мультифрактальні вейвлет моделі 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. 9.7-9.9 та в табл. 9.3.

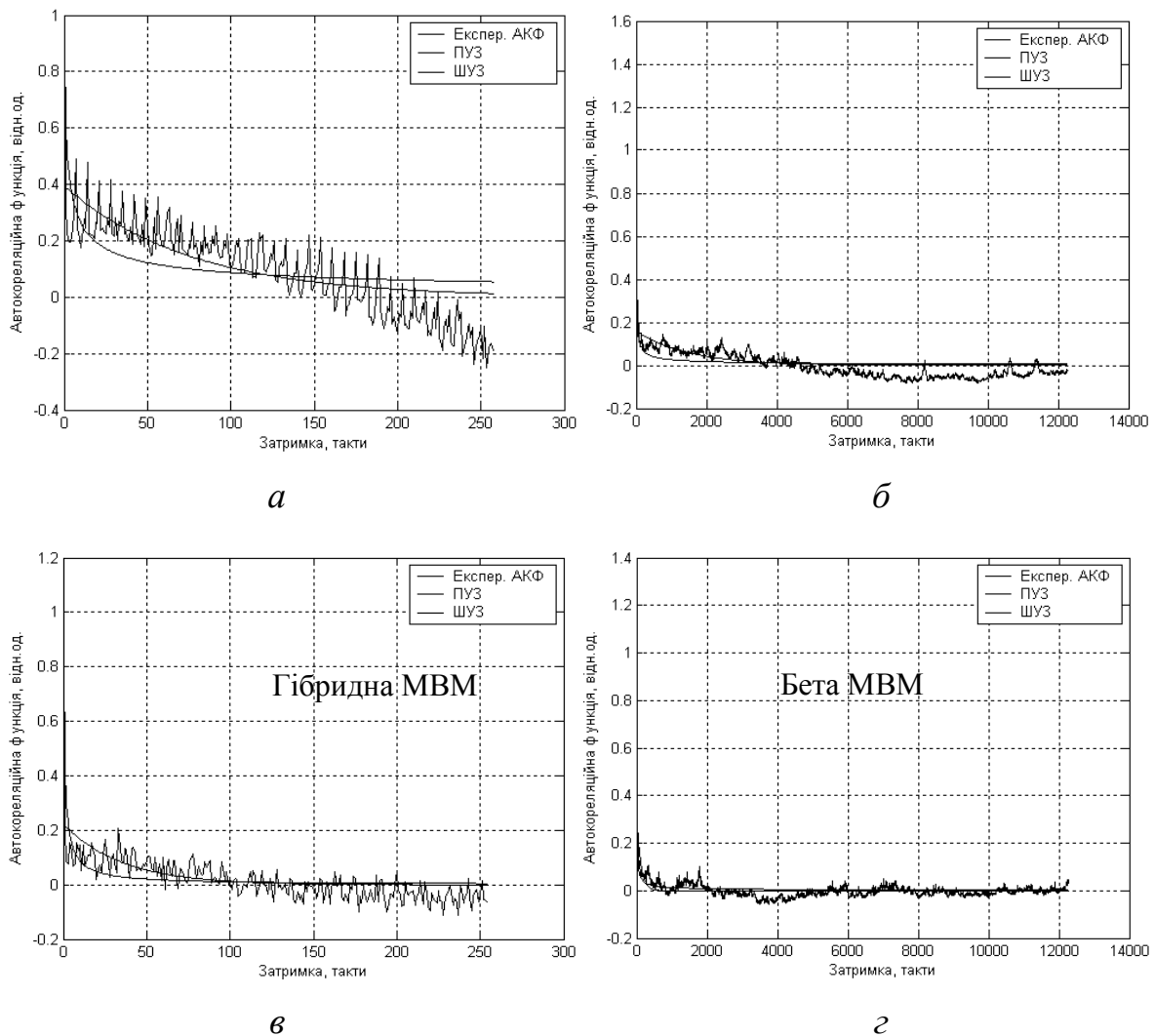


Рис. 9.7. АКФ експериментальних сигналів 1 (*a*) і 2 (*б*) та модельних сигналів 1 (*в*) і 2 (*г*)

За видом АКФ (рис. 9.7, *a* і *б*), спектрів (рис. 9.8, *a* і *б*) та дисперсій (рис. 9.9, *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 поволі (ПУЗ) та швидко (ШУЗ) убуваючих залежностей).

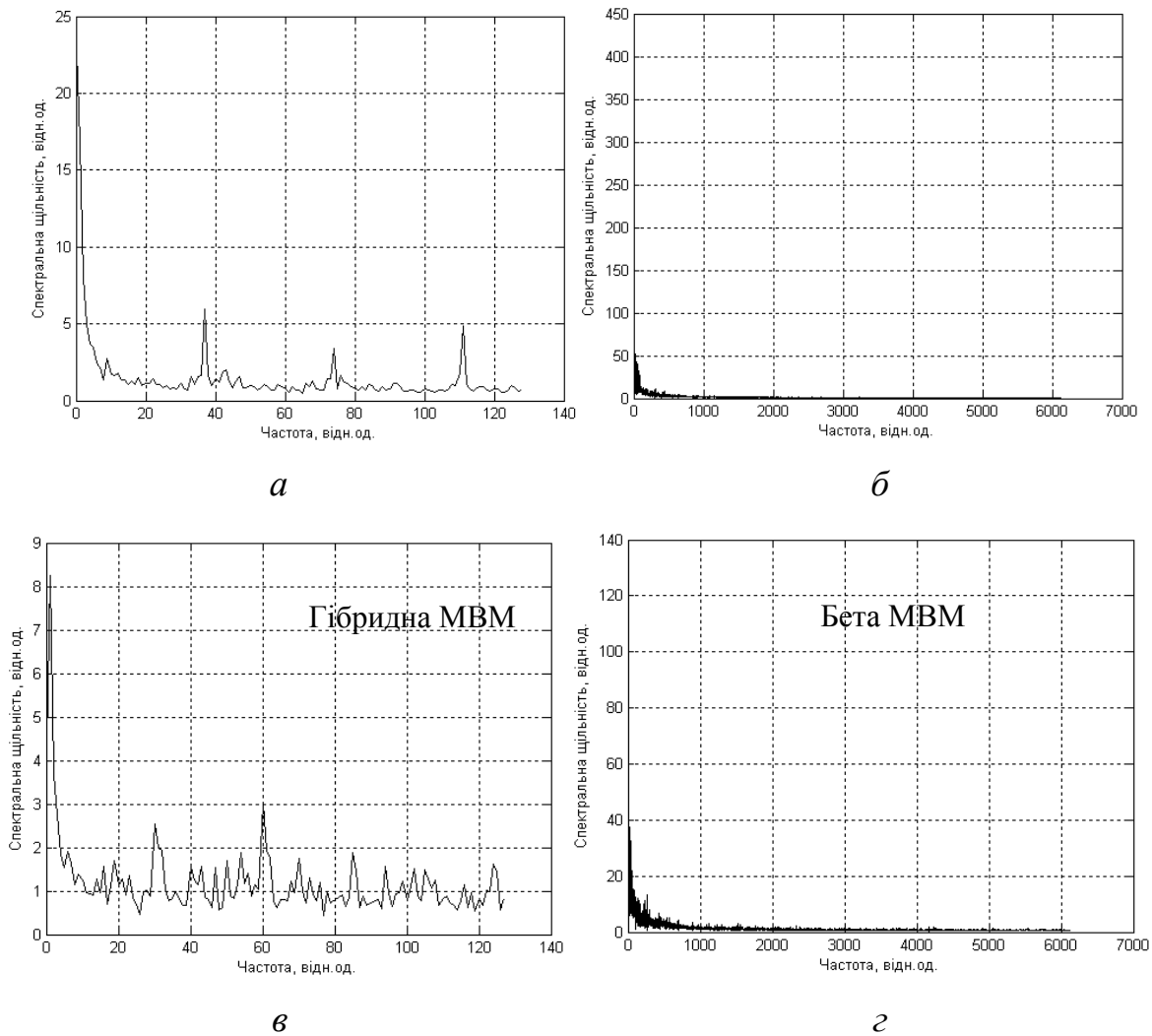


Рис.9.8. Спектри експериментальних сигналів 1 (а) і 2 (б)
та модельних сигналів 1 (в) і 2 (г)

Моделювання сигналів трафіку мультифрактальними вейвлет моделями підтверджує їх якісні оцінки (рис. 9.7, в і г; рис. 9.8, в і г; рис. 9.9, 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. 9.3), що експериментальні і модельні сигнали 1 і 2 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 %).

Числові значення характеристик модельних сигналів в цілому відповідають характеристикам експериментальних сигналів, при цьому бета МВМ мають меншу похибку визначення характеристик, ніж гібридні МВМ.

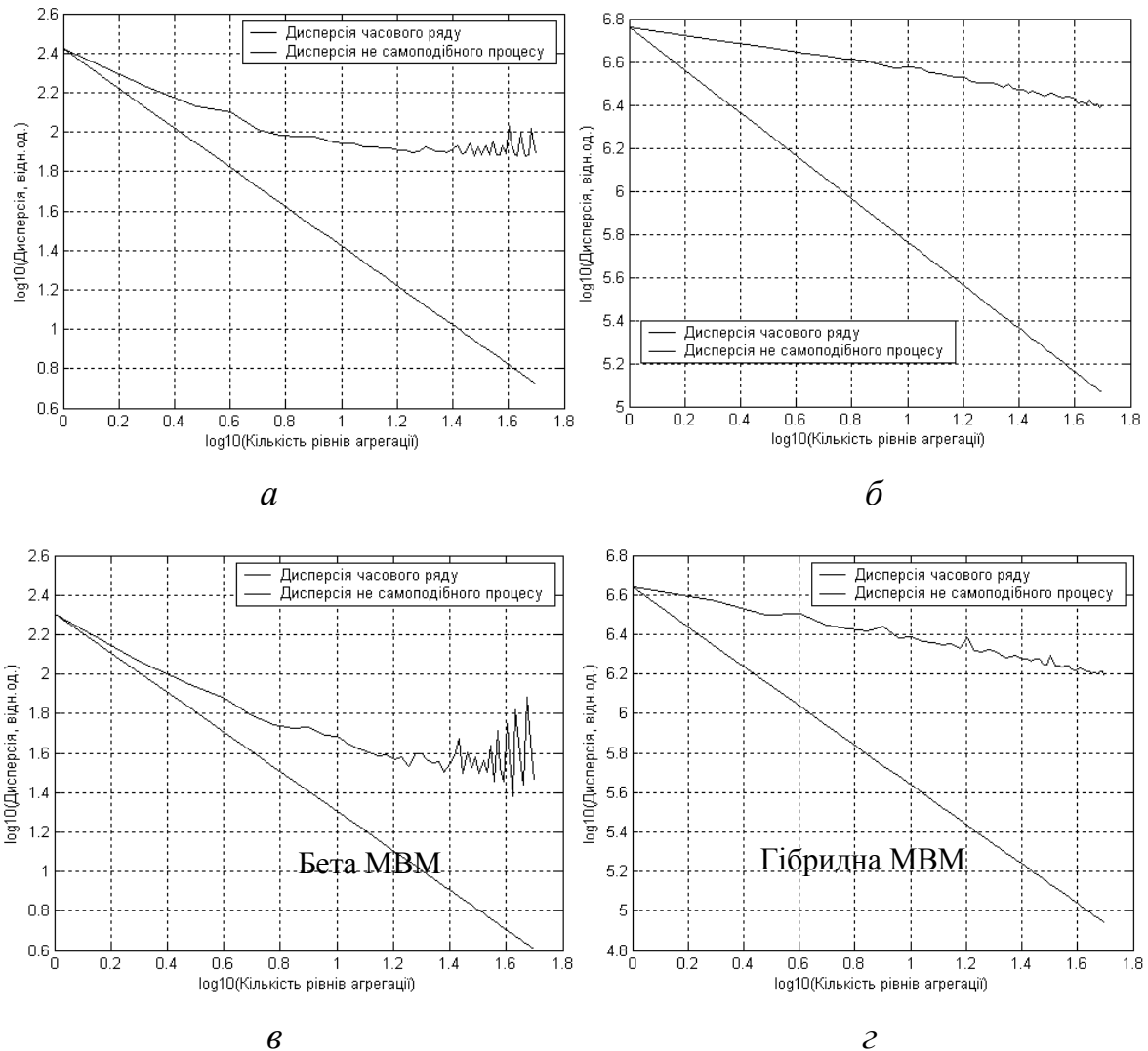


Рис. 9.9. Дисперсії експериментальних сигналів 1 (*a*) і 2 (*б*) та модельних сигналів 1 (*в*) і 2 (*г*)

Проведений аналіз детрендових флуктуацій дозволив отримати наступні результати. Поведінка сигналу 1 (рис. 9.10, *a*) на інтервалі від 10 до 86 тактів некорельована, оскільки показник скейлінга $\alpha_1 = 0,5$.

Показник	Сигнал 1			Сигнал 2		
	Експериментальний	Бета МВМ	Гібридна МВМ	Експериментальний	Бета МВМ	Гібридна МВМ
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						
- періодограмний метод	0,923	0,891	0,936	0,980	0,940	0,884
- метод агрегованої дисперсії	0,904	0,843	0,842	0,888	0,843	0,869
- R/S аналіз	0,814	0,910	0,776	0,913	0,858	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 2 значення $\alpha = 0,98$ близьке до 1 (рис. 9.10, б), що властиво для 1/f-шума (флікер-шума), для якого енергія спектра зворотно пропорційна частоті.

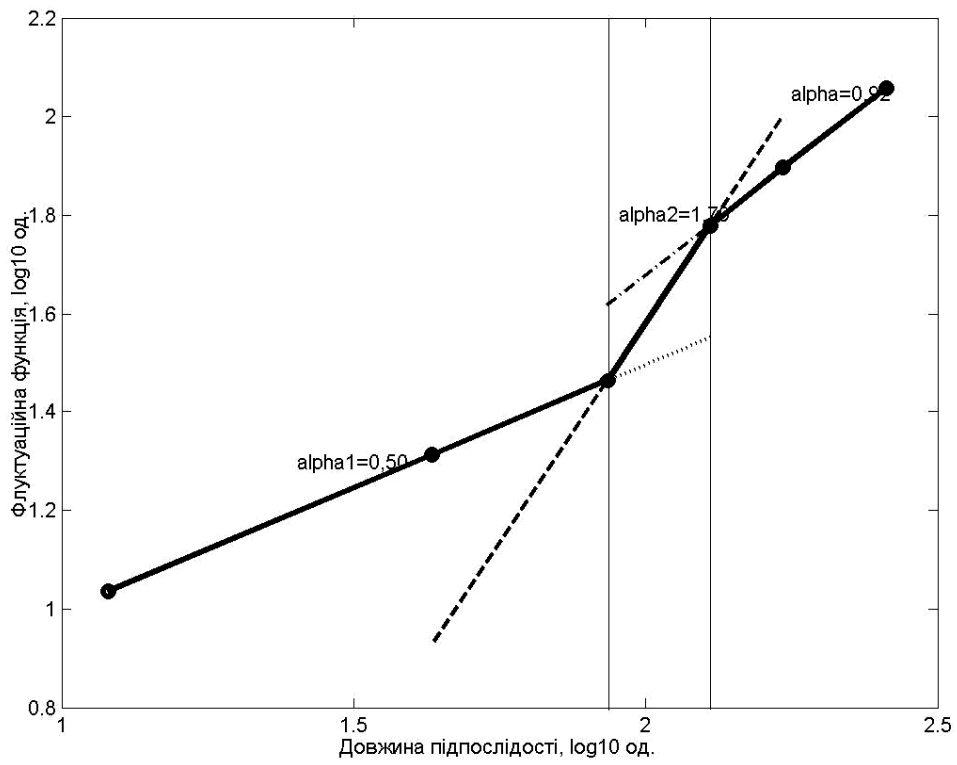
Слід відмітити, що для експериментальних сигналів отримані значення показника скейлінга α співпадають зі значеннями показника Херста (табл. 9.3), отриманими періодограмним методом.

Результати мультифрактального АДФ наведені на рис. 9.11. При обмеженні значень порядку $-5 \leq q \leq 5$ на масштабах $s < 2^5 = 32$ трафіки (сигнали 1 та 2) можна розглядати як монофрактальні процеси.

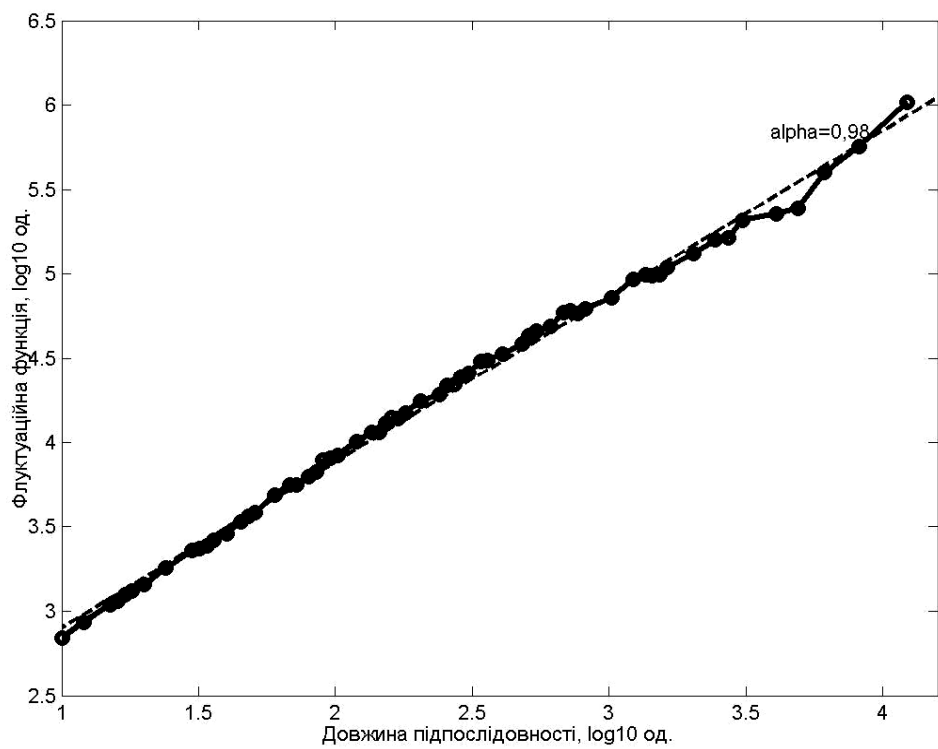
В результаті проведеного дослідження сигналів з використанням ММВП були отримані наступні результати. Графіки часткових функцій $Z(q,s)$ (рис. 9.12) показують, що при варіації значення їх порядку $-5 < q < 5$ на масштабі $s \leq 2^3 = 8$ експериментальні сигнали виявляють монофрактальні властивості.

Графіки скейлінгових експонент $\tau(q)$, показників Гельдера (узагальнених показників Херста) $h(q)$ та спектрів сингулярності $D(h)$ експериментальних сигналів наведені на рис. 9.13-9.15, 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.

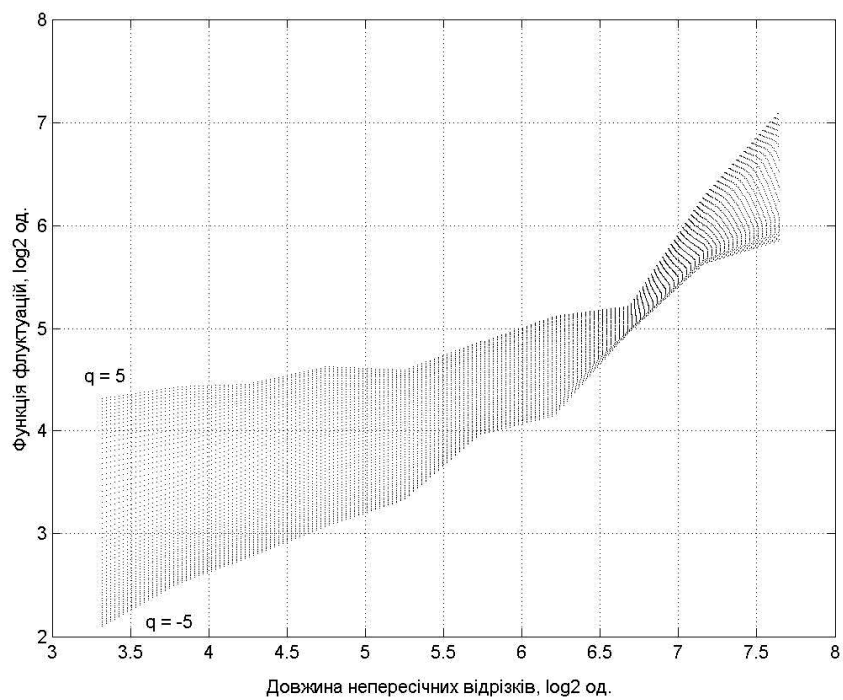


a

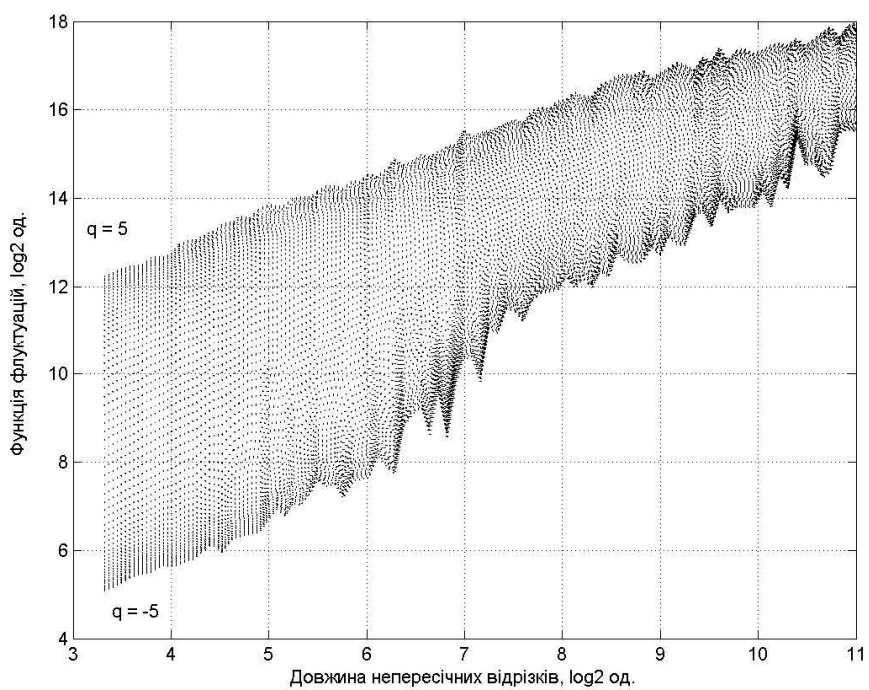


б

Рис. 9.10. Флуктуаційні функції експериментальних сигналів 1 (*a*) і 2 (*б*) за АДФ

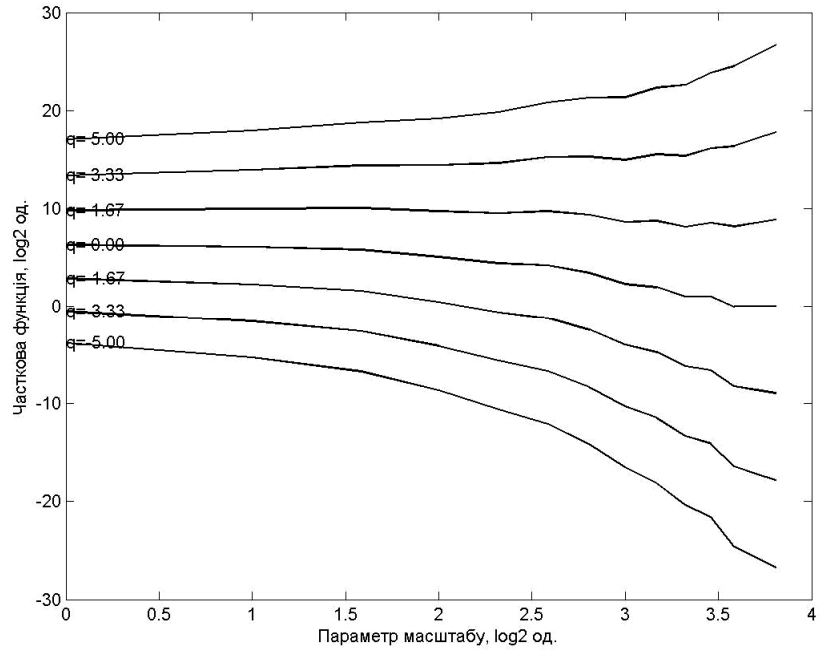


a

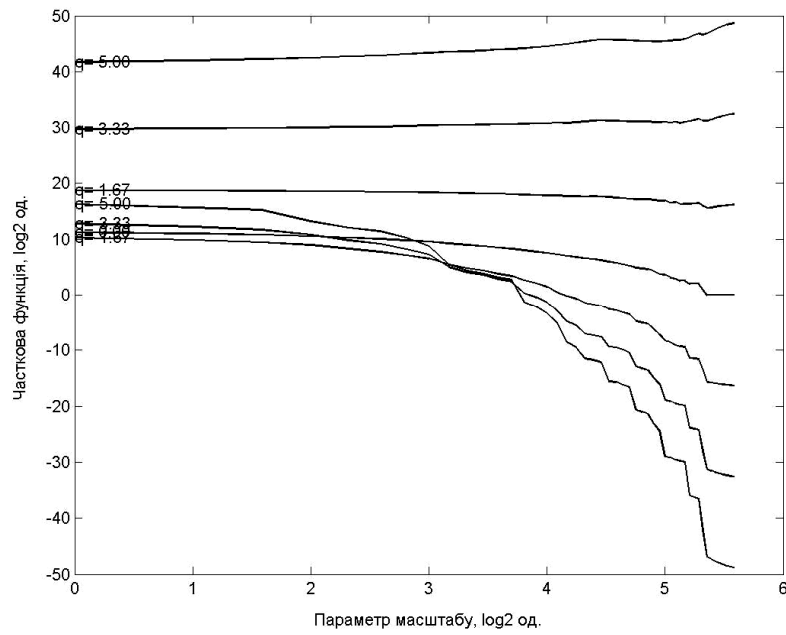


б

Рис. 9.11. Флуктуаційні функції експериментальних сигналів 1 (*a*) і 2 (*б*) за мультифрактальним АДФ



a



б

Рис. 9.12. Часткові функції експериментальних сигналів 1 (*a*) і 2 (*б*)

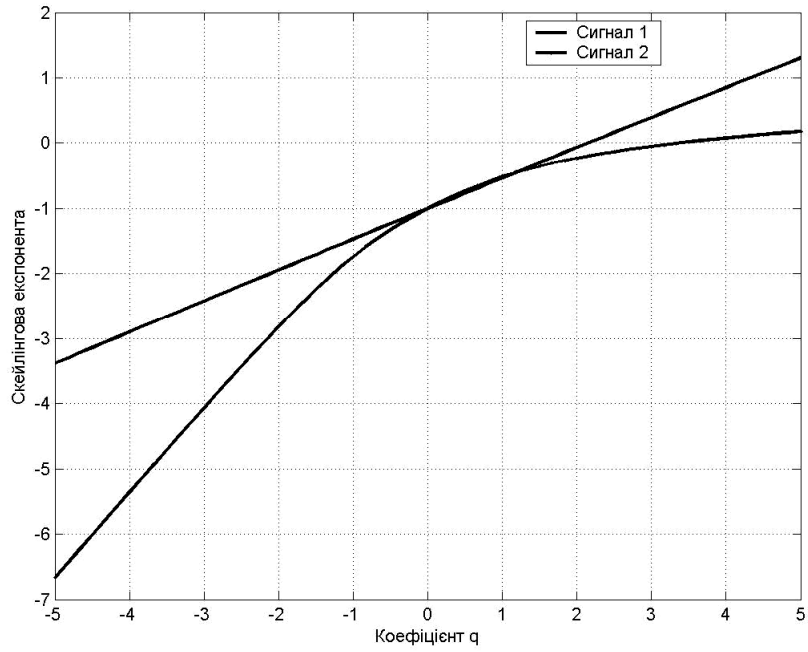


Рис. 9.13. Скейлінгові експоненти експериментальних сигналів

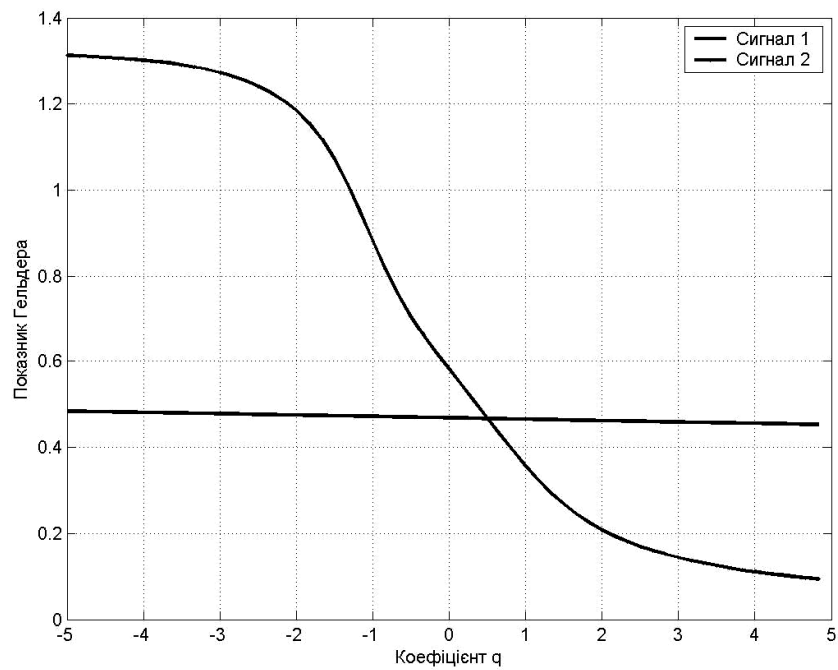


Рис. 9.14. Показники Гельдера експериментальних сигналів

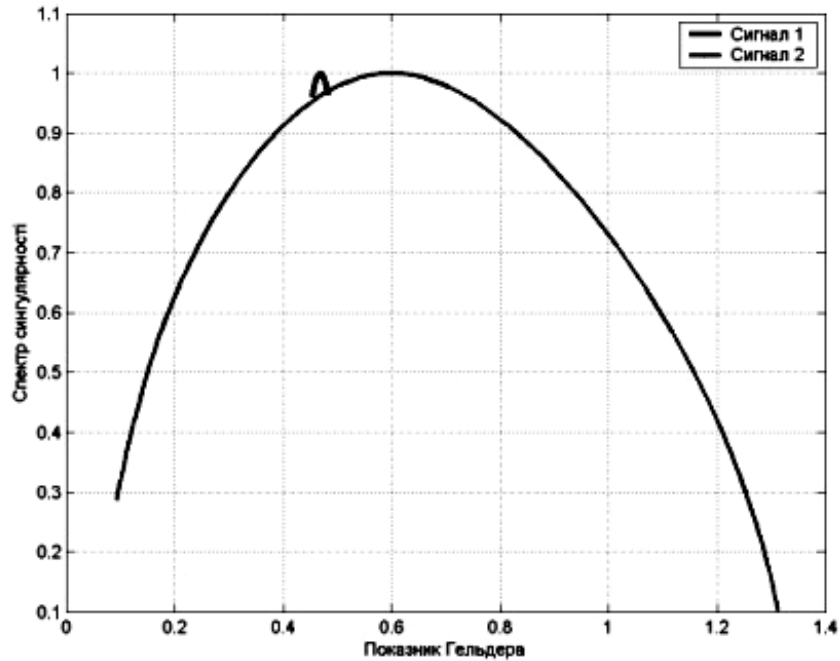


Рис. 9.15. Спектри сингулярностей експериментальних сигналів

Сигнал 2 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 2 може бути апроксимований монофрактальними моделями.

9.4. Структурно-параметрична ідентифікація та прогнозування трафіку

9.4.1. Приклад структурно-параметричної ідентифікації трафіку.

Моделювання розв'язання задачі структурно-параметричної ідентифікації виконувалося за допомогою розроблених програм в середовищі Matlab на основі експериментальних даних трафіку, що передається через мережу Інтернет. Дані являють собою залежність розміру Ethernet кадрів в байтах від часу. Для їх нормування по часовій осі була проведена процедура агрегації з кроком 5 с.

Глибина прогнозу була прийнята 1 такту, а глибина пам'яті за різними входами від 1 до 4. В якості критерію структурної оптимізації обраний критерій незміщеності (8.56). У якості глобальних методів оптимізації застосовувалися: генетичний алгоритм (ГА), ГА з багатокритеріальною оптимізацією (БО), прямий випадковий пошук (ПВП), метод імітації відпалу (МІВ) та метод порогового прийняття (МПП). При цьому використовувалася структура моделей Гаммерштейна-Вінера з базисними функціями НМ прямого розповсюдження (8.58), НМ із РБФ (8.59) і Anfis (8.60).

Результати глобальної оптимізації структури моделі трафіку в ІТМ наведені на рис. 9.16.

Метод БО використовував ГА для знаходження оптимальних по Парето рішень. ПВП мав адаптивний крок пошуку і повний пошук навколо поточної ітерації, МІВ та МПП – обмежену область перевідпалу, ГА – одноточкове схрещування, селективний вибір батьків, формування нової популяції з витісненням. Кількість ітерацій для ПВП, МПП і МІВ (для ГА і БО поколінь) обмежувалося на рівні 100, а розмір простору пошуку для ПВП (для ГА і БО розмір популяції, для МПП і МІВ розмір області перевідпалу) – 30.

При глобальній оптимізації варіювалися наступні структурні характеристики моделі:

- тип базисної функції - НМІР(8.58), НМ із РБФ (8.59) і Anfis (8.60);
- кількість нейронів в прихованому шарі;
- тип функцій активації та приладдя прихованого шару;
- тип алгоритму параметричної оптимізації.

В результаті моделювання (див. рис. 9.16) встановлено, що ГА має найвищу швидкість збіжності (ГА виходить в область оптимальних рішень на перших поколіннях, МПП - в середньому після 5 ітерацій, МІВ - після 15, а БО і ПВП - після 20 ітерацій). Алгоритм МПП виявив найкращу швидкодію (0,6 с на ітерацію при 0,7 с на ітерацію в МІВ, 2,2 с на ітерацію в ПВП, 19,3 с на покоління в ГА і 27,4 с на покоління в БО).

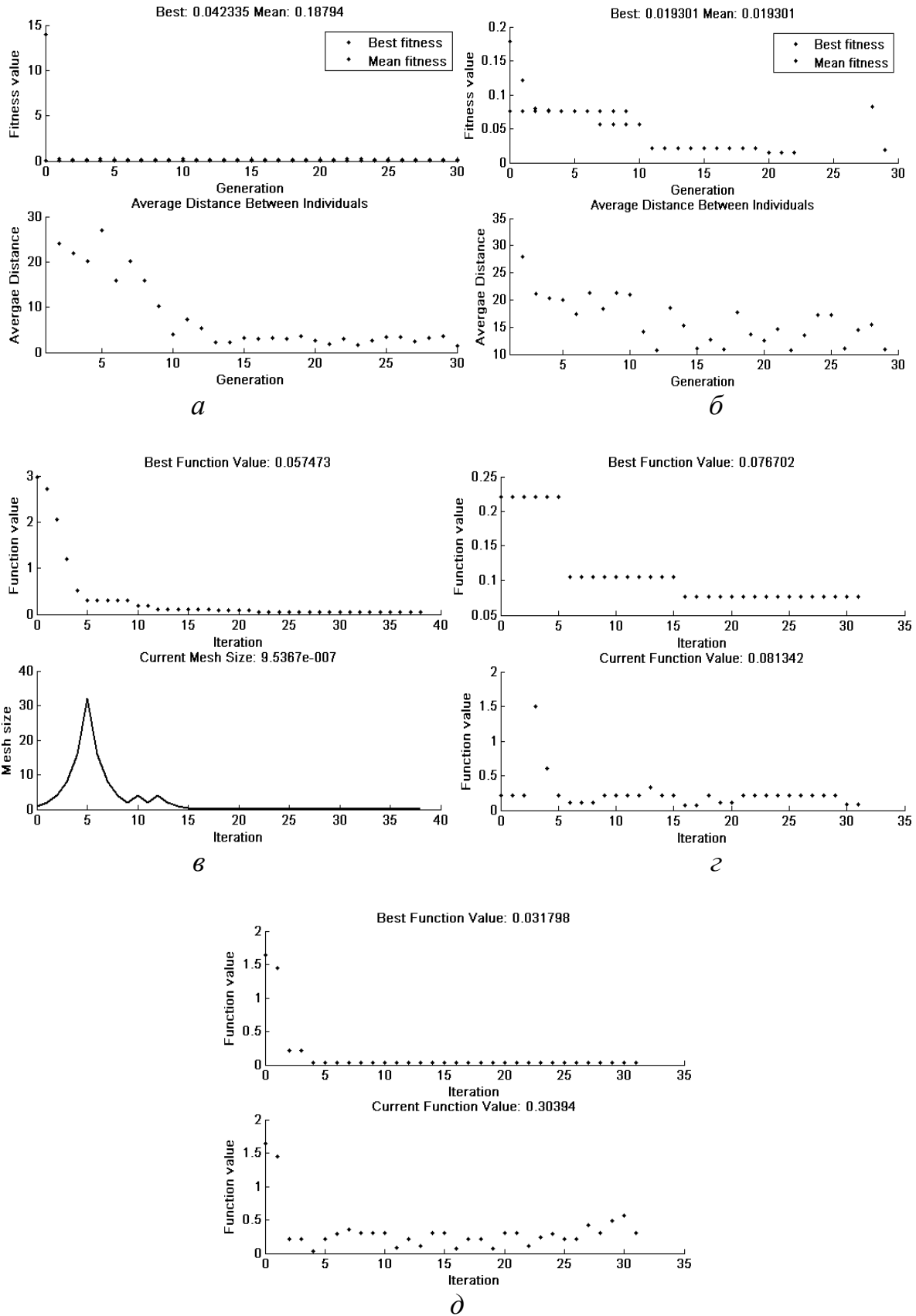


Рис. 9.16. Результати глобальної оптимізації структури моделі трафіку за допомогою ГА (а), БО (б), ПВП (в), МІВ (г) і МПП (д)

Алгоритм БО виявив найкращу збіжність (значення критерію похибки на перевірочній послідовності (8.55) склало 0,019 при його використанні, на відміну від 0032 – при МПП, 0,042 – при ГА, 0,057 – при ПВП і 0,077 – при МІО).

Результат параметричної ідентифікації трафіку в ІТС наведено на рис. 9.17.

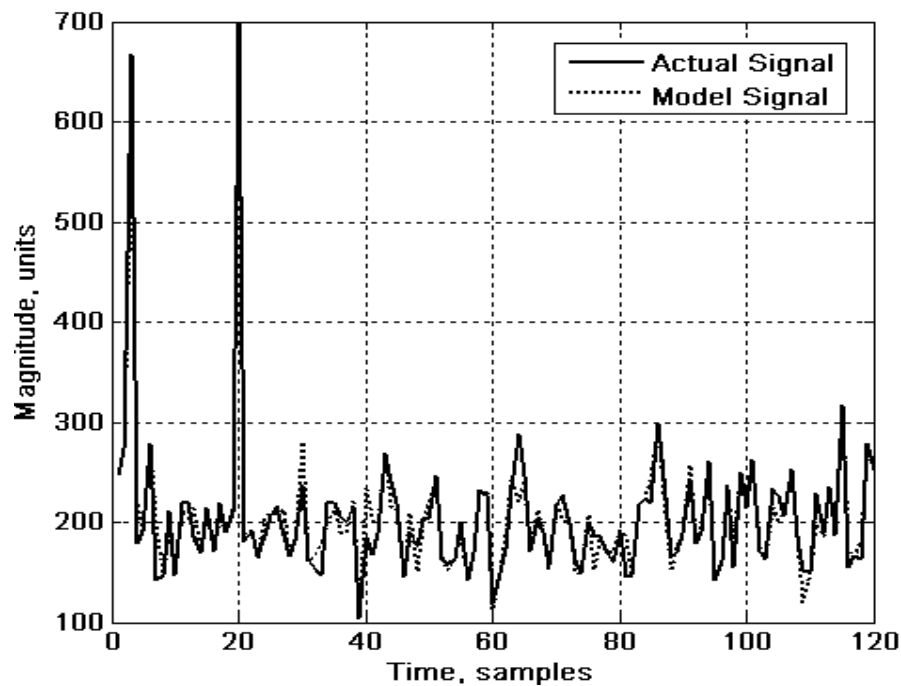


Рис. 9.17. Результат ідентифікації трафіку

Встановлено, що мінімуму критерію незміщенності (8.56) відповідають базисні функції у вигляді каскадної НМПП. При цьому кількість нейронів в прихованому шарі становить 73, функція активації прихованого шару - конкуруюча з м'яким максимумом, вихідного шару - лінійна, алгоритм навчання НМ - градієнтний спуск з вибором параметра швидкості настроювання.

За міру точності параметричної ідентифікації моделі оптимальної структури використовувався критерій похибки (8.55), значення якого склало - 0,0311, що менше похибок відомих моделей трафіку в ІТМ.

Адекватність отриманої інтелектуальної прогнозуючої моделі трафіку перевірялась за непараметричним критерієм знаків. Було встановлено, що для

рівня значущості 0,01 розроблена модель з ідентифікованими структурою і параметрами адекватна експериментальним реалізаціям.

9.4.2. Приклад моделювання трафіку із використанням блочно-орієнтованих структур.

Розв'язання задачі обґрунтування блочно-орієнтованих структур моделей мережевого трафіка виконувалось за допомогою розроблених програм в середовищі Matlab. Експериментальними даними був трафік Internet, наведений в п. 9.4.1. Глибина прогнозу була прийнята 3 такти, а глибина пам'яті за різними входами від 1 до 4. Як критерій оцінки моделей обрано критерій незміщенності (мінімуму зсуву), що дозволяє обрати модель найменш чутливу до зміни множини точок, за якими вона отримана. Така модель повинна давати однакові результати на навчальній та перевірочній вибірках. Тому, цей критерій рекомендується для структурної ідентифікації.

Як типи структур розглядались моделі Гамерштейна-Вінера, ARX, Вінера і Гамерштейна з НСБ (базисними функціями) у вигляді каскадної НМ прямого розповсюдження, вейвнету і поліному Колмогорова-Габора.

Результати ідентифікації трафіка наведені в табл. 9.4 та на рис. 9.18.

Похибки незміщенності моделей, відн. од.

Таблиця 9.4

Базисна функція	Тип структури моделі			
	Гамерштейна-Вінера	ARX	Вінера	Гамерштейна
Каскадна НМПР		0,0619		
Вейвнет	0,0682	0,0679	0,0631	
Поліном Колмогорова-Габора	0,1004			0,1262

Із цієї таблиці видно, що перевагу (у сенсі критерію незміщенності (8.56)) мають мережеві базисні функції (каскадна НМ та вейнет).

Час обчислень на комп'ютері з процесором Pentium IV за моделями Гамерштейна-Вінера, Вінера і Гамерштейна становить 7-10 мс на цикл прогнозу, а по ARX – 0,2 мс, що не вносить часових обмежень на застосування цих моделей для ідентифікації трафіку в ІТМ.

Адекватність отриманих моделей мережевого трафіку перевірялася за непараметричним критерієм знаків. Було встановлено, що для рівня значущості 0,01 розроблені моделі, які адекватні експериментальним реалізаціям.

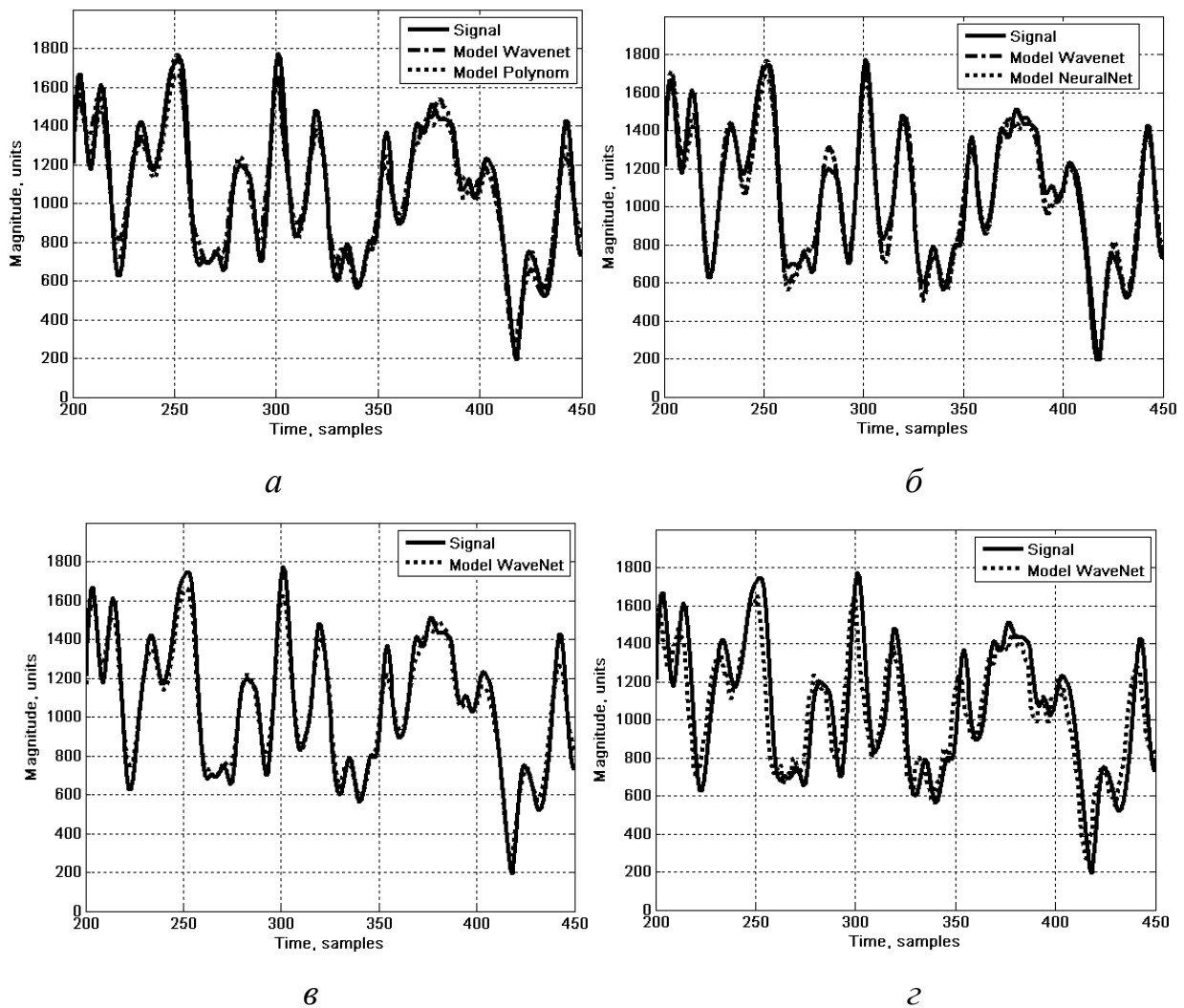


Рис. 9.18. Результати ідентифікації мережевого трафіку у вигляді моделей Гамерштейна-Вінера (*a*), ARX (*б*), Вінера (*в*) і Гамерштейна (*з*)

9.4.3. Приклад адаптивного нечіткого прогнозування трафіку.

Моделювання процесів прогнозування виконувалося за допомогою програм, розроблених в середовищі Matlab на основі експериментальних даних трафіку Internet, наведений в п. 9.4.1. Глибина прогнозу була прийнята у 4 такти, а глибина пам'яті за різними входами від 1 до 4.

Результати глобальної оптимізації для знаходження оптимальної структури (типу) і параметрів АФА наведені на рис. 9.19.

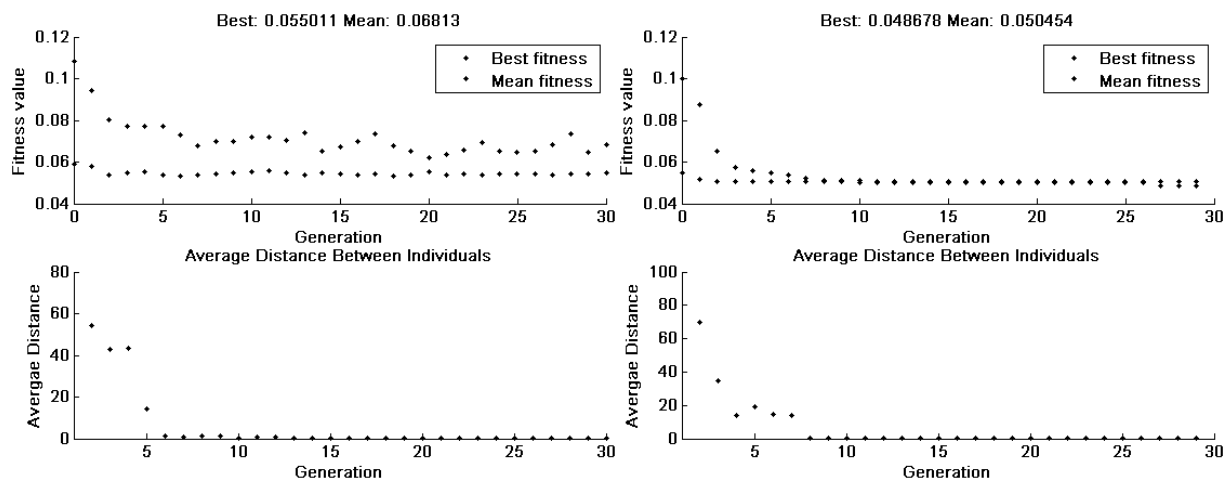
У якості глобальних методів оптимізації застосовувалися ГА, БО, ПВП, МІВ і МПП.

Як критерій параметричної оптимізації використовувався критерій регулярності (8.55), який вираховується на перевірочній вибірці, а для структурної - критерій незміщенності (8.56).

Метод БО використовував ГА для знаходження оптимальних по Парето рішень. ПВП мав адаптивний крок пошуку і повний пошук навколо поточної ітерації, МІВ та МПП - обмежену область перевідпалу, ГА - однотокове схрещування, селективний вибір батьків, формування нової популяції з витісненням. Кількість ітерацій для ПВП, МПП і МІВ (для ГА і БО поколінь) обмежувалося на рівні 100, а розмір простору пошуку для ПВП (для ГА і БО розмір популяції, для МПП і МІВ розмір області перевідпалу) - 30.

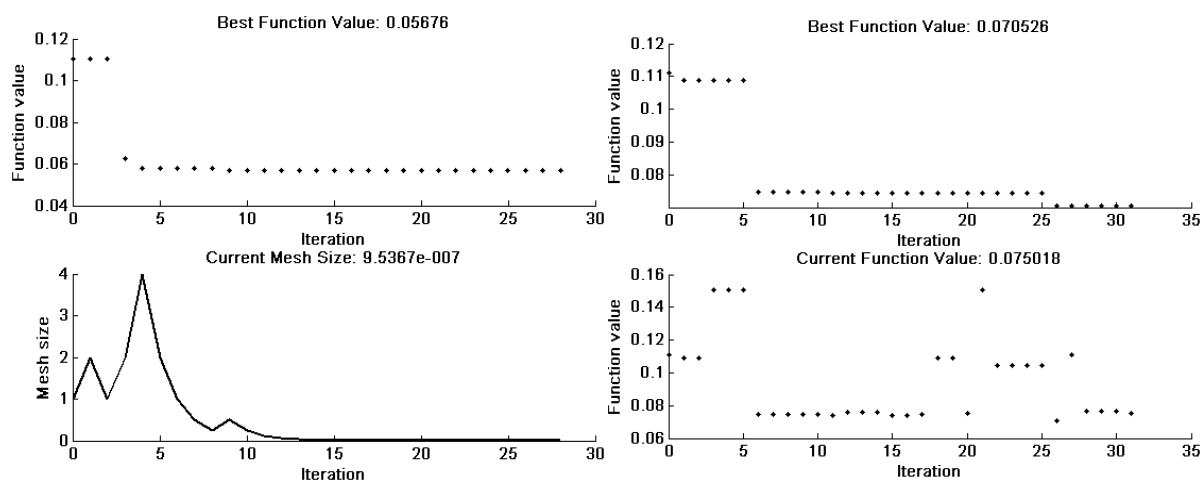
При глобальній оптимізації варіювалися наступні параметри АФА:

- тип АФА - Anfis, Genfis2 і Genfis3;
- для Anfis - кількість нейронів в прихованому шарі, його тип функцій належності і тип алгоритма параметричної оптимізації;
- для Genfis2 - діапазон впливу кластерного центру;
- для Genfis3 - алгоритм нечіткої логіки (Мамдані або Сугено) і кількість кластерів, що визначає число правил і функцій належності.



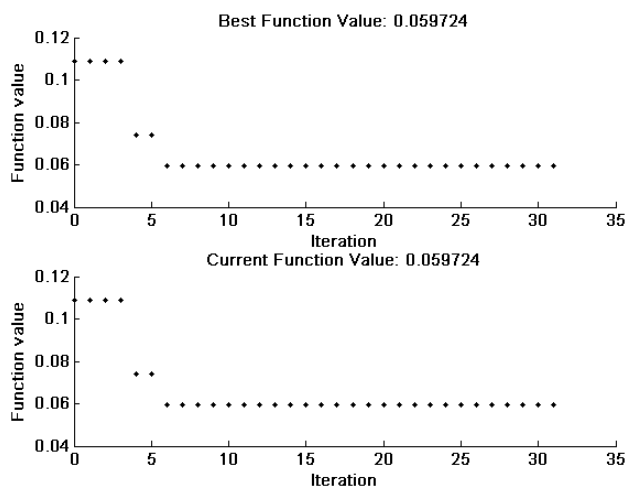
a

б



в

г



д

Рис. 9.19. Результати глобальної оптимізації структури і параметрів АФА для прогнозування трафіку за допомогою ГА (*a*), МО (*б*), ПСП (*в*), МІО (*г*) і МПП (*д*)

В результаті моделювання встановлено, що ГА має найвищу швидкість збіжності (ГА виходить в область оптимальних рішень на перших поколіннях, БО, МПП і ПВП - в середньому після 10 ітерацій, МІВ - після 25 ітерацій).

Алгоритм МІВ виявив найкращу швидкодію (1,9 с на ітерацію при 3 з на ітерацію в МПП, 7,8 с на ітерацію в ПВП, 17,3 с на покоління в ГА і 21,2 с на покоління в БО). При цьому алгоритм БО виявив найкращу збіжність (значення критерію незміщеності (8.56) і при його використанні склали 0,049, на відміну від 0,055 при ГА, 0,057 при ПВП, 0,06 при МПП і 0,07 при МІВ).

Результат прогнозування трафіку в ІТМ наведено на рис. 9.20.

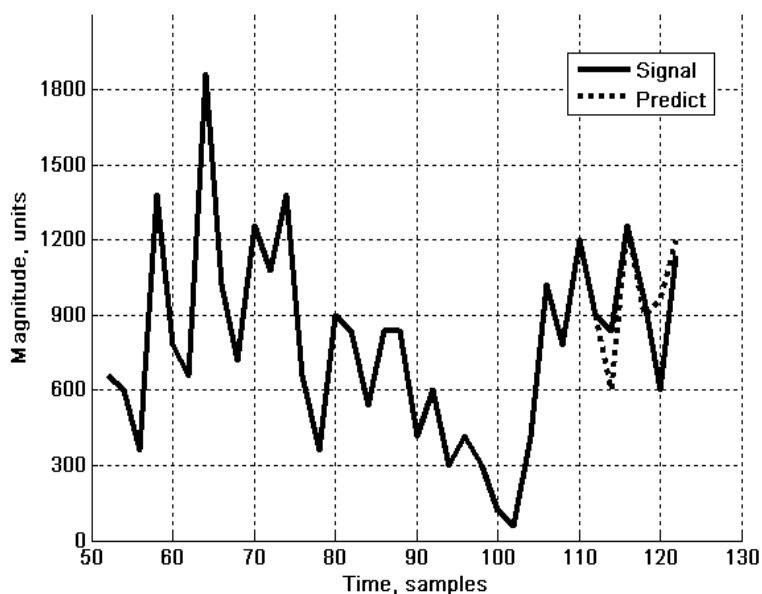


Рис. 9.20. Результат прогнозування мережевого трафіку

Встановлено, що мінімуму критерію регулярності відповідають АФА Genfis2 з діапазоном впливу кластерного центру рівним 0,31. Значення критерію параметричної оптимізації склало - 0,0329.

Адекватність отриманих АФА перевірялась за непараметричним критерієм знаків. Було встановлено, що для рівня значущості 0,01 розроблені АФА адекватні експериментальним реалізаціям.

Контрольні запитання

1. В чому полягає комплексна оцінка характеристик трафіка.
2. Розкрийте зміст часо-частотного аналізу трафіка.
3. Розкрийте зміст статистичного аналізу трафіка.
4. Розкрийте зміст фрактального аналізу трафіка.
5. Розкрийте зміст методики оцінки характеристик мережевого трафіка.
6. Розкрийте зміст моделей мережевого самоподібного трафіка.
7. В чому полягає мультифрактальний аналіз трафіка.
8. В чому полягає мультифрактальний аналіз детрендових флуктуацій.
9. Розкрийте зміст методу максимумів модулів вейвлет-перетворення для аналізу мережевого трафіку.
10. Які особливості моделювання трафіка із використанням блочно-орієнтованих структур?

Глава 10. ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНИХ МЕТОДІВ ПРИ РІШЕННІ ПРИКЛАДНИХ ЗАВДАНЬ

10.1. Адаптивне оцінювання та ідентифікація процесів за допомогою вейвнет

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 – **Wavelet Network**). Такі АФА апроксимують нелінійні функції з високою точністю, але мають недостатні прогностичні властивості, що особливо важливо для процесів із значними запізненнями.

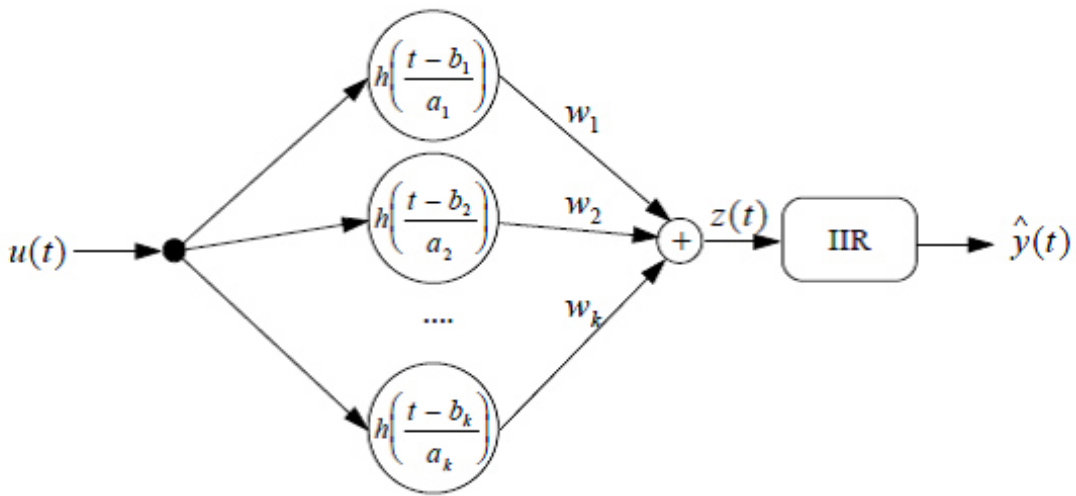
Для подолання цього недоліку було запропоновано будувати АФА у вигляді комбінації вейвнету і фільтра з нескінченною імпульсною характеристикою (блок НІХ), які утворюють структуру Гаммерштейна нелінійної динамічної системи.

10.1.1. Структура вейнету з блоком із нескінченною імпульсною характеристикою.

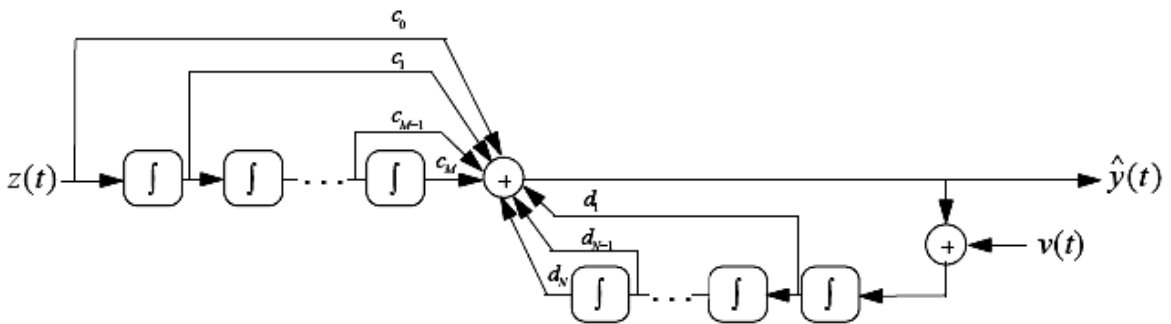
Вейвнет є мережею, в якій функція виходу добре локалізована в часовій і в частотній областях. Комбінування каскадної побудови вейвнет з блоком НІХ забезпечує обчислювально ефективно навчання системи і зменшує час досягнення мінімально допустимої похибки системи. В вейвнеті (рис. 10.1) формується оцінка 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:

$$\hat{y}(t) = \sum_{i=0}^M c_i z(t-i)u(t) + \sum_{j=1}^N d_j \hat{y}(t-j)v(t), \quad (10.1)$$

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

Рис. 10.1. Структура вейвнету із блоком НІХ (a) 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:

$$\bar{c}(n+1) = \bar{c}(n) + \mu_c \Delta \bar{c}; \quad \bar{c} = \{w, a, b, c, d\}, \quad (10.2)$$

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 застрявання 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.

10.1.2. 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».

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. 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. 10.2 і 10.3 та в табл. 10.1.

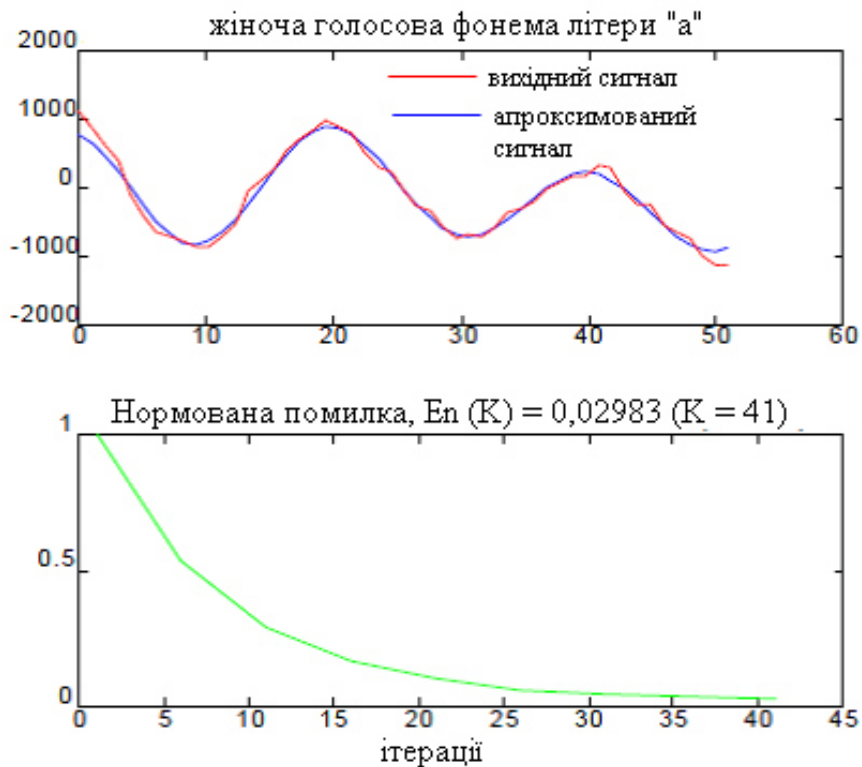


Рис. 10.2. Процес навчання вейвнету без блоку НІХ

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, вейнет з блоком НІХ має в 1,41 раз більшу швидкість збіжності.

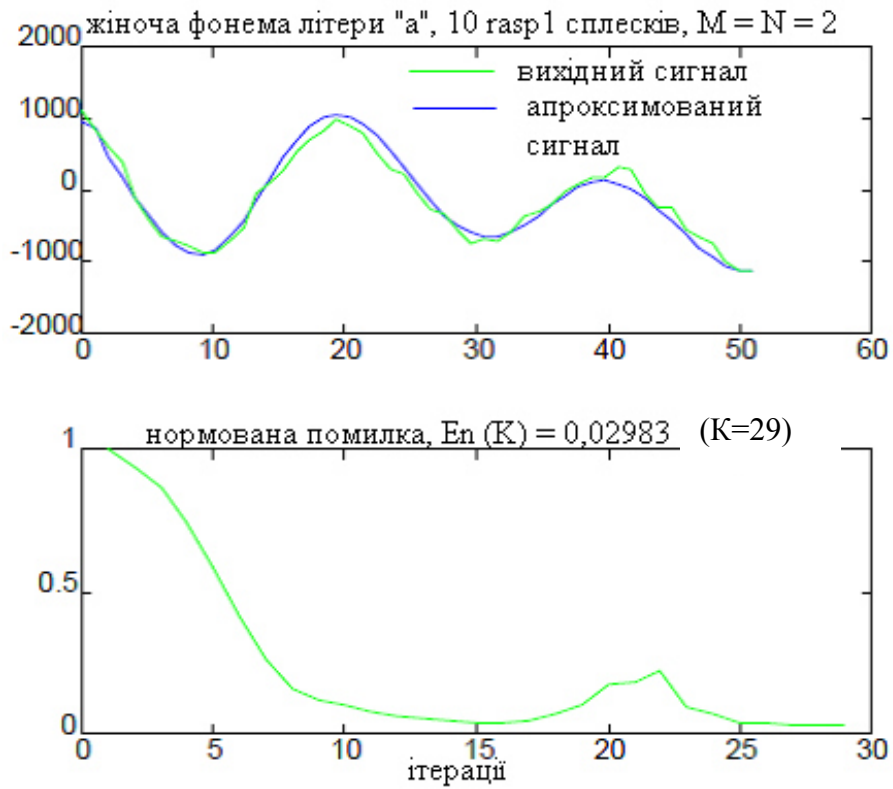


Рис. 10.3. Процес навчання вейвнету з блоком НІХ

Кількість ітерацій для досягнення бажаної похибки

Таблиця 10.1

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

10.1.3. Фільтрація голосу при імпульсних завадах.

Однією з переваг використання адаптивних НМ є їх стійкість до завад, наприклад, коли сигнали забруднені негаусовими шумами чи завадами. 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. 10.4. Тут оригінал сигналу представлений переривчастою лінією, забруднений сигнал – тонкою лінією і синтезований мовний сигнал з використанням шести вейвлет Морле – жирною лінією.

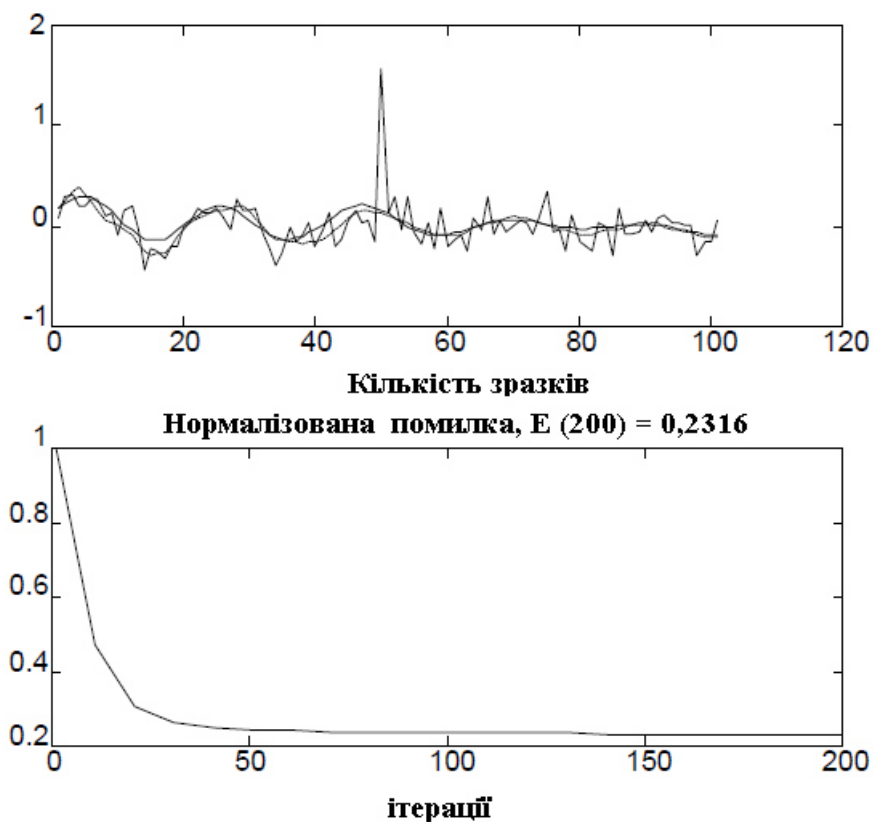


Рис. 10.4. Моделювання мовних голосових даних, спотворених шумом та імпульсною завадою

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.

10.1.4. Апроксимація складних функцій.

Розглянемо навчання та порівняємо ефективність моделювання складних нелінійних функцій за допомогою НМПР із алгоритмом зворотного поширення похибки та вейнету без блоку НІХ.

Хай складна кусково-лінійна функція, яку треба апроксимувати, має вигляд системи рівнянь:

$$f(x) = \begin{cases} -2.186x & -10 \leq x < -2; \\ 4.246x & -2 \leq x < 0; \\ 10e^{-0.05x-0.5} \cdot \sin((0.03x + 0.7)x) & 0 \leq x \leq 10. \end{cases} \quad (10.3)$$

НМПР 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.

Результати навчання та моделювання НМПР та вейнету наведені на рис. 10.5 і 10.6, відповідно.

Із графіків видно, що сума квадратів похибки для НМПР склала на трьохтисячній ітерації величину 0.7872, а для вейнету – 0.0663 всього лиш на сотій ітерації. Тобто вейвнет при апроксимації складних функцій має суттєво кращу швидкість збіжності та саму збіжність.

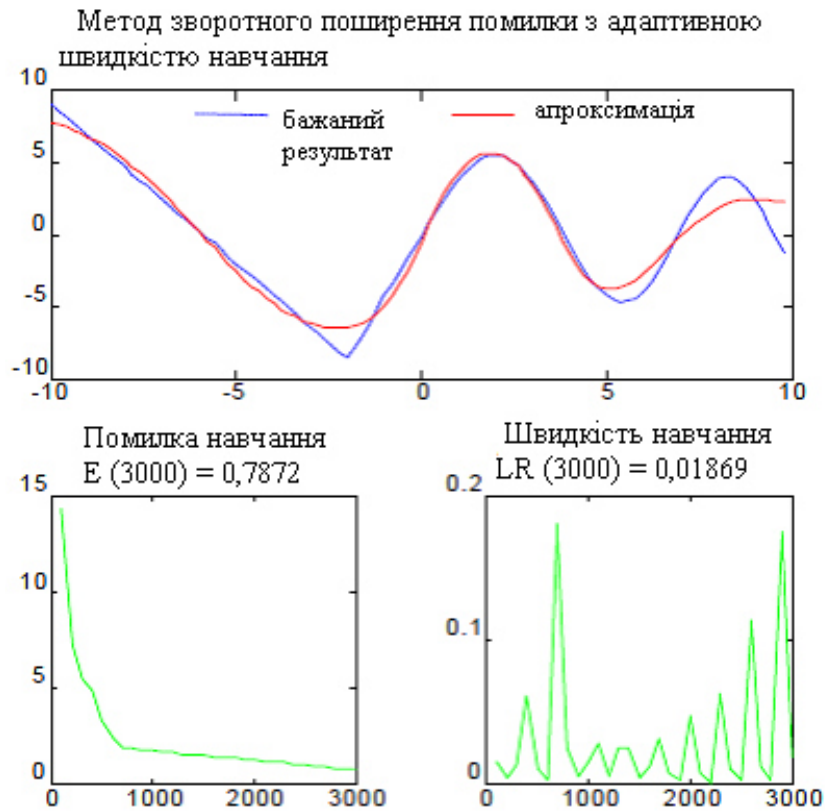


Рис. 10.5. Апроксимація функції за допомогою НМПР

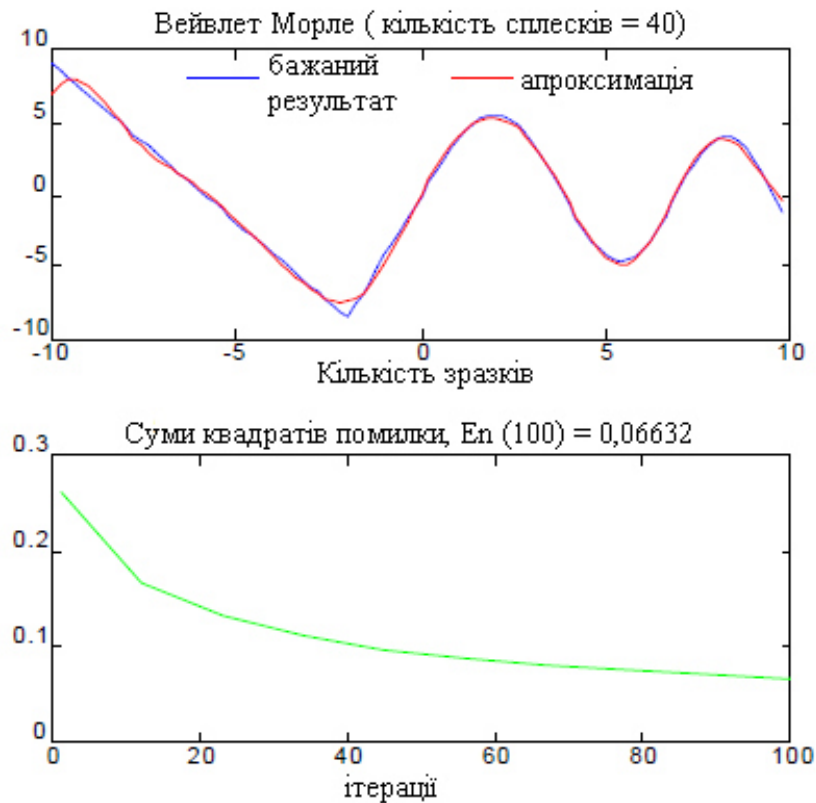


Рис. 10.6. Апроксимація функції за допомогою вейвнету без блока НІХ

10.1.5. Ідентифікація нелінійного динамічного процесу із нормально розподіленим шумом.

Розглянемо ідентифікацію складного динамічного процесу (системи) із шумом за допомогою вейвнету із блоком НІХ. При цьому процес описується рівнянням:

$$y(n) = \frac{y(n-1)y(n-2)}{1 + y^2(n-1) + y^2(n-2)} + u(n), \quad (10.4)$$

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 в рівнянні (10.4), результати чого показано на рис. 10.7. При цьому, нормована похибка склала 0.02352 на 50-й ітерації.

Для оцінки завадостійкості ідентифікації за допомогою вейвнету з блоком НІХ у виразі (10.4) додано шум з нормальним розподілом $\sigma^2 = 0.1$. Всі мережеві параметри були прийняті такими ж, як і вище.

Результати моделювання представлені на рис. 10.8.

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. 10.8.) Це моделювання підкреслює високу завадостійкість і надійність мережі на основі вейвнету з блоком НІХ.

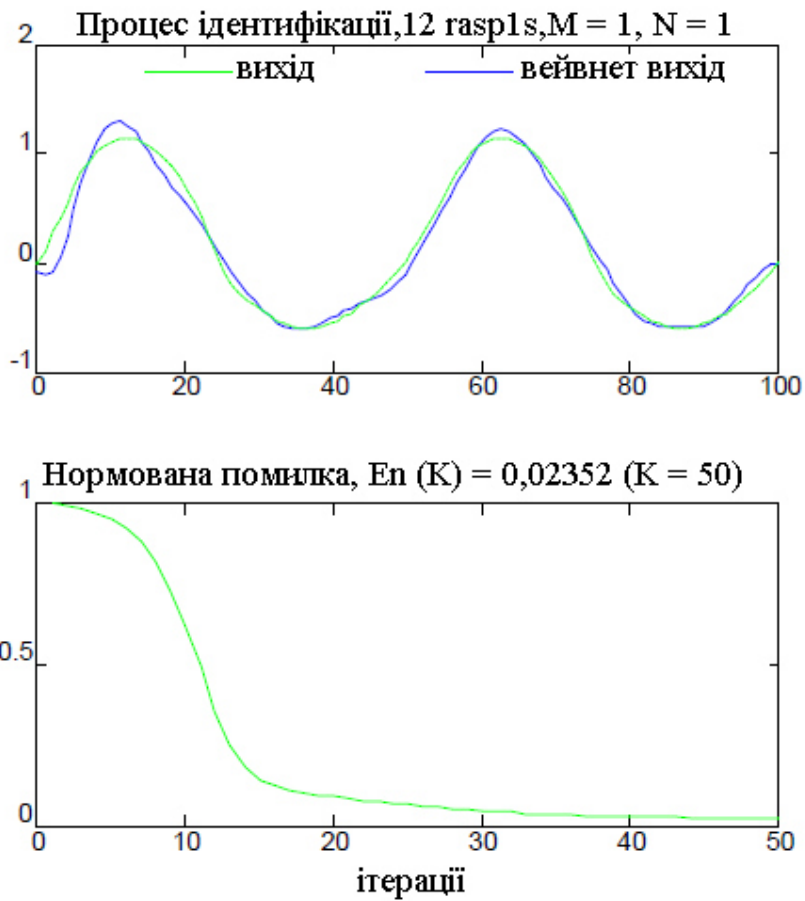


Рис. 10.7. Ідентифікація нелінійної динамічної системи без шуму

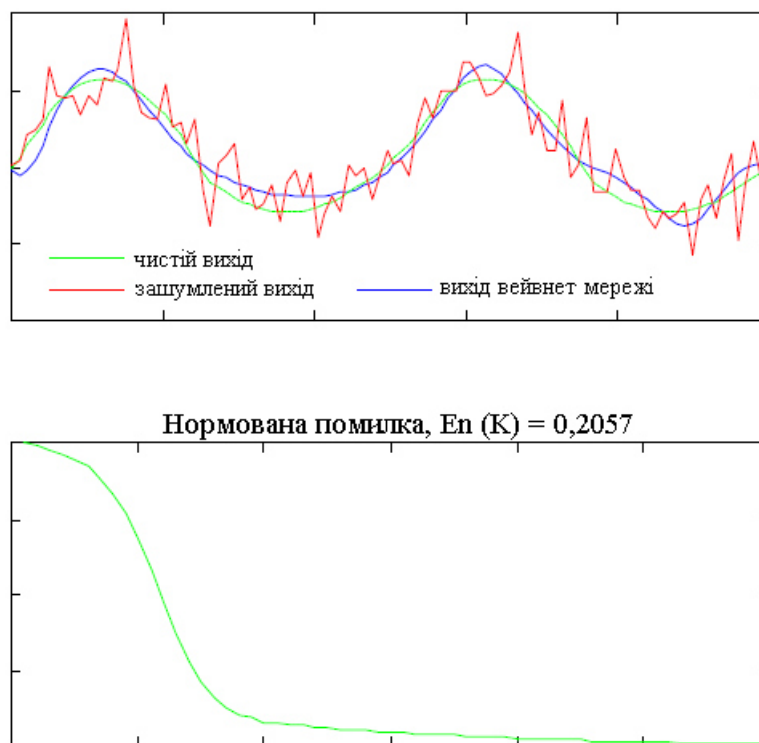


Рис. 10.8. Ідентифікація нелінійної динамічної системи з шумом

10.2. Оцінка систем захисту інформації від витоку

10.2.1. Оцінка DLP-систем з використанням засобів нечіткої логіки.

У зв'язку з достатньо великим різноманіттям програмних систем захисту інформації від її витоків (DLP-систем), представлених на ринку програмного забезпечення, і їх високою вартістю, доцільним є вибір тієї системи, яка найповніше задовольняє запитам конкретної організації (компанії) і забезпечує необхідний рівень захисту інформації від витоків.

Для оцінки DLP-систем використовуються ряд критеріїв, кожен з яких описується за допомогою лінгвістичної змінної і є вхідної змінної системи нечіткого висновку:

1. Технології розпізнавання конфіденційної інформації (код лінгвістичної змінної **A1**). На сьогодні існує декілька базових технологій, кожна з яких підходить для аналізу інформації різних видів у тому числі і для пошуку інформації в зашифрованих даних.

2. Повнота контрольованих каналів (**A2**). Кожен канал передачі інформації — це потенційний канал витоків, тому важливо блокувати всі невживані для роботи канали, а що залишилися — контролювати.

3. Зручність управління (**A3**). Для збереження якості розпізнавання конфіденційної інформації на належному рівні необхідно здійснювати регулярне обслуговування, аудит і корегування налаштувань. З цією метою DLP-система повинна гарантувати простоту встановлення, налаштування та експлуатації.

4. Ведення журналів і звіти (**A4**). У журналах повинні акумулюватися події і об'єкти, що фіксуються системою в ході роботи. Звіти на основі зібраної інформації можуть застосовуватися для аналізу дій користувачів, для розслідування інцидентів інформаційної безпеки, для контролю стану захищеності інформації, коректності налаштувань тощо.

5. Цінова доступність (**A5**). Вживані рішення повинні бути адекватними вартості інформації, що захищається, і наявним фінансовим можливостям компанії.

6. Рівень ресурсоемності (**A6**). Найчастіше замовник наполягає на програмному рішенні, яке легко впроваджується в існуючу інфраструктуру без зміни архітектури мережі, використовуваних прикладних програм і апаратних засобів.

Таким чином, вхідними параметрами бази правил системи нечіткого висновку є 6 вхідних змінних **A1-A6**. Значення лінгвістичної змінної **A1** задаються за допомогою терм-множини **A1**=[«недостатні», «достатні»], а **A2-A6**=[«низька», «середня», «висока»].

Вихідною змінною бази правил системи нечіткого висновку є комплексна оцінка DLP-системи, яка може бути складною нелінійною функцією від вхідних параметрів. З метою комплексного оцінювання системи лінгвістичні змінні були розбиті на дві групи: експлуатаційну і функціональну. Експлуатаційна група оцінюється лінгвістичною змінною **V1**, функціональна група — змінною **V2**. Значення лінгвістичних змінних **V1** і **V2** задаються терм-множиною [«низька», «середня», «висока»].

Для лінгвістичної змінної **V1** сформульовано 6 продукційних правил:

1. Якщо **A1**=«недостатні», То **V1**= «низька».
2. Якщо **A2**=«низька», То **V1**=«низька».
3. Якщо **A1**=«достатні» І **A2**=«середня» І **A6**=«низька», То **V1**=«середня».
4. Якщо **A1**=«достатні» І **A2**=«середня» І (**A6**=«середня» АБО **A6**=«висока»), То **V1**=«низька»;
5. Якщо **A1**=«достатні» І **A2**=«висока» І (**A6**=«середня» АБО **A6**=«низька»), То **V1**=«висока».
6. Якщо **A1**=«достатні» І **A2**=«висока» І **A6**=«висока», То **V1**=«середня».

Для лінгвістичної змінної **V2** сформульовано 10 продукційних правил:

1. Якщо **A4** = «низька», То **V2** = «низька».
2. Якщо **A5** = «низька», То **V2** = «низька».
3. Якщо **A3** = «висока» І **A4** = «висока», То **V2** = «висока».
4. Якщо **A3** = «висока» І **A4** = «висока» І (**A5** = «середня» АБО **A5** = «висока»), То **V2** = «висока».
5. Якщо **A3** = «середня» І **A4** = «середня» І **A5** = «середня», То **V2** = «середня».
6. Якщо **A3** = «низька» І **A4** = «висока» І **A5** = «висока», То **V2** = «висока».
7. Якщо **A3** = «низька» І **A4** = «середня», То **V2** = «низька»;
8. Якщо **A3** = «висока» І **A4** = «висока» І (**A5** = «середня» АБО **A5** = «висока»), То **V2** = «висока».
9. Якщо **A3** = «висока» І **A4** = «середня» І (**A5** = «середня» АБО **A5** = «висока»), То **V2** = «середня».
10. Якщо **A3** = «середня» І **A4** = «висока» І (**A5** = «середня» АБО **A5** = «висока»), То **V2** = «висока».

Значення вихідної лінгвістичної змінної **R**, що показує комплексну оцінку, задаються терм-множиною [«низька», «середня», «висока»]. Продукційні правила комплексної оцінки DLP-системи визначені через лінгвістичні змінні **V1** і **V2**:

1. Якщо **V1** = «низька», То **R** = «низька».
2. Якщо **V1** = «середня» І **V2** = «низька», То **R** = «низька».
3. Якщо **V1** = «середня» І (**V2** = «середня» АБО **V2** = «висока»), То **R** = «середня».
4. Якщо **V1** = «висока» І **V2** = «низька», То **R** = «середня»
5. Якщо **V1** = «висока» І (**V2** = «середня» АБО **V2** = «висока»), То **R** = «висока».

Для реалізації бази нечітких правил використовують, наприклад, алгоритм Мамдані.

10.2.2. Оцінка DLP-систем з використанням штучних нейронних мереж.

Як вхідні сигнали НМ можуть використовуватися значення критеріїв, що характеризують особливості DLP–системи і які затребувані в організації-споживачеві:

- оцінка форматів розпізнавання інформації;
- оцінка контрольованих каналів;
- оцінка складності супроводження;
- оцінка повноти і зручності створення звітів;
- оцінка ціни придбання і супроводу.

Значення критеріїв задаються на основі експертних оцінок в інтервалі від 0 до 1. Кількість виходів відповідає кількості DLP–систем, що розглядаються для придбання. Навчання НМ здійснюється так, щоб максимальний відгук НМ був отриманий на виході, відповідному той DLP–системі, яка найповніше задовольняє вимогам, що пред'являються організацією.

Наприклад, на перший вхід подається значення, відповідне рівню технології розпізнавання інформації, яке потрібне підприємству для забезпечення інформаційної безпеки. На одному з виходів системи сигнал буде активним. По номеру активного виходу визначається DLP–система, яка найбільш підходить для використання на підприємстві.

Структура системи комплексної оцінки DLP-системи складається з двох рівнів. Перший рівень містить у собі модулі врахування протидії ризику витоку інформації:

- мережею;
- у кінцевих точках;
- різних типів.

До першого рівня також входить модуль врахування складності налаштування, обслуговування та вимог до ресурсів.

Другий рівень містить модуль комплексної оцінки.

Залежно від наявних характеристик DLP–системи формуватимуться вхідні сигнали модулів першого рівня. Для реалізації кожного модуля використовується НМ. Число входів кожної НМ обумовлене числом показників, що характеризують оцінку рівня ефективності DLP–системи. На виходах НМ першого рівня формуються значення оцінки відповідного параметра, а другий рівень забезпечує формування комплексної оцінки

Дворівневе обчислення спрощує збір даних для навчання НМ, а також процес формування експертних оцінок для формування входів НМ.

У якості НМ переважним тут є використання нейронних мереж зустрічного розповсюдження (LVQ-network), що складаються з вхідного шару і шарів нейронів Кохонена і Гросберга. Окрім хороших здібностей до узагальнення, ці мережі мають на 2 порядки менший час навчання.

10.3. Методика нечіткої оцінки технічного стану комп'ютерної системи в задачах інформаційної безпеки

Для комп'ютерних систем випадкові збої і відмови практично неминучі. Їх виникнення може бути обумовлене як внутрішніми технологічними причинами, так і зовнішніми впливами (механічними, кліматичними електромагнітними, біологічними, термічними тощо). В результаті збоїв або відмов може відбутися спотворення і навіть знищення даних, що зберігаються і оброблюються в системі.

Враховуючи, що корпоративні комп'ютерні системи конструктивно менш захищені, ніж промислові, вони можуть бути чутливі до різного роду зовнішніх впливів. Важливим є відстежування їх фізичного стану на етапі зберігання і обробки інформації.

Для комп'ютерної системи на цілісність інформації і апаратного забезпечення впливають температури центрального процесора (ЦП), системного блоку і жорсткого диска, а також вологість в корпусі.

У загальному випадку температура ЦП в режимі простою складає до 45°C, а при навантаженні доходить до 65°C. При цьому температура 70°C і вище вважається для процесора критичною, оскільки можливе зниження продуктивності і перевантаження комп'ютера. Даний показник залежить від технології виготовлення і, в залежності від виробника, допустимий тепловий діапазон може складати 30–35°C, а при навантаженні — 70–75°C.

Температура жорсткого диска також є іррегулярною величиною, залежною від ряду обставин. Вважається, що температура нижче 25°C і 45–52°C небажані, а температура вище 55°C є вже критичною для жорсткого диска.

Складніше формалізується вплив вологості в системному блоці. Низька вологість на рівні 15–20% приводить до накопичення в повітрі статичної електрики і, як наслідок електризації пилу, забрудненню апаратури і утворенню струмопровідних доріжок. Недостатня вологість (до 30%) приводить до руйнування лаку на електронних друкарських платах, висиханню і твердінню ізоляції проводів системного блоку з подальшим їх розтріскуванням. Вологість вище 60% є причиною корозії і окислення контактів, що може викликати коротке замикання.

Розглянуті параметри взаємопов'язані. Температура ЦП може залежати від стани термопасти і частоти обертання вентилятора, що охолоджує ЦП. На температури жорсткого диска і системного блоку впливає встановлений в ньому вентилятор. Частота обертання вентиляторів залежить від температури і регулюється материнською платою. Вологість в корпусі залежить від стану приміщення, для її оцінки потрібна установка в корпус датчика вологості.

Розглянемо методику нечіткої оцінки стану системного блоку комп'ютера.

Введемо вхідні лінгвістичні змінні:

T1 — температура ЦП;

T2 — температура жорсткого диска;

T3 — температура в системному блоці;

V1 — швидкість обертання вентилятора ЦП;

V2 — швидкість обертання вентилятора в системному блоці;

H — вологість в системному блоці.

Вихідною лінгвістичною змінною є **R** — рівень ризику інформаційної безпеки.

Визначимо терм-множину для вхідних і вихідної лінгвістичних змінних. Для вхідних лінгвістичних змінних **T2**, **V1**, **V2**, **H** введемо терм-множину [«низька», «середня», «висока»], для **T1**, **T3** — [«критично низька», «низька», «середня», «висока», «критично висока»]. Терм-множину вихідної лінгвістичної змінної описує рівень інформаційного ризику **R**=[«відсутній», «маловірогідний», «низький», «достатньо високий», «критично високий»].

В результаті нечітка база знань має 27 продукційних правил:

1. Якщо **T1**=«критично низька», То **R**=«критично високий».
2. Якщо **T1**=«критично висока», То **R**=«критично високий».
3. Якщо **T1**=«низька», То **R**=«відсутній».
4. Якщо **T1**=«середня» І **V1**=«низька», То **R**=« відсутній».
5. Якщо **T1**=«середня» І **V1**=«середня», То **R**=«відсутній».
6. Якщо **T1**=«середня» То **R**=«маловірогідний».
7. Якщо **T1**=«висока» І **V1**=«низька», То **R**=«достатньо високий».
8. Якщо **T1**=«висока» І **V1**=«середня», То **R**=«низький».
9. Якщо **T1**=«висока» І **V1**=«висока», То **R**=«маловірогідний».
10. Якщо **H**=«середня», То **R**=«відсутній».
11. Якщо **H**=«висока», То **R**=«критично високий».
12. Якщо **H**=«низька», То **R**=«критично високий».
13. Якщо **T3**=«критично низька», То **R**=«критично високий».
14. Якщо **T3**=«критично висока», То **R**=«критично високий».

15. Якщо $T3 = \text{«низька»}$ І $V2 = \text{«низька»}$, То $R = \text{«відсутній»}$.
16. Якщо $T3 = \text{«середня»}$ І $V2 = \text{«низька»}$, То $R = \text{«маловірогідний»}$.
17. Якщо $T3 = \text{«середня»}$ І $V2 = \text{«середня»}$, То $R = \text{«відсутній»}$.
18. Якщо $T3 = \text{«середня»}$ І $V2 = \text{«висока»}$, То $R = \text{«маловірогідний»}$.
19. Якщо $T3 = \text{«висока»}$ І $V2 = \text{«низька»}$, То $R = \text{«критично високий»}$.
20. Якщо $T3 = \text{«висока»}$ І $V2 = \text{«середня»}$, То $R = \text{«низький»}$.
21. Якщо $T3 = \text{«висока»}$ І $V2 = \text{«висока»}$, То $R = \text{«низький»}$.
22. Якщо $T2 = \text{«критично низька»}$, То $R = \text{«критично високий»}$.
23. Якщо $T2 = \text{«висока»}$, То $R = \text{«критично високий»}$.
24. Якщо $T2 = \text{«низька»}$ І $V2 = \text{«низька»}$, То $R = \text{«відсутній»}$.
25. Якщо $T2 = \text{«середня»}$ І $V2 = \text{«низька»}$, То $R = \text{«маловірогідний»}$.
26. Якщо $T2 = \text{«середня»}$ І $V2 = \text{«середня»}$, То $R = \text{«відсутній»}$.
27. Якщо $T2 = \text{«середня»}$ І $V2 = \text{«висока»}$, То $R = \text{«маловірогідний»}$.

На практиці форма кривих функції належності обирається виходячи із експертних оцінок.

Як механізм нечіткого логічного висновку системи використовувався алгоритм Мамдані, що полягає в застосуванні мінімаксної композиції нечітких множин. Система реалізована в середовищі Matlab і дозволяє оцінити ризик інформаційної безпеки в комп'ютерній системі на підставі введеної вихідної лінгвістичної змінної R .

Розглянемо приклад проведення подібної оцінки. Для цього припустимо, що на підставі вимірювань в системному блоці були отримані наступні вхідні дані: температура ЦП склала 45°C , температура жорсткого диска — 44°C , температура в системному блоці — 45°C , при цьому швидкість обертання вентилятора ЦП — 2000 об/хв, швидкість обертання вентилятора в системному блоці — 2500 об/хв, вологість в системному блоці — 50%.

Згідно виконаному розрахунку рівень ризику інформаційної безпеки при виміряних параметрах є $R = \text{«низький»}$, що відповідає початковим вхідним даним.

Проведені аналогічні дослідження представленої нечіткої моделі для різних наборів початкових даних також показали прийнятні результати.

10.4. Методика виявлення DoS-атак в інформаційно-комунікаційних мережах з використанням штучних нейронних мереж

Одним з найбільш руйнівних інструментів зловмисників в останні роки стали атаки типу «відмова в обслуговуванні» – Denial of Service (DoS) в інформаційно-комунікаційних мережах (табл. 10.2).

Особлива небезпека цих атак полягає у простоті їх організації і в тому, що для їхньої реалізації зловмисникам не потрібно володіти якими-небудь особливими знаннями та вміннями.

Задача виявлення атак (мережевих аномалій) є досить актуальною і полягає в пошуку ефективних методів їх виявлення для побудови систем виявлення та попередження атак (Intrusion Detection and Prevention System – IDPS). Основними вимогами до таких IDPS є високий рівень виявлення атак поряд з невеликою кількістю хибних спрацьовувань, швидкість роботи, робота в режимі реального часу тощо.

Дослідження властивостей НМ для вирішення задачі класифікації даних при виявленні DoS-атак передбачає:

- визначення критеріїв виявлення DoS-атак;
- отримання тестових даних;
- визначення критеріїв порівняння НМ;
- визначення топології НМ;
- навчання НМ різних архітектур та апробація результатів їх роботи;
- отримання результатів виявлення атак, їх аналіз;
- порівняльний аналіз результатів роботи НМ та вибір оптимальної архітектури НМ для заданих умов функціонування IDPS.

DoS-атаки			Методи виявлення
1	Атаки великою кількістю сфальсифікованих пакетів, спрямовані на виснаження ресурсів вузла або мережі	<ul style="list-style-type: none"> ▪ Smurf ▪ ICMP flood ▪ UDP flood ▪ TCP flood ▪ TCP SYN flood 	<ul style="list-style-type: none"> ▪ Сигнатурний аналіз ▪ Штучні нейронні мережі
2	Атаки спеціально сконструйованими пакетами, що викликають загальний збій системи через помилки в програмах	<ul style="list-style-type: none"> ▪ Ping of death ▪ Teardrop ▪ Land ▪ Nuke 	<ul style="list-style-type: none"> ▪ Експертні системи ▪ Статистичний аналіз
3	Атаки сфальсифікованими пакетами, що викликають зміни в конфігурації або стані системи	<ul style="list-style-type: none"> ▪ Десинхронізація TCP-з'єднання ▪ Скидання TCP-з'єднання ▪ Примус на зменшення швидкості передачі даних 	<ul style="list-style-type: none"> ▪ Кластерний аналіз

10.4.1. Інформативні ознаки DoS-атак.

Для визначення критеріїв виявлення DoS-атак необхідно розглянути інформативні ознаки DoS-атак, що наведені в табл. 10.3.

З метою зменшення кількості вхідних змінних НМ доцільно об'єднати інформативні ознаки у загальні групи.

Обсяг трафіку, що циркулює в корпоративній мережі, змінюється протягом дня й залежить від часу доби. Ця характеристика показує активність бізнес-процесів, нормальне протікання яких забезпечує інформаційна система, що надає доступ до своїх ресурсів. Зазвичай, типова залежність кількості звернень від часу доби має вигляд гаусової кривої.

Ознаки DoS-атаки

Таблиця 10.3

Інформативна ознака	DoS-атака, що виявляється	Група, до якої належить ознака
Інтенсивність вхідних UDP-пакетів	UDP-flood	UDP
Інтенсивність вхідних SYN-сегментів TCP	TCP SYN flood	TCP
Інтенсивність вхідних ACK-сегментів TCP	Десинхронізація TCP-з'єднання	
Інтенсивність вихідних ACK-сегментів TCP	Десинхронізація TCP-з'єднання	
Інтенсивність RST-сегментів TCP	Розрив TCP- з'єднання	
Інтенсивність вхідних повідомлень ICMP DestinationUnreachable	Розрив TCP- з'єднання	ICMP
Інтенсивність вхідних повідомлень ICMP SourceQuench	Примус на зменшення швидкості передачі даних	
Інтенсивність вхідних повідомлень ICMP Echo-Reply	Smurf	

Таким чином, спостереження за станом потоку звернень до інформаційної системи може дати інформацію про нормальний її стан. При DoS-атаці на систему закон її нормального функціонування буде спотворений через зовнішній вплив.

Визначення «нормального» стану інформаційно-комунікаційної мережі дає змогу встановити допустиму межу відповідно до визначених ознак виявлення DoS-атак.

Для цього можна використати пакетний сніфер CommView – програму перехоплення й аналізу трафіку Інтернет та локальної мережі. Вона дає змогу:

- зібрати інформацію про дані, що проходять через мережеві пристрої та декодувати ті з них, що необхідні для аналізу;
- зафіксувати перелік мережевих з'єднань, IP-статистику;
- захоплювати/ігнорувати конкретні пакети, які потім зберігаються у файл для наступного аналізу й вибору інформативних даних для навчання НМ.

Стан DoS-атаки моделюється за допомогою пакетного сніферу, що дозволяє генерувати пакети заданого розміру, власноруч встановлювати управляючі прапори тощо, а потім надсилати їх на машину-жертву із заданою інтенсивністю.

У процесі експерименту пакетний сніфер записував дані про пакети у log-файл. Відповідно до визначених групових ознак збережено 3 файли, що містять інформацію про пакети UDP, TCP та ICMP відповідно. Проаналізувавши ці дані, можна отримати кількісні характеристики циркулюючого трафіку за період часу 20 хвилин з інтервалом у 10 секунд в стані «норма» та у стані «атака».

Вихідною змінною є «CLASS», що приймає 2 значення: «норма» або «атака».

На вхід НМ подавалось 120 спостережень, що властиві імітаційній мережі у стані «норма» та така ж кількість спостережень у стані «атака» (кількість спостережень обирається однаковою або із близькими значеннями, щоб

кількісна різниця змінних різних класів була невеликою для більшої стійкості НМ).

Середні значення визначених загальних критеріїв виявлення DoS-атак сформованої навчальної вибірки представлена у вигляді гістограми (рис. 10.9). Це дає змогу оцінити, наскільки вони відрізняються у стані мережі «норма» та стані «атака» для заданих умов проведення експерименту.

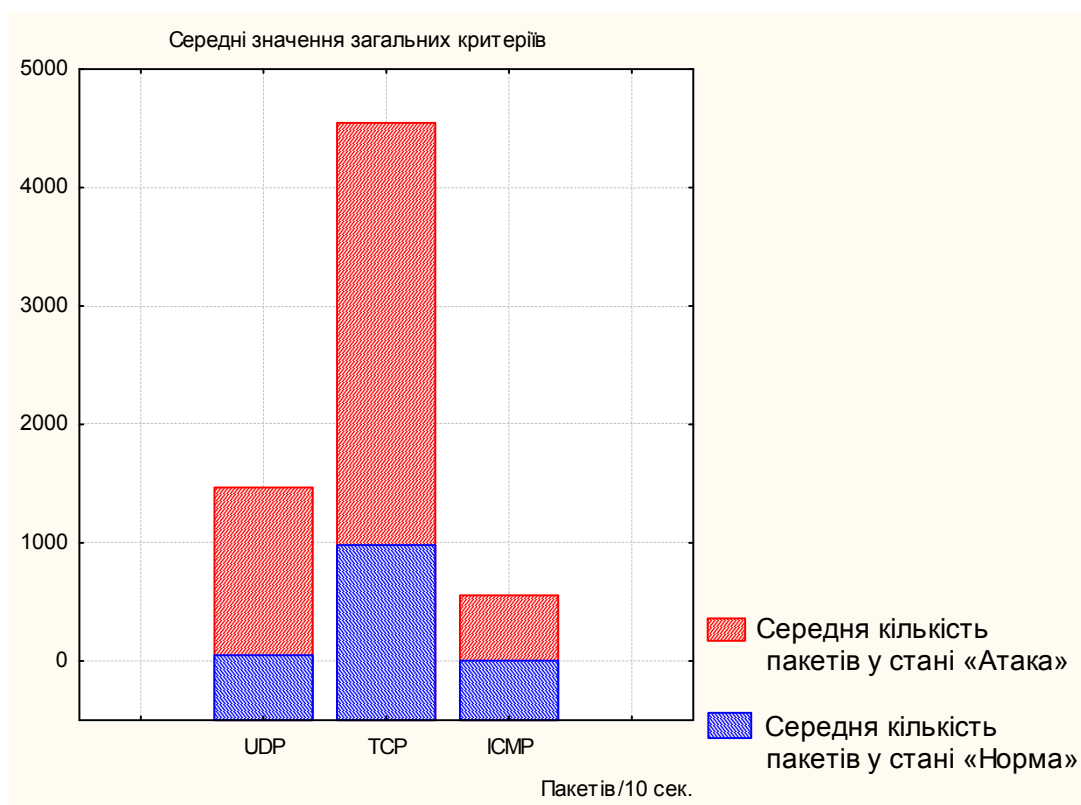


Рис. 10.9. Середні значення загальних критеріїв у стані «норма» та «атака»

10.4.2. Критерії ефективності нейронних мереж.

Для порівняльного аналізу НМ різних архітектур, що використовуються для класифікації даних при виявленні DoS-атак, обрані наступні критерії:

1. Вірогідність виявлення DoS-атак. Даний критерій характеризує достовірність виявлення реальної атаки системою IDPS і визначає ефективність роботи IDPS в цілому.

2. Рівень хибних спрацьовувань. Даний критерій визначає рівень виявлення системою IDPS DoS-атак, коли насправді вони відсутні. Це може статися, наприклад, внаслідок зростання мережевої активності легітимних користувачів. Якщо рівень хибних спрацьовувань дуже великий, то IDPS, відповідно, сповіщує адміністратора безпеки або формує реакцію у відповідь, внаслідок чого нормальна робота мережі з точки зору користувачів теж порушується.

3. Стійкість НМ. Даний критерій характеризує незалежність виходу НМ від системи, що захищається (для однакових ситуацій на вході, НМ повинна давати однаковий вихід незалежно від захищеної системи).

4. Швидкість роботи НМ. Даний критерій визначає кількість перевіреного IDPS трафіку за одиницю часу. Сучасна IDPS повинна забезпечувати пропускну здатність від 100 Мб/с до 1 Гб/с і більше. Тому цей критерій може впливати на здатність системи до масштабування.

5. Швидкість навчання НМ. Даний критерій характеризує здатність системи IDPS оперативно навчатися, перенавчатися або донавчатися. У процесі функціонування IDPS, що використовує в якості аналізатора НМ, її адміністратору доведеться аналізувати та відповідним чином корегувати здатність системи виявляти атаки або ігнорувати потоки трафіку при значній активності легітимних користувачів. Це робиться саме через процедуру перенавчання НМ відповідно до нових даних.

10.4.3. Дослідження ефективності нейронних мереж для виявлення DoS-атак.

Задача виявлення мережевих атак відноситься до задачі класифікації даних за 2 групами – «норма» або «атака».

Для вирішення задачі класифікації даних розглянемо використання НМ у вигляді багат шарового перцептрона (БП), НМ із РБФ, ймовірнісну НМ та лінійну НМ.

Для БП шляхом імітаційного моделювання встановлено, що для даної задачі класифікації достатньо одного проміжного шару з 7 нейронами, оскільки похибка навчання та контрольна похибка при збільшенні числа нейронів зменшуються незначно у процесі навчання.

При цьому функція активації для першого шару лінійна, для другого і третього – логістична. Мінімальне вихідне значення, при якому спостереження вважатиметься приналежним класу «атака» дорівнює 0,5, так само як і максимальне вихідне значення, при якому спостереження буде віднесене до класу «норма» дорівнює 0,5. Результати роботи БП представлені у таблиці 10.4.

Результати класифікації за допомогою НМ БП 3-7-1

Таблиця 10.4

Показник	Клас	
	«норма»	«атака»
Кількість випробувань	120	120
Розпізнано:		
- вірно	119	118
- помилково	1	2
- не відомо	0	0
Доля розпізнаних (%):		
- вірно	99,17	98,33
- помилково	0,83	1,67
- не відомо	0	0

НМ топології БП демонструє прийнятні показники виявлення нормального стану мережі, а також стану мережі при атаці. Тестування даної НМ БП 3-7-1 на екзаменаційній вибірці, що містить 40 спостережень у стані «норма» та «атака», показало, що вона правильно класифікувала всі спостереження.

Для НМ із РБФ в ході моделювання встановлено, що для даної задачі класифікації достатньо 11 нейронів в проміжному шарі. Мережа навчалася за алгоритмом С-середніх. Прийнятний баланс віднесення до класу «атака» вдалося досягти, встановивши поріг прийняття 0,35, а поріг відхилення відповідно 0,65 (табл. 10.5).

Результати класифікації НМ із РБФ 3-11-1

Таблиця 10.5

Показник	Клас	
	«норма»	«атака»
Кількість випробувань	120	120
Розпізнано:		
- вірно	120	107
- помилково	0	13
- не відомо	0	0
Доля розпізнаних (%):		
- вірно	100	89,17
- помилково	0	10,83
- не відомо	0	0

НМ із РБФ демонструє дещо гірші показники виявлення атак, ніж БП, проте вона більш стійка у випадку визначення нормального стану мережі.

Тестування даної НМ РБФ 3-11-1 на екзаменаційній вибірці показало, що НМ із РБФ помилилась 1 раз, класифікувавши стан «атака», як стан «норма».

При вирішенні задачі класифікації ймовірнісна НМ оцінює щільність ймовірності для кожного класу, порівнює між собою ймовірності належності різним класам і обирає з них найбільш ймовірний.

Шляхом моделювання встановлено, що найбільш прийнятні результати ймовірнісна НМ дає при значенні параметра згладжування 0,01, а також порозі прийняття 0,4 та відхилення 0,6 (табл. 10.6).

Показник	Клас	
	«норма»	«атака»
Кількість випробувань	120	120
Розпізнано:		
- вірно	116	81
- помилково	4	36
- не відомо	0	3
Доля розпізнаних (%):		
- вірно	96,67	67,50
- помилково	3,33	30,00
- не відомо	0	2,50

Ймовірнісна НМ добре визначає стан мережі «норма», але стан «атака» було виявлено помилково у третині випадків. Крім цього 2,5 % спостережень, що належать до класу «атака» не класифіковано взагалі.

Тестування ймовірнісної НМ на екзаменаційній вибірці показало, що з 40 спостережень ймовірнісна НМ правильно класифікувала 30, а в 10 помилилася.

Лінійна мережа являє собою НМ без проміжних шарів, що містить в вихідному шарі тільки елементи з лінійною функцією активації.

У процесі навчання найкраще співвідношення показників виявлення станів «норма» та «атака» досягнуто при встановленому порозі прийняття рівному 0,47 та відхилення 0,57 (табл. 10.7).

Тестування лінійної НМ на екзаменаційній вибірці показало, що з 40 тестових спостережень вона неправильно класифікувала 12, що є досить поганим результатом.

Проаналізувавши отримані результати роботи чотирьох НМ (рис. 10.10), можна зробити висновок, що найбільш прийнятну класифікацію даних при

виявленні DoS-атак отримано за допомогою БП. Саме ця НМ дозволяє з високою долею ймовірності (біля 98%) виявити факт наявності атаки і при цьому забезпечує досить низький рівень хибних виявлень (біля 1%), доволі швидкий процес навчання, обробки даних та стійкість роботи.

Результати класифікації лінійною НМ 3-1

Таблиця 10.7

Показник	Клас	
	«норма»	«атака»
Кількість випробувань	120	120
Розпізнано:		
- вірно	86	94
- помилково	34	26
- не відомо	0	0
Доля розпізнаних (%):		
- вірно	71,67	78,33
- помилково	28,33	21,67
- не відомо	0	0

Наступною за сукупністю показників роботи є НМ із РБФ. Ця мережа дозволяє виявляти атаку з ймовірністю біля 89% у сукупності з нульовим рівнем хибних виявлень. До переваг РБФ перед БП можна віднести те, що вона навчається на порядок швидше, а до недоліків – меншу стійкість.

Згідно з отриманими результатами, ймовірнісна НМ здатна виявляти наявність атаки з точністю біля 67,5 %. При цьому рівень хибних виявлень біля 3,5%. Дана мережа має найкращі показники за швидкістю навчання, але потребує багато пам'яті та повільно працює.

Рівень класифікації факту атаки лінійною НМ становить біля 78%, а рівень хибних виявлень – 28%. Слід зазначити, що відносна простота лінійної НМ не дає прийнятних результатів при виявленні атак.

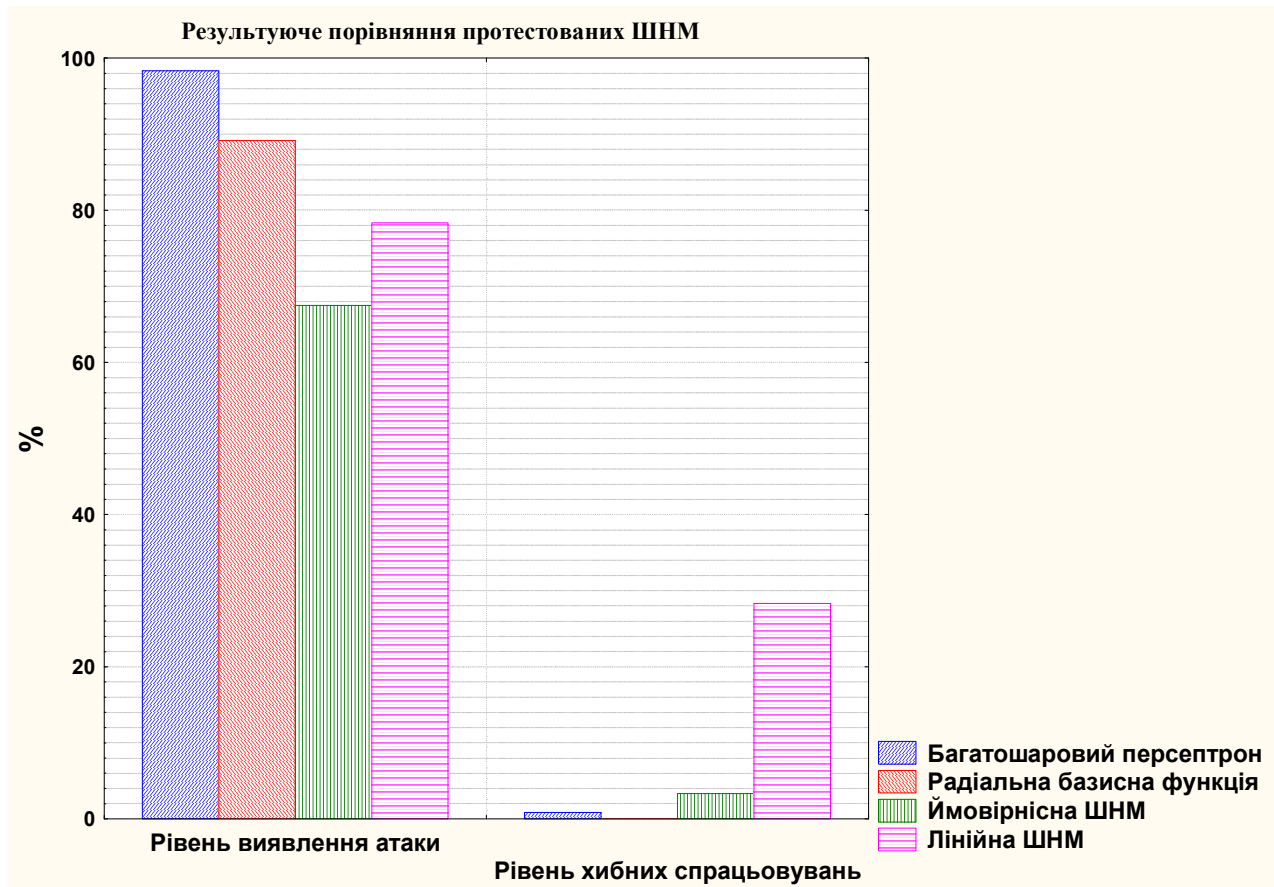


Рис. 10.10. Результати тестування НМ

10.5. Нечітка маршрутизація пакетного трафіка з використанням мурашиного алгоритму навчання

Вдосконалення методів керування в телекомунікаційних мережах грає значну роль в забезпеченні якості обслуговування при передачі даних. У сукупності з елементами керування потоками даних маршрутизація визначає функціонування всієї мережі щодо заданої якості і кількості послуг, що надаються. Кожен вузол мережі має маршрутизатор, який містить таблицю доступних напрямів, по яких може бути переданий пакет даних на шляху до вузла призначення.

Критеріями при виборі маршруту є час передачі даних, надійність передачі, ширина смуги пропускання каналу, завантаження мережевого буфера тощо.

Необхідність обліку чинників зводить процес маршрутизації до рішення задачі оптимального розподілу ресурсів мережі, одним з перспективних методів якого є використання систем нечіткого висновку з оптимізацією її параметрів на основі алгоритмів ройового інтелекту, зокрема, мурашиних алгоритмів (МА).

10.5.1. Протоколи маршрутизації і алгоритми керування трафіком в IP-мережах.

Архітектура протоколів TCP/IP призначена для мережі, що складається із сполучених маршрутизаторами один з одним окремих різнорідних пакетних підмереж, кожна з яких працює відповідно до своїх специфічних характеристик.

Кожен пакет включає інформацію, що ідентифікує вузол призначення (термінал), тобто кожен пакет інформаційного потоку обробляється (маршрутизується) окремо від інших. Таким чином, маршрутизація в технології IP є основним чинником, що впливає на продуктивність і ефективність мережі в цілому.

Кожен маршрутизатор використовує таблицю маршрутизації, в якій відбита топологія мережі на даний момент часу. У загальному випадку таблиця містить адреси мережі призначення, адреси наступного вузла на шляху до цієї мережі і метрику (вартість) маршруту. Формування і оновлення таблиці здійснюється за допомогою відповідних протоколів і алгоритмів маршрутизації.

Найбільш популярні протоколи динамічної маршрутизації, в яких таблиця маршрутизації заповнюється відповідним алгоритмом маршрутизації. При цьому, маршрутизатори підтримують свої таблиці маршрутизації шляхом

обміну різними службовими повідомленнями, за рахунок чого відбувається періодичне оновлення всіх таблиць.

Алгоритми маршрутизації ґрунтуються на множині показників, згорнутих в один інтегральний показник — метрику. Метрика — специфічна структура даних, що містить в собі у вигляді компонент показники якості каналу. Оскільки компоненти метрики (показники) вельми різнорідні, то метрика, як правило, є величиною безрозмірною.

Метрики формуються, зазвичай, на основі наступних показників (табл. 10.8):

- довжина маршруту;
- надійність;
- затримка;
- ширина смуги пропускання;
- вартість зв'язку.

Склад метрик популярних протоколів маршрутизації Таблица 10.8

№	Показники	Алгоритми (протоколи)				
		RIP	BGP	OSPF	IGRP/ EIGRP	IS/IS
1	Довжина маршруту (число переходів)	*				
2	Надійність				x	x
3	Затримка				*	x
4	Ширина смуги пропускання			*	*	
5	Навантаження (завантаженість каналу)				x	
6	Вартість зв'язку					x
7	Задаються довільно		*			x

Тут «*» — враховується за умовчанням, а «х» — може бути врахований.

Аналіз табл. 10.8 показує, що протоколи маршрутизації не використовують всі допустимі показники маршруту для обчислення метрики. Це обумовлено складністю обліку показників при формуванні метрики.

10.5.2. Алгоритм маршрутизації на основі систем нечіткої логіки.

Поширеним протоколом, заснованим на дистанційно-векторному алгоритмі (DVA) є протокол RIP (Routing Information Protocol).

Його специфікація визначає діапазон значень метрики від 0 до 16. Якщо значення метрики рівне 0, то мережа призначення безпосередньо підключена до даного інтерфейсу, а значення метрики 16 відповідає недосяжності мережі призначення.

Тут метрика чисельно рівна значенню відстані між вузлами зв'язку. Іншим чинником, що робить вплив на вибір маршруту, може служити ступінь завантаженості буферної пам'яті вихідного інтерфейсу (маршрутизатора): іноді доцільно відправити пакет даних по довшому маршруту, на якому буфери відповідних інтерфейсів не переобтяжені. Тобто при маршрутизації доцільно розглядати два параметри: відстань і завантаженість буферної пам'яті відповідного маршрутизатора. Оскільки ступінь впливу кожного чинника на формування метрики складно обрати у вигляді конкретної функції, то цю залежність можна сформуванати на основі системи нечіткого висновку.

Критерії значущості параметрів, що враховуються при формуванні метрики, зазвичай, ґрунтуються на базі експертних оцінок.

10.5.3. Навчання системи нечіткого висновку на основі мурашиного алгоритму.

Вибір форми і виду функцій належності (ФН) системи нечіткого висновку залежить від апріорного знання ступеня впливу даного параметра на точність обчислюваної метрики. Виходячи з цього, можна виділити два способи вибору

форми ФН: експертне оцінювання, яке припускає вибір форми ФН, виходячи із ступеня важливості того або іншого параметра, а також навчання системи нечіткого висновку шляхом оптимізації параметрів ФН.

Проблема навчання параметрів нечітких моделей вирішується методами оптимізації нелінійних функцій.

Труднощі застосування класичних методів оптимізації нелінійних функцій змусили фахівців звернутися до таких метаевристичних засобів, як алгоритм імітації відпалу, генетичні і мурашині алгоритми тощо (табл. 10.9).

Метаевристичні алгоритми оптимізації

Таблиця 10.9

Алгоритм	Параметри	Точність апроксимації	Час пошуку
Алгоритм відпалу	Температура, закон охолодження, критерій допуску	Низька	Дуже велике
Дискретний МА	Феромон, число мурашок, емпіричні коефіцієнти	Невисока	Невелике
Безперервний МА	Число мурашок, розмір архіву рішень, емпіричні коефіцієнти	Висока	Невелике
Генетичний алгоритм	Розмір популяцій, селекція, мутація, схрещування	Невисока	Велике
Алгоритм частинок, що рояться	Розмір риючи, коефіцієнт інерції, константи прискорення	Висока	Велике

На основі аналізу табл. 10.9, для ідентифікації параметрів ФН був обраний дискретний МА. Його основу складає орієнтований граф із зваженими дугами, які є нормованими значеннями параметрів ФН.

10.5.4. Оцінка ефективності алгоритму маршрутизації.

В імітаційній моделі мережі (рис. 10.11) пакетний трафік направляється із мережі А в мережу В. Вхідна послідовність пакетів поступає на порівнювані моделі маршрутизаторів R1 і R2. Пакети передаються по трьом маршрутах з відомими відстанями. Термінальні пристрої мережі В (T1, T2) фіксують час доставки пакетів від маршрутизаторів.

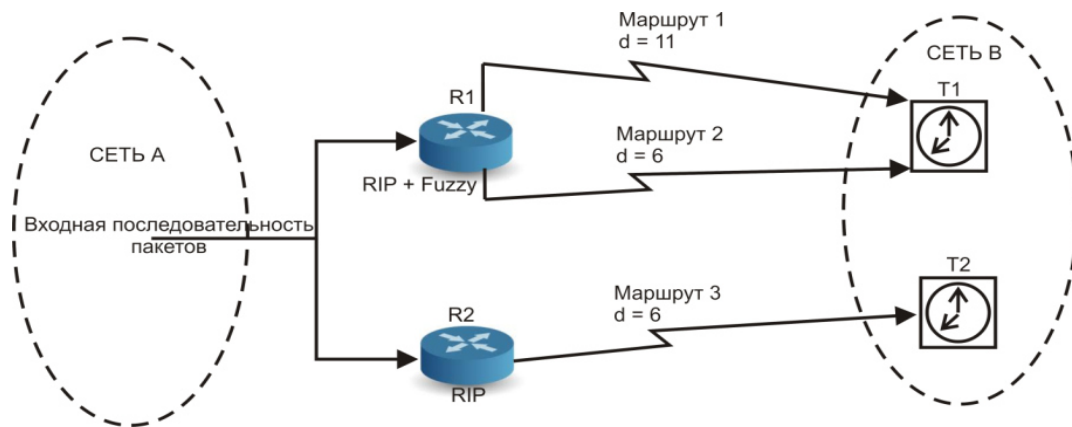


Рис. 10.11. Схема імітаційної моделі мережі

Тут R1 — модель маршрутизатора з системою нечіткого висновку, доповнюючою протокол RIP, а R2 — модель маршрутизатора, що працює з протоколом RIP.

Згідно протоколу RIP маршрутизатор R2 передає всі пакети по найкоротшому маршруту (маршрут 3 – $d = 6$). У свою чергу R1 може передавати пакети по двох маршрутах (маршрут 1 – $d = 11$, маршрут 2 – $d = 6$) в залежності від значення метрики, обчисленої за допомогою системи нечіткого висновку.

Результати моделювання показали, що середній час проходження пакетів через ненавчений маршрутизатор R1 на 5,5% менше, ніж через R2, а після навчання системи нечіткого висновку за допомогою дискретного МА середній час проходження пакетів через маршрутизатор R1 на 80% менше, ніж через R2.

Використання системи нечіткого висновку з МА навчання підвищує ефективність роботи маршрутизатора. При цьому додатковий інтерфейс передачі даних знижує завантаженість буфера очікування.

Істотною перевагою запропонованого алгоритму є його інваріантність до спостережуваних і керованих маршрутизаторів.

10.6. Нечіткі алгоритми IP-маршрутизації в телекомунікаційних мережах

10.6.1. Маршрутизація як процес керування.

Процес маршрутизації пакета пов'язаний (рис. 10.12) з керуванням пристроєм, що забезпечує вибір вихідного інтерфейсу маршрутизатора (перемикання трафіку). Керуючий вплив на маршрутизуючий пристрій подається у вигляді метрики (M). Метрика, в свою чергу залежить від параметрів маршрутизації пакета.

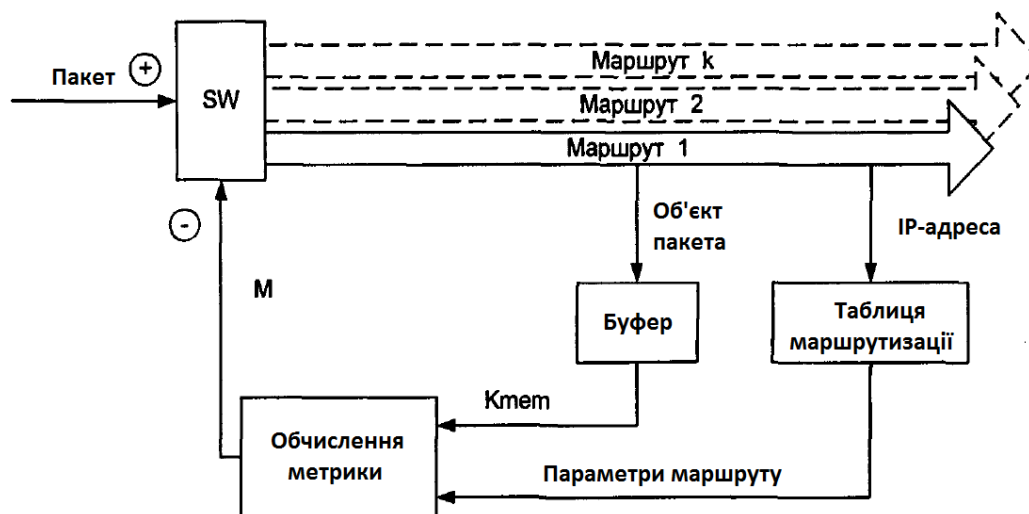


Рис. 10.12. Узагальнена схема керування трафіком

Тут комбінована метрика обчислюється як з урахуванням величини завантаженості буфера, так і з урахуванням параметрів маршруту. Така схема дозволяє проводити більш гнучке керування трафіком, що в кінцевому підсумку повинно підвищити ефективність функціонування мережі в цілому.

Керування трафіком на основі маршрутизації слід розглядати як взаємодію двох процесів: процесу передачі інформації і процесу вибору маршруту, що забезпечує його процес передачі. Ці процеси використовують загальні мережеві ресурси і, в першу чергу, смугу пропускання каналів.

10.6.2. Нечітка маршрутизація в рамках протокола RIP.

При моделюванні маршрутизатора в рамках протоколу RIP була прийнята наступна логічна структура контролера інтерфейсу (рис. 10.13), який складається з двох модулів: визначення параметрів і обчислення метрики.

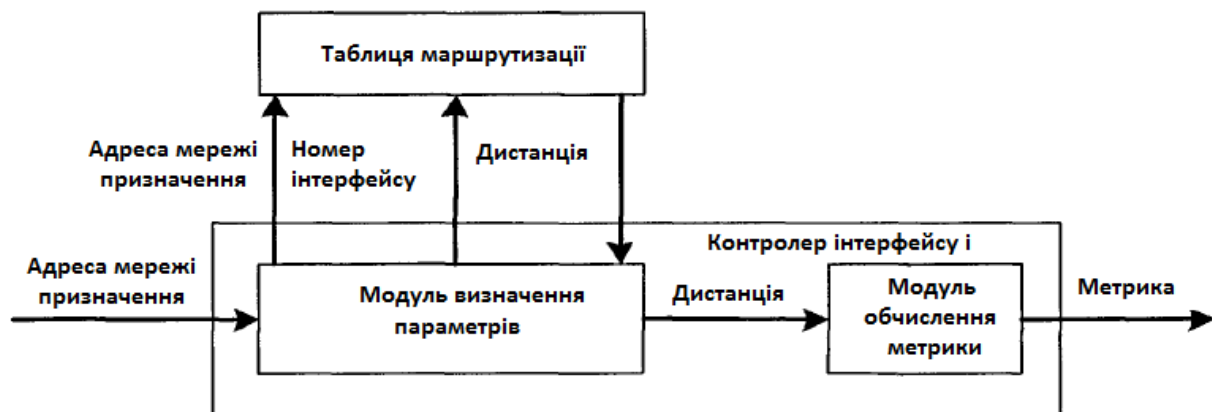


Рис. 10.13. Структура контролера вихідного інтерфейсу

Модуль визначення параметрів на основі адреси мережі призначення маршрутизації пакета опитує таблицю маршрутизації і вираховує значення параметрів, необхідних для обчислення метрики для даного вихідного інтерфейсу. На підставі цих параметрів модуль обчислення метрики визначає

значення метрики, виходячи з закладеного алгоритму. В даному випадку такий параметр один – дистанція.

Другим фактором, що суттєво впливає на вибір маршруту, може бути ступінь завантаженості буферної пам'яті вихідного інтерфейсу. Цей фактор визначає доцільність відкидання пакету. Таким чином, при маршрутизації розглядається два вхідних параметра: дистанцію і завантаженість буферної пам'яті відповідного інтерфейсу.

Алгоритм вибору рішення формується на основі нечіткої логіки, яка дозволяє враховувати вплив багатьох чинників, пов'язаних з проблемою маршрутизації. При цьому структура контролера вихідного інтерфейсу має вигляд, представлений на рис. 10.14.

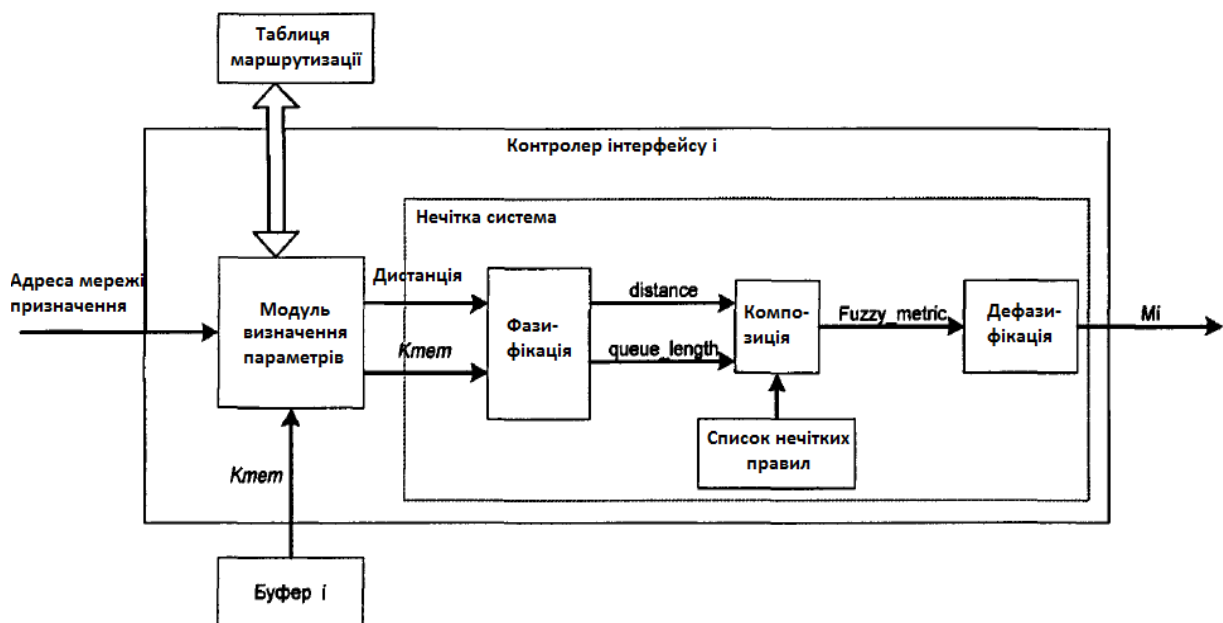


Рис. 10.14. Структура контролера вихідного інтерфейсу з використанням нечіткої логіки

Модуль обчислення метрики набуває характеру нечіткої системи, а сам процес визначення метрики складається з основних етапів формування нечіткого висновку: фазифікації, композиції та дефазифікації.

Відповідність "чітких" вхідних і вихідних змінних системи лінгвістичним змінним наведено в табл. 10.10.

Для кожної лінгвістичної змінної задається набір термів і відповідні їм нечіткі множини. Лінгвістична змінна *Fuzzy_metric* має в своєму складі два терма:

$$\text{Fuzzy_metric} = \{Little, Big\} \quad (10.5)$$

Нечіткі множини, що відповідають термам лінгвістичної змінної *Fuzzy_metric* визначаються відповідними характеристичними функціями, які в свою чергу задаються аналітично:

$$\text{Fuzzy_metric} (Big) = M/16; \quad (10.6)$$

$$\text{Fuzzy_metric} (Little) = 1-M/16; \quad (10.7)$$

де *M* – метрика (змінюється в діапазоні 0...16).

Змінні нечіткої системи маршрутизації

Таблиця 10.10

Ім'я змінної	Тип змінної	Діапазон значень	Відповідна лінгвістична змінна
Дистанція	Вхідна	0...16	distance
Кмет (коефіцієнт використання буферної пам'яті)	Вхідна	0...100%	queue_length
Метрика, М	Вихідна	0...16	Fuzzy_metric

Нечіткі правила зведені в табл. 10.11, а схема алгоритму з нечіткими множинами представлена на рис. 10.15.

Запропонована таблиця рішень потребує довізначення із залученням додаткових чинників, що уточнюють і доповнюють ситуацію. Це може бути врахування пропускну здатності каналу (пов'язаного з інтерфейсом), швидкості передачі даних, середнього коефіцієнту використання пам'яті тощо.

Fuzzy_metric		Distance	
		<i>Little</i>	<i>Big</i>
Oueue_length	<i>Little</i>	<i>Little</i>	<i>Big</i>
	<i>Big</i>	<i>Big</i>	<i>Big</i>

Крім того, термальність лінгвістичних змінних може бути збільшена, наприклад $distance = \{little, middle, big\}$, що дозволяє більш детально реагувати на ситуацію, яка складається в маршрутизаторі.

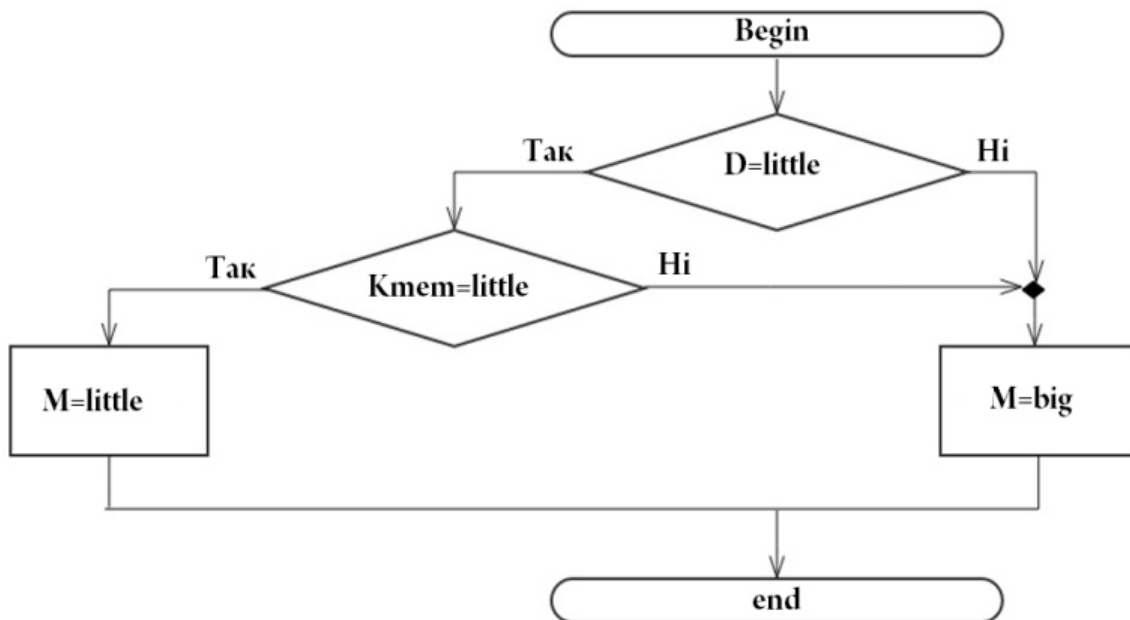


Рис. 10.15. Блок-схема алгоритму з нечіткими множинами

На основі розроблених правил проводиться обчислення значення метрики M . Залежність вихідної змінної від вхідної (поверхня відгуку нечіткої системи) при використанні алгоритму Мамдані наведена на рис. 10.16.

На цьому рисунку точка A визначає позитивне тривіальне рішення по вибору маршруту, точки B, C, D – негативні тривіальні рішення. Тривіальність

інтерпретується як абсолютне дотримання правил табл. 10.11 (з достовірністю 1).

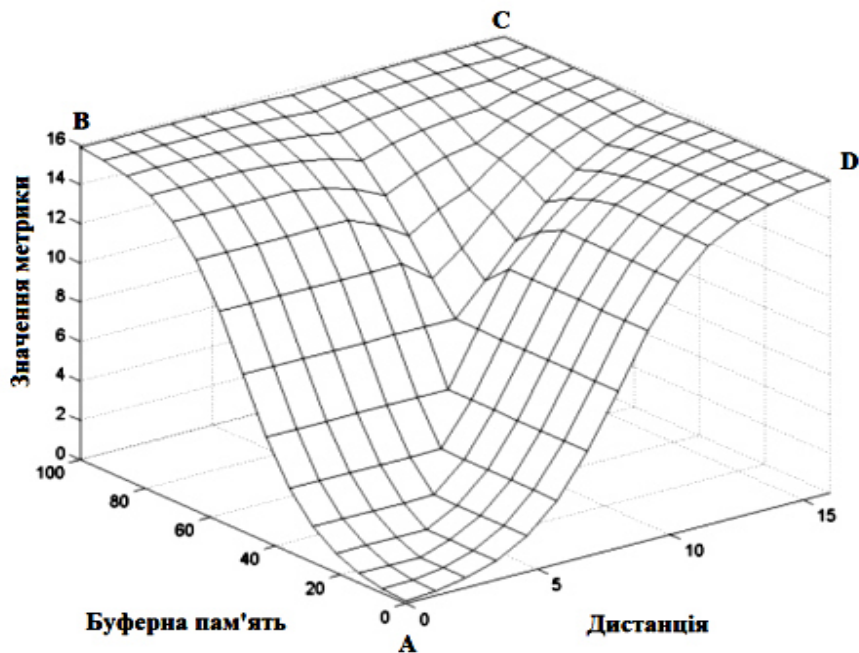


Рис. 10.16. Поверхня відгуку нечіткої системи протоколу RIP

10.6.3. Нечітка маршрутизація в рамках протоколу SNMP

Основною концепцією протоколу SNMP (Simple Network Management Protocol) є те, що вся необхідна для керування пристроєм інформація зберігається на самому пристрої в так званій Адміністративній Інформаційній Базі (MIB – Management Information Base), що являє собою набір змінних, які характеризують стан об'єкта керування. Ці змінні можуть відображати такі параметри, як кількість пакетів, оброблених пристроєм, стан його інтерфейсів, часу функціонування пристрою тощо.

SNMP – це протокол типу «запит-відповідь», тобто на кожен запит, що надійшов від контролера, агент повинен передати відповідь. Агент в протоколі SNMP – це обробляючий елемент, який забезпечує контролерам, розміщеним на керуючих станціях мережі, доступ до значень змінних MIB і тим самим дає їм можливість реалізовувати функції з керування та нагляду за пристроєм.

SNMP-протокол, включає мінімальний набір команд, проте дозволяє виконувати практично весь спектр завдань керування мережевими пристроями – від отримання інформації про місцезнаходження конкретного пристрою, до можливості виконувати його тестування.

Блок-схема даної програми (рис. 10.17) працює наступним чином.

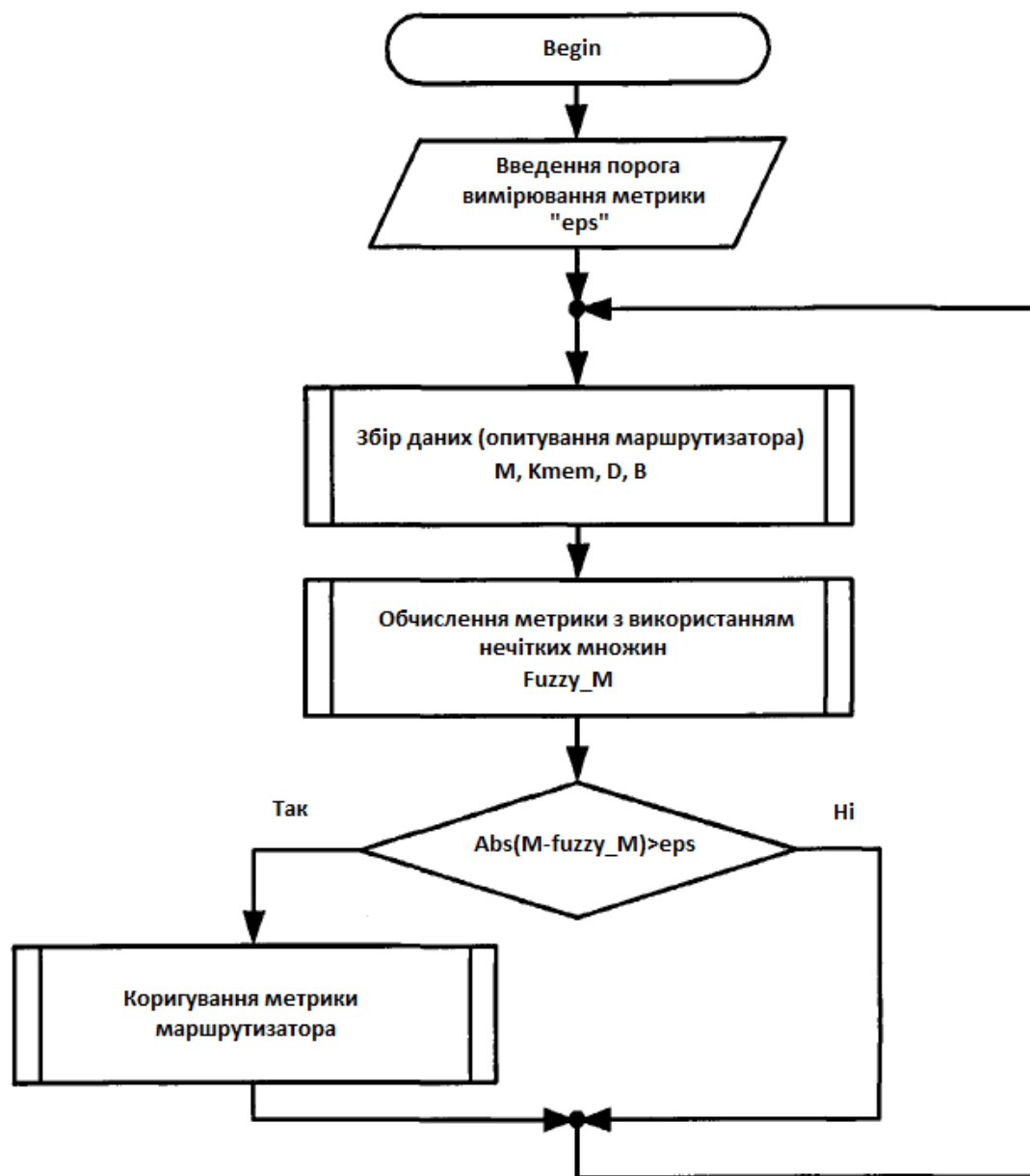


Рис. 10.17. Блок-схема алгоритма за протоколом SNMP

За допомогою протоколу SNMP проводиться опитування маршрутизатора і здійснюється збір параметрів маршруту. Далі за отриманими даними за

допомогою алгоритму нечіткого висновку розраховується метрика. Потім проводиться порівняння розрахованої метрики і метрики, отриманої від маршрутизатора. Якщо різниця цих метрик перевершує заданий вручну поріг, то за допомогою протоколу керування SNMP робиться коригувальний запит на маршрутизатор.

10.7. Система нечіткого вибору міжсегментного інтервалу в транспортному протоколі телекомунікаційної мережі

Інформаційні повідомлення, що передаються в сучасних телекомунікаційних мережах з комутацією пакетів, утворюють мультимедійний трафік (мовні повідомлення, відео) та трафік даних (сигнали керування, текстові повідомлення, різні зображення, інша графічна та таблична інформація). Для якісної передачі потокового трафіку необхідно мінімізувати затримку пакетів та її дисперсію (джитер), що вимагає гарантованого надання відповідної пропускної здатності мережі.

Керування інтенсивністю відправки даних джерелом здійснюється в умовах неповної, неточної інформації про стан елементів телекомунікаційної мережі, що для забезпечення ефективності доставки даних вимагає використання систем нечіткого висновку.

Аналіз існуючих методів керування інтенсивністю відправки даних джерелом показує, що, по-перше, жоден з методів не дозволяє визначити значення інтенсивності відправки даних, яке б відповідало миттєвим змінам доступної пропускної здатності мережі. По-друге, будь-яка втрата пакету інтерпретується протоколом транспортного рівня як мережне перевантаження. По-третє, аналітичні вирази для визначення величини інтенсивності відправки даних підібрані експериментально, а тому, враховують особливості передачі даних лише в окремих випадках. Виходячи з цього, обґрунтованим є застосування методу вибору міжсегментного інтервалу, який дозволяє

раціонально використовувати доступну для трафіка даних пропускну здатність мережі та підвищити оперативність доставки даних.

Метод адаптивної швидкості, заснований на керуванні міжсегментним інтервалом, передбачає виконання таких режимів: прискореного старту, мультиплікативного скидання, відновлення, тонкої настройки, другого варіанту мультиплікативного скидання.

Результати одного із імітаційних експериментів, виконаних із моделями, що реалізують фіксовану зміну міжсегментного інтервалу, керування вікном перевантаження та метод адаптивної швидкості представлені на рис.10.18.

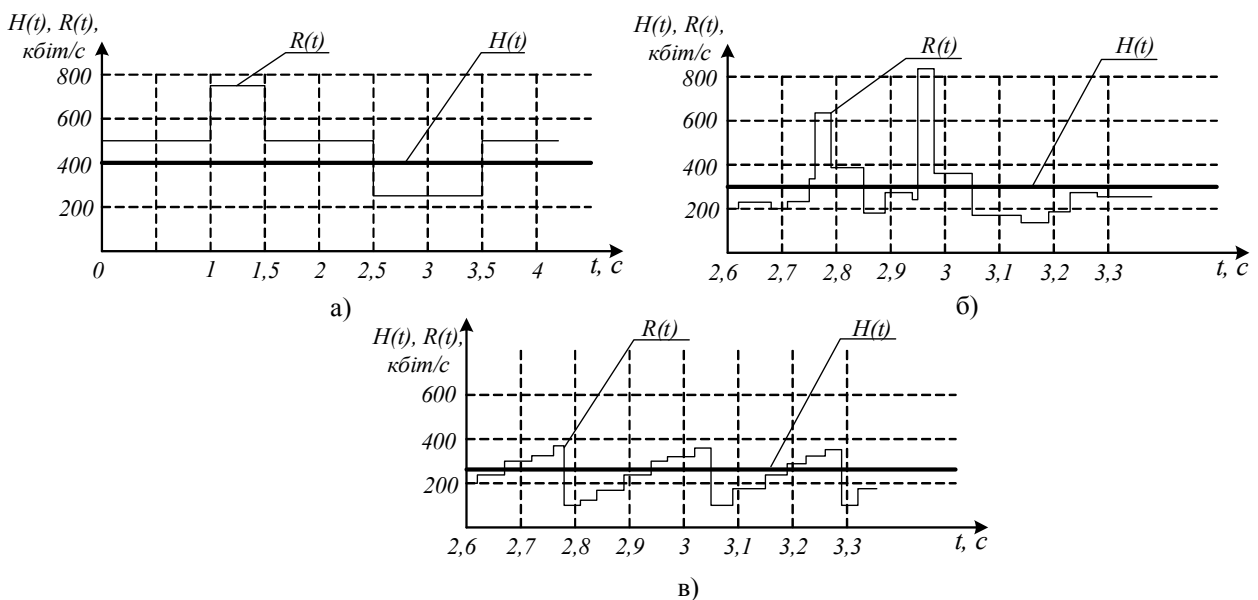


Рис.10.18. Доступна пропускну здатність $R(t)$ та інтенсивності відправки даних $H(t)$ при керуванні: із фіксованою зміною міжсегментного інтервалу (а); за вікном перевантаження (б); за методом адаптивної швидкості (в)

Вдосконалений адаптивний метод передбачає визначення поточного значення міжсегментного інтервалу як вихідного параметру системи нечіткого висновку. Особливості побудови такої системи визначаються складом її вхідних сигналів, аналіз яких дозволяє розрахувати поточне значення вихідної змінної.

Визначено, що найбільш чутливими до зміни стану мережі є такі параметри:

- попереднє значення величини ковзного середнього часу очікування квитанції;
- поточне значення величини ковзного середнього часу очікування квитанції;
- попереднє значення міжсегментного інтервалу.

Ці параметри використані в якості вхідних змінних системи нечіткого висновку на основі алгоритму Мамдані.

Синтез бази правил системи нечіткого висновку міжсегментного інтервалу здійснюється на основі чисельних даних, отриманих в результаті експериментальних досліджень. При цьому формування навчальної таблиці виконується із застосуванням різних варіантів спеціально створеного в мережі тестового навантаження.

Результати нечіткого вибору міжсегментного інтервалу в транспортному протоколі представлені на рис. 10.19.

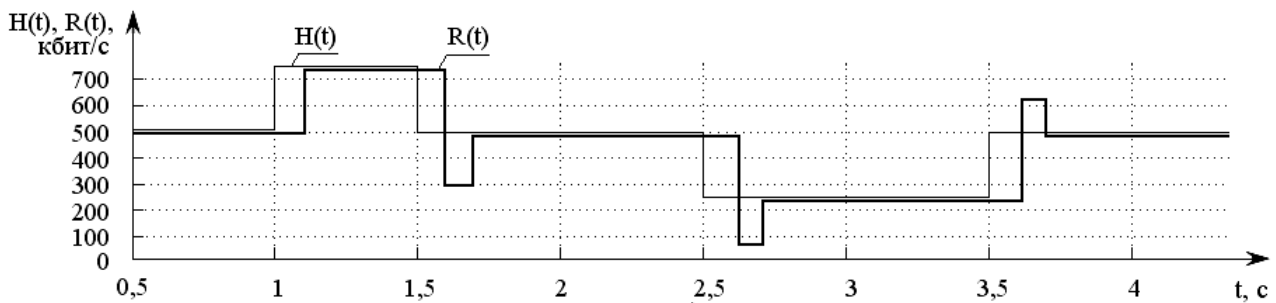


Рис. 10.19. Пропускна здатність $H(t)$ та інтенсивність відправки даних $R(t)$ при нечіткому виборі міжсегментного інтервалу

Аналіз результатів вибору міжсегментного інтервалу на основі системи нечіткого висновку показав, що інтенсивність відправки даних добре узгоджується з доступною пропускнуою здатністю телекомунікаційної мережі. Крім того, встановлено, що використання методу нечіткого вибору дозволяє

зменшити кількість повторних передач сегментів внаслідок мережних перевантажень та знизити на 16...39 % середній час передачі повідомлення заданого обсягу у порівнянні з іншими методами керування інтенсивністю відправки даних джерелом.

При цьому, використання нечіткого вибору міжсегментного інтервалу в транспортному протоколі телекомунікаційної мережі зменшує середній час доставки повідомлення T_n . Графіки залежності середнього часу доставки повідомлення T_n від середнього часу його передачі t_0 при різній кількості спроб його доставки n наведені на рис. 10.20.

Аналіз цих графіків показує, що із збільшенням кількості спроб n , необхідних для успішної доставки повідомлення, суттєво зростає середній час T_n . Тому, чим більше кількість спроб доставки повідомлення, тим меншим повинен бути середній час його передачі t_0 для забезпечення своєчасної доставки.

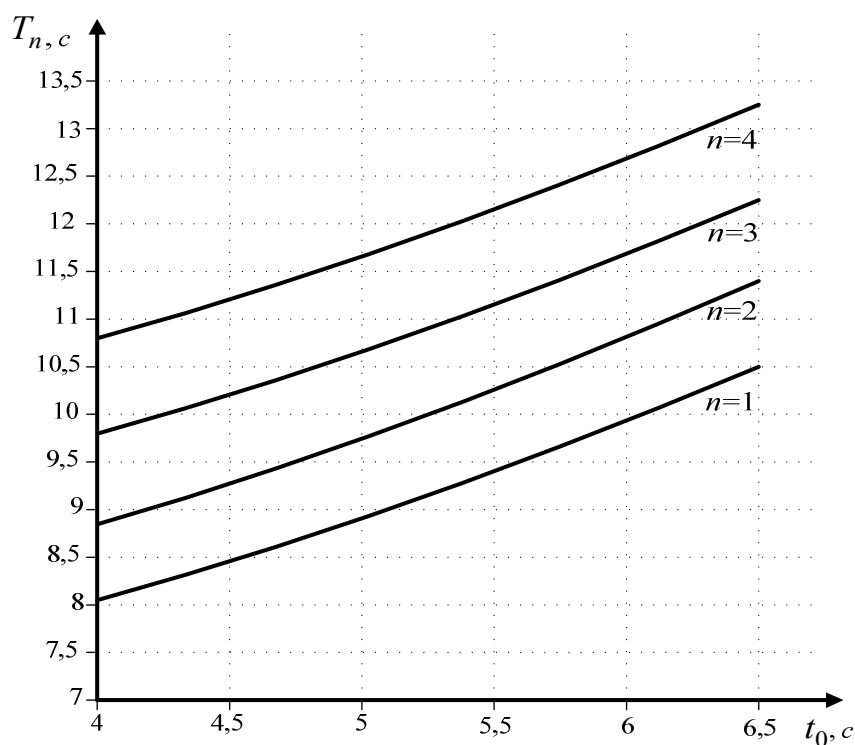


Рис. 10.20. Залежності середнього часу доставки інформаційного повідомлення від середнього часу його передачі

Загалом визначено, що нечіткий вибір міжсегментного інтервалу в транспортному протоколі телекомунікаційної мережі забезпечує зниження середнього часу доставки повідомлення на 6,2...20,5 % у порівнянні з результатами застосування інших методів керування інтенсивністю відправки даних джерелом.

10.8. Придушення шумів сигналів мобільного зв'язку з нейронечіткою ідентифікацією 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 широко використовуються методи систем штучного інтелекту, зокрема, адаптивні нейронні системи нечіткого висновку (Anfis). 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. 10.21.

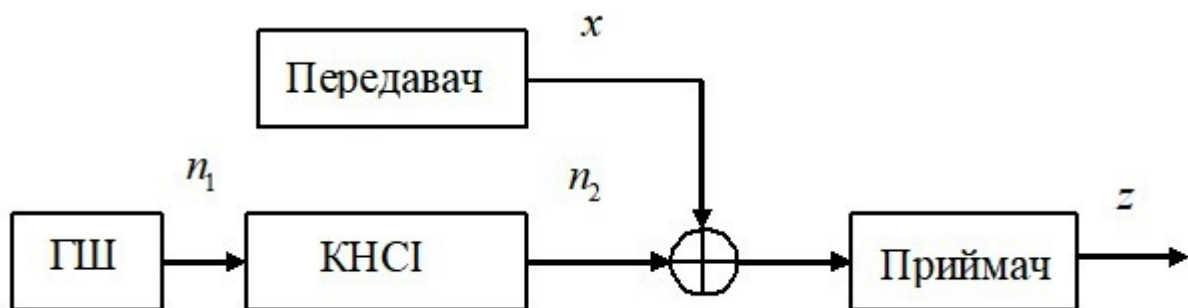


Рис. 10.21. Схема передачі радіосигналу мобільного зв'язку

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:

$$\varepsilon = \frac{\|x - \hat{x}\|}{\|x\|}. \quad (10.8)$$

Результати моделювання системи прийому з нейронечіткою ідентифікацією для різних типів передаваних сигналів наведені в табл. 10.12.

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. 10.13.

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

В моделях цієї таблиці k – це такт часу. Аналіз цієї таблиці показує, що використання систем з нейронечіткою ідентифікацією більш ефективно при нелінійних трактах розповсюдження сигналів.

Похибки прийому для різних моделей тракту розповсюдження Таблиця 10.13

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.

10.9. Модифіковані мурашині алгоритми оптимізації у бездротових сенсорних мережах

Останніми роками в області інформаційно-комунікаційних технологій велика увага приділяється бездротовим сенсорним мережам (БСМ). Такі мережі мають різні області застосування, наприклад, промисловий моніторинг, системи «Розумний будинок», системи розподіленого збору інформації тощо. При проектуванні і реалізації БСМ необхідно вирішувати безліч завдань, що відносяться до різних областей досліджень. Одним з основних завдань є забезпечення високої відмовостійкості, низького енергоспоживання мережі і збільшення тривалості функціонування БСМ.

З метою збереження енергії вузлами БСМ розробники розглядають питання позиціонування вузлів, кластеризації мережі і синтезу енергоефективної моделі збору інформації в БСМ. Для вирішення завдання автономності БСМ використовують також методи визначення часу включення передачі, багатоланкову передачу, попередню обробку і скорочення об'єму даних за рахунок їх компресії тощо.

Для скорочення часових витрат і ресурсоемності існуючих способів автоматичного вибору маршруту в мережах збору інформації використовуються алгоритми побудови структури БСМ, що самоорганізуються, засновані на застосуванні мурашиних, бджолиних і генетичних алгоритмів. Розглянемо тут модифікації МА, направлені на підвищення ефективності використання МА при оптимізації БСМ.

10.9.1. Мурашиний алгоритм із використанням semaфорів.

Метод semaфорів використовується з метою обмеження доступу до деяких вузлів, в першому випадку – із заданою кількістю потоків, в другому – до вузлів, що приймають аварійну сигналізацію, яка повинна передаватися з мінімальними затримками і втратами.

У МА введемо принцип semaфора з трьома станами. Хай вузол j БСМ передає аварійну сигналізацію, тоді мурашка не може на нього піти – semaфор «Червоний» і мурашка не відкладає по цій дузі феромон. У разі обслуговування вузлом потоку і semaфор «Жовтий», то мурашка може піти і відкласти феромон, але з меншим коефіцієнтом. У разі відсутності обслуговування потоку і semaфор «Зелений», то мурашка застосовує класичний метод.

У класичний вираз визначення вірогідності переходу від вузла i до вузла j для мурашки вводиться принцип semaфора шляхом правила:

$$P_{i,j} = \frac{\tau_{i,j}^\alpha \cdot w_{i,j}^{-\beta}}{\sum_{j \in M} \tau_{i,j}^\alpha \cdot w_{i,j}^{-\beta}} \cdot S, \quad (10.9)$$

де $\tau_{i,j}$ – кількість феромонів на ребрі від вузла i до вузла j ; $w_{i,j}$ – відстань між вузлами i та j ; M – кількість вузлів, які можна відвідати мурашці при переході від вузла i до j ; α – коефіцієнт впливу кількості феромонів $\tau_{i,j}$; β – коефіцієнт впливу відстані між вузлами i та j на вибір мурашки; S – стан semaфора, який приймає значення: $S = 0$ при «Червоному» сигналі; $S = 1$ при «Зеленому» сигналі; $0 < S < 1$ при «Жовтому» сигналі.

Запропонований принцип роботи МА дозволяє організувати передачу аварійної інформації з найвищим пріоритетом.

У класичній реалізації МА не враховується можлива завантаженість вузла. Для усунення цього недоліку використовується коефіцієнт S , який залежить від завантаженості вузла інформаційними потоками:

$$S_j(f) = \sum_{m \in \Omega} f_{j,m}(\varphi_{j,m}, r_{j,m}) = \sum_{m \in \Omega} \varphi_{j,m} \cdot r_{j,m} \quad (10.10)$$

де S_j – стан семафора на вузлі j ; $f_{j,m}$ – потік, який приходить на вузол j від вузла m ; Ω – множина вузлів, з яких приходить потік на вузол j ; $\varphi_{j,m}$ – пропускна спроможність каналу між вузлом j та m ; $r_{j,m}$ – середня інтенсивність потоку, що приходить по каналу j,m на вузол j .

У БСМ характеристики трафіку, а отже і інформаційних потоків, значною мірою залежать від моделі збору інформації або сценарію роботи і принципу передачі даних. Враховуючи, що області застосування БСМ різні, то трафік усередині мережі часто приймає стохастичний характер і має ознаки самоподоби.

Як видно з виразу (10.10) на інтенсивність потоку впливає φ – пропускна спроможність каналу між вузлами і r – середня інтенсивність потоку, яка приходить по каналу.

Для забезпечення необхідної якості обслуговування (QoS) необхідне виконання умови:

$$\sum_{m \in \Omega} \varphi_{j,m} \cdot r_{j,m} \leq \sum_{m \in N} \psi_{j,m}, \quad (10.11)$$

де $\psi_{j,m}$ – задана пропускна спроможність каналу від вузла j до вузла m ; N – канали, по яких приходить потік на вузол j .

Слід зазначити, що якщо у виразі (10.9) $\alpha = 0$, то буде обраний найближчий вузол, що відповідає жадібному алгоритму в класичній теорії МА. Якщо $\beta = 0$, тоді працює лише феромонне посилення, що спричиняє за собою швидку звироднілість маршрутів до одного субоптимального рішення.

Запропонований принцип роботи МА дозволяє організувати передачу аварійної інформації в мережі, а також врахувати при маршрутизації потокове навантаження на вузли БСМ.

10.9.2. Мурашиний алгоритм із розділенням часу збирання інформації.

У МА окрім вірогідності переходу від одного вузла до іншого, для заданої групи мурашок-дослідників, вводять поняття слід феромонів.

Після того, як мурашка успішно проходить маршрут по вузлах, вона залишає на всіх пройдених ребрах слід феромонів, обернено пропорційний довжині пройденого шляху:

$$\tau_{ij,k} = \begin{cases} \frac{\chi}{L_k}, & (i, j) \in T; \\ 0, & (i, j) \notin T, \end{cases} \quad (10.12)$$

де χ – регульований параметр; k – порядковий номер мурашки; L_k – довжина шляху, що пройдена мурашкою k ; $\tau_{ij,k}$ – кількість відкладеного феромону мурашкою k .

Оновлення феромонів в класичному випадку відбувається за правилом:

$$\tau_{ij} = (1 - p)\tau_{i,j} + \Delta\tau_{ij}, \quad (10.13)$$

де p – швидкість випаровування феромону; $\tau_{i,j}$ – кількість феромонів на ребрі від вузла i до вузла j .

Запропоновано в класичний МА ввести поняття «Ранок», «День», «Вечір» та «Ніч». «День» – стан, коли вузол може передати дані, а вузол-приймач може прийняти ці дані. «Ніч» – вузол-одержувач не може прийняти дані, оскільки можливо зайнятий прийомом інших даних, вузлу-відправникові забороняється передавати дані. «Вечір» – вузлу-одержувачеві прийшла нова заявка або він йде в «сплячий» режим. «Ранок» – вузол-відправник зібрав необхідні дані і готовий їх передати.

Дану ідею реалізовано в МА таким чином. Нехай мурашки проводять дослідження в мережі тільки «Вдень», а «Вранці» йде процес оновлення феромонів (10.13) :

$$p_k = \left(\sum_{\delta=1}^{R_k} s_{\delta k} \right)^{-1} \cdot \gamma, \quad (10.14)$$

де p_k – швидкість випаровування феромону мурашки k ; $s_{\delta k}$ – червоний семафор на шляху мурашки k ; R_k – кількість семафорів, що зустрічалися

мурашці k ; γ – коефіцієнт впливу кількості червоних семафорів на шляху мурашки k .

У режимах «Вечір» і «Ніч» дослідження мережі не проводяться. Якщо використовувати локальний час для кожного сегменту мережі, то виходить, що в різних сегментах мережі процес оновлення феромонів буде несинхронним. Такий підхід можна використовувати у разі, коли БСМ є достатньо великою мережею, оскільки МА ефективні при on-line оптимізації процесів в розподілених нестационарних системах, якими є трафік в БСМ.

Контрольні запитання

1. Розкрийте структуру побудови вейнету з блоком із нескінченною імпульсною характеристикою.
2. Які особливості фільтрації голосу за допомогою вейвнет?
3. Які особливості апроксимації складних функцій за допомогою вейвнет?
4. Які особливості ідентифікація нелінійного динамічного процесу із нормально розподіленим шумом?
5. В чому полягає оцінка DLP-систем із використанням засобів нечіткої логіки?
6. В чому полягає оцінка DLP-систем із використанням штучних нейронних мереж?
7. Розкрийте зміст методики нечіткої оцінки технічного стану компютерної системи в задачах інформаційної безпеки.
8. Які інформативні ознаки DoS-атак в інформаційно-комунікаційних мережах?
9. Які критерії ефективності нейронних мереж щодо виявлення DoS-атак в інформаційно-комунікаційних мережах?
10. Яка ефективність різних типів нейронних мереж для виявлення DoS-атак?

11. В чому полягає нечітка маршрутизація пакетного трафіку з використанням мурашиного алгоритму навчання?
12. Розкрийте сутність нечітких алгоритмів IP-маршрутизації в телекомунікаційних мережах.
13. Яким чином працює система нечіткого вибору міжсегментного інтервалу в транспортному протоколі телекомунікаційної мережі?
14. Як здійснюється придушення шумів сигналів мобільного зв'язку з нейронечіткою ідентифікацією тракту передачі?
15. Які модифіковані мурашині алгоритми оптимізації використовуються в безпроводних сенсорних мережах?

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Анищенко В.С. Лекции по нелинейной динамике: Учеб. пособие для студ. вузов / В.С. Анищенко, Т.Е. Вадивасова. – М. – Саратов: Изд-во Сарат. Ун-та, 2010. – 322 с.
2. Анищенко В.С. Нелинейная динамика хаотических и стохастических систем / Анищенко В.С., Астахов В. В., Т.Е. Вадивасова. – Саратов: Изд-во Сарат. ун-та, 1999. – 312 с.
3. Даник Ю.Г. Основи кібербезпеки та кібероборони: підручник / Ю.Г Даник, П.П. Воробієнко, В.М. Чернега. – О.: ОНАЗ ім. О.С. Попова, 2018. – 228 с.
4. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем / А.Г. Ивахненко – Киев: Наук. думка, 1981. – 296 с.
5. Интеллектуальные системы автоматического управления / под. ред. И. М. Макарова, В. М. Лохина. – М. : Физматлит, 2001. – 425 с.
6. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учебное пособие / А.П. Карпенко. – М.: Издательство МГТУ им. Н.Э. Баумана, 2017. – 446 с.
7. Климонтович Ю. Л. Турбулентное движение и структура хаоса. Новый подход к статистической теории открытых систем. – М.: Наука, 1990. – 320 с.
8. Корнієнко В.І. Теорія систем керування: підручник / В.І. Корнієнко, О.Ю. Гусєв, О.В. Герасіна, В.П. Щокін. – Дніпро : НГУ, 2017. – 497 с.
9. Короновский А.А. Непрерывный вейвлетный анализ и его приложения / А.А. Короновский, А.Е. Храмов. — М.: Физматлит, 2003. – 175 с.
10. Кроновер Р. М. Фракталы и хаос в динамических системах. Основы теории / Р.М. Кроновер. – М.: Постмаркет, 2000. 265 с.
11. Кузнецов С.П. Динамический хаос / С.П. Кузнецов. – М.: Физматлит, 2002. – 296 с.
12. Курейчик В.В. Теория эволюционных вычислений / В.В. Курейчик, В.М. Курейчик, С.И. Родзин. – М.: Физматлит, 2012. – 260 с.
13. Круглов В.В. Нечеткая логика и искусственные нейронные сети / В.В. Круглов, М.И. Длин, Р.Ю. Голунов. – М.: Физматлит, 2001. – 224 с.
14. Магницкий Н.А. Хаос, структуры, вычислительный эксперимент / Н.А. Магницкий. – М.: ЛЕНАНД, 2011. – 320 с.
15. Нейрокомпьютеры и интеллектуальные роботы / Под ред. Н. М. Амосова. – Киев.: Наукова думка, 1991. – 412 с.
16. Прикладные нечеткие системы / Под ред. Т. Тэрано, К. Асаи, М. Сугено. – М.: Мир, 1993. – 297 с.
17. Розенблатт Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга / Ф. Розенблатт. – М.: Мир, 1965. – 304 с.
18. Ротштейн А. П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети / А. П. Ротштейн. – Винница: «УНІВЕРСУМ-Вінниця», 1999. – 320 с.

19. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский // пер. с польск. И Л Рудлинского – М.: Горячая линия-Телеком 2006 – 452 с.
20. Тенехов В.А. Нейросетевые системы управления / В.А. Тенехов, Д. В. Ефимов, И. Ю. Тюкин. – М.: ИПРЖР, 2002. – Кн. 8. – 650 с.
21. Уоссермен Ф. Нейрокомпьютерная техника / Ф. Уоссермен. – М.: Мир, 1992. – 240 с.
22. Федер Е. Фракталы / Е. Федер. – М.: Мир, 1991. – 254 с.
23. Хакен Г. Информация и самоорганизация / Г. Хакен. — М.: Мир, 1991. – 432 с.
24. Чуличков А.И. Математические модели нелинейной динамики / А.И. Чуличков. – М.: Физматлит, 2003. – 357 с.
25. Шеннон К.Э. Работы по теории информации и кибернетике / К. Э. Шеннон. – М.: ИЛ, 1963. – 829 с.
26. Шредер М. Фракталы, хаос, степенные законы / М. Шредер. – М.-И.: R&C Dynamics, 2001. – 527 с.
27. Штовба С.Д. Проектирование нечетких систем средствами Matlab / С.Д. Штовба. – М.: Горячая линия-Телеком, 2007. – 288 с.
28. Шустер Г. Детерминированный хаос. Введение / Г. Шустер. – М.: Мир, 1988. – 256 с.
29. Cohen L. Time-Frequency Analysis / L. Cohen. – Prentice Hall, 1995. – 299 pp.
30. Diks C. Nonlinear Time Series Analysis: Methods and Applications / C. Diks. – World Scientific Press, 1999. – 180 pp.
31. Flandrin P. Time-Frequency/Time-Scale Analysis / P. Flandrin. – Academic Press, 1999. – 386 pp.
32. Gusev O.Yu. Theory of adaptive filtration: tutorial / O.Yu. Gusev, V.M. Gorev, V.I. Korniienko. – Dnipro: NTU “DP”, 2019. – 156 pp.
33. Hurst H.E. Long-term storage capacity of reservoirs. / H.E. Hurst // Trans. Amer. Soc. Civ. Eng. – 1951. – Vol. 116. – Pp. 770-808.
34. Kantelhardt J.W. Multifractal detrended fluctuation analysis of nonstationary time series / J.W. Kantelhardt, S.A. Zschiegner, E. Koncsienly-Bunde, S. Havlin, A. Bunde and H. E. Stanley // Physica A. – 2002. –316. – Pp. 87–114.
35. Mandelbrot B.B. The Fractal Geometry of Nature / B.B. Mandelbrot. – W.H. Freeman. – 1983. – 537 pp.
36. Nelles O. Nonlinear System Identification: From Classical Approaches to Neural and Fuzzy Models / O. Nelles. – Berlin: Springer, 2001. – 785 pp.
37. The Hilbert-Huang Transform and Its Applications / Eds By Norden E. Huang, Samuel S. Shen. – World Scientific Publishing Company. – 2005. – 324 pp.

Навчальне видання

Корнієнко Валерій Іванович

Гусєв Олександр Юрійович

Герасіна Олександра Володимирівна

**ІНТЕЛЕКТУАЛЬНЕ МОДЕЛЮВАННЯ
НЕЛІНІЙНИХ ДИНАМІЧНИХ ПРОЦЕСІВ
У СИСТЕМАХ КЕРУВАННЯ, КІБЕРБЕЗПЕКИ, ТЕЛЕКОМУНІКАЦІЙ**

Підручник

За загальною редакцією професора В.І. Корнієнка

Видано в редакції авторів.

Підписано до друку 05.10.2020. Формат 30х42/4.
Папір офсетний. Ризографія. Ум. друк. арк. 29,5.
Обл.-вид. арк. 29,5. Тираж 60 пр. Зам. №

Підготовлено до друку та надруковано
у Національному технічному університеті
«Дніпровська політехніка».

Свідоцтво про внесення до Державного реєстру ДК № 1842 від 11.06.2004.

49005, м. Дніпро, просп. Д. Яворницького, 19.