Міністерство освіти і науки України
Національний університет «Одеська політехніка»
Інститут штучного інтелекту та робототехніки
Кафедра програмних та комп'ютерно-інтегрованих технологій

КОНСПЕКТ ЛЕКЦІЙ
з дисципліні:
Сучасні системи керування

2024

Міністерство освіти і науки України
Національний університет «Одеська політехніка»
Інститут штучного інтелекту та робототехніки
Кафедра програмних та комп'ютерно-інтегрованих технологій

КОНСПЕКТ ЛЕКЦІЙ
з дисципліні:
Сучасні системи керування

Затверджено
на засіданні кафедри програмних та
комп'ютерно-інтегрованих технологій
Протокол № 2 від 26.01.2024 р.

2024

Конспект лекцій до курсу «Сучасні системи керування» для здобувачів другого (магістерського) рівня вищої освіти за спеціальністю G7 - Автоматизація, комп'ютерно-інтегровані технології та робототехніка . / Укладач: Беглов К.В. – Одеса: НУ «Одеська політехніка», 2024. – 136 с.

Укладач: Беглов К.В, канд . техн. наук, доцент

Зміст

Тема 1. Експеріментальні дослідження об'єктів автоматизації

Лекція № 1. Планування та проведення активного експерименту

1. Отримання апріорної інформації про об'єкт дослідження
2. Проведення однофакторного дослідження
3. Проведення багатофакторного дослідження

1.1 Моделювання та експериментальні виміри.

Однією з головних завдань експерименту є отримання та перевірка математичної моделі об'єкта, що описує в кількісній формі взаємозв'язку між вхідними та вихідними параметрами об'єкта. Вхідні параметри, що можуть бути змінені, називають факторами. Для кожного фактора до вимірювання встановлюється область визначення, яка може бути безперервною та дискретною. Часто безперервна область визначення штучно дискретизується. Теоретично планування експерименту об'єкт досліджень прийнято представляти як «чорного ящика», яке математична модель описує функціональні зв'язок між вхідними і вихідними параметрами. Головними вимогами до математичних моделей об'єктів є *зручність математичного використання* та *інтерпретованість* моделі. Крім того, завжди повинні бути позначені *межі застосування* моделі. Якщо ці вимоги не виконуються, то при використанні та експериментальній перевірці моделей неминуче виникають *методичні похибки* та *похибки адекватності* , які будуть розглянуті в наступному розділі.

Можна виділити такі завдання перевірки моделей (рис.1.1):

1. Побудувати «чорну скриньку», яка буде відповідним чином відгукуватися на заданий вхідний вплив.

2. Маючи «чорну скриньку», знаючи вхідні та вихідні сигнали, отримати (змоделювати) його вміст.

Мал. 1.1

Суть процесу моделювання можна пояснити з прикладу аналізу електронної схеми, у якого будуть отримані певні вихідні сигнали. Можна перевірити модель, зібравши експериментальну схему та знявши реальні вихідні сигнали. При цьому неминучі розбіжності між сигналами модельними та реальними. Щоб з'ясувати причини розбіжності, потрібні експерименти з окремими елементами схеми.

Необхідне коригування моделі може бути виконане таким чином:

1. Перевірка розбіжностей — експериментальна перевірка характеристик всіх елементів та його порівняння з модельними.

2. Виправлення характеристик окремих елементів у вихідній моделі.

3. Зіставлення отриманих залежностей з експериментальними (вихідними).

Таким чином, побудова та перевірка моделі, що адекватно описує роботу електронної схеми, у загальному випадку вимагає дуже великої кількості експериментальних вимірювань. Планування експерименту дозволяє оптимізувати кількість вимірів.

Наприклад, електронна схема складається з транзисторів, резисторів, конденсаторів та котушок індуктивності. Якщо номінальні значення пасивних електронних елементів (резисторів, конденсаторів тощо) збігаються з їхніми реальними значеннями з необхідною точністю, то розбіжність між модельними та реальними сигналами найчастіше виникає через невідповідність реальних робочих характеристик активних елементів (транзисторів, мікросхем тощо). Тому досвідчені схемотехніки перевіряють лише окремі вузли схеми, по суті інтуїтивно плануючи експеримент виходячи зі свого досвіду і використовуючи апріорну інформацію.

Розглянемо приклад моделювання найпростішого чотириполюсника, що здійснює виділення огинаючої (детектування) радіосигналу (рис.1.2).

Чотирьохполюсник складається з двох найпростіших схем:

1. детектора на діоді Д з вихідним резистором R_1.

2. інтегруючого ланцюга R_2C.

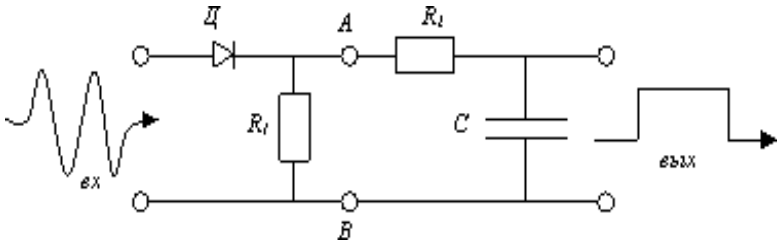
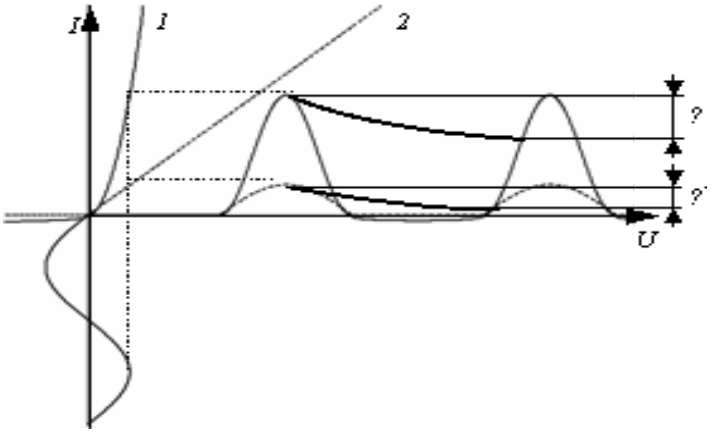


Рис.1.2

Сигнали на виході детектора *AB* та виході інтегруючого ланцюга показані на рис.1.3 . Тут криві 1 і 2 відповідають різним вольтамперним характеристикам (ВАХ) діода. Детектор відрізає негативні напівперіоди сигналу, а інтегруючий ланцюг – виділяє його огинаючу. Якість виділення огинаючої визначатиметься відхиленням Δ від «ідеального» сигналу.

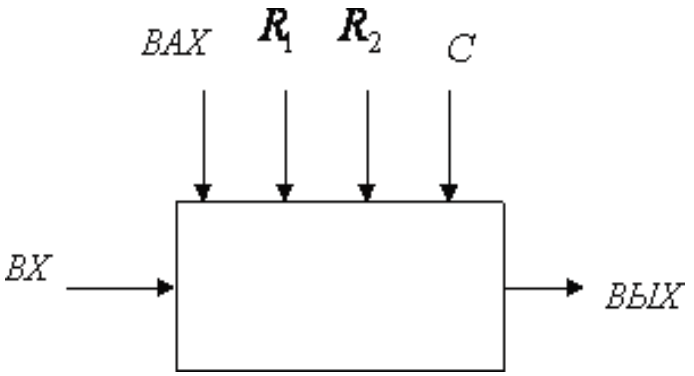


Мал. 1.3

Величина Δ у свою чергу залежить від характеристик, як детектора, так і ланцюга, що інтегрує. У детекторі вона визначатиметься вольтамперною характеристикою (ВАХ) діода $Д$, а в інтегруючому ланцюгу - співвідношенням між ємністю конденсатора C і опором R_2.

Як видно з рис.1.3 *амплітуда* вихідного сигналу детектора, відповідна ВАХ-1, вище, що неминуче призведе до збільшення Δ в результуючому сигналі. З іншого боку, зменшення ємності конденсатора інтегруючого ланцюга також призводить до збільшення Δ. При моделюванні схеми розбіжність між розрахунковими та реальними сигналами вимагає внесення коригування в характеристики, що задаються в моделі.

У випадку чотириполюсник може розглядатися як об'єкт, схема якого показано на рис.1.4. Характеристики окремих елементів схеми (ВАХ діода та величини інших пасивних елементів) можуть вважатися фіксованими параметрами (керуючими). Залежно від плану експерименту ці параметри можна розглядати як вхідні (чинники), які задаються дискретно.



Мал. 1.4

Експериментальні виміри прийнято розділяти на три основні види:

1) *прямі вимірювання* , при яких безпосередньо реєструються значення вимірюваної величини (наприклад, вимірювання напруги *U* вольтметром);

2) *непрямі виміри* (наприклад, вимірювання сили струму *I* амперметром, активного опору *R* омметром і розрахунок *U = IR*);

Тобто непрямі виміри - це отримання величини $y = f(x_1, x_2, \ldots)$ за виміряними значеннями $x_1, x_2, \ldots$.

3) *спільні виміри* (наприклад, вимірювання напруги U та сили струму I при різних значеннях I та побудова результуючої залежності $U = U(I)$);

Тобто спільні виміри — це виміри двох чи кількох неодноєменних величин для побудови залежності між ними.

Планування експерименту передбачає як оптимізацію числа вимірів, а й зменшення експериментальних похибок. Тому значну частину математичного апарату теорії планування експерименту становлять теорія помилок, теорія ймовірностей та математична статистика.

3.2. Пасивний та активний експеримент

Теорія передбачає, що експеримент може бути пасивним та активним.

При *пасивному експерименті* інформація про об'єкт, що досліджується, накопичується шляхом пасивного спостереження, тобто інформацію отримують в умовах звичайного функціонування об'єкта. Активний експеримент проводиться із застосуванням штучного на об'єкт за спеціальною програмою.

При пасивному експерименті існують лише чинники як вхідних контрольованих, але некерованих змінних, і експериментатор перебуває у становищі пасивного наблюдателя. Завдання планування в цьому випадку зводиться до оптимальної організації збору інформації та вирішення таких питань, як вибір кількості та частоти вимірів, вибір методу обробки результатів вимірів.

Найчастіше метою пасивного експерименту є побудова математичної моделі об'єкта, яка можна розглядати або як добре, або як погано організований об'єкт. У добре організованому об'єкті мають місце певні процеси, у яких взаємозв'язки вхідних та вихідних параметрів встановлюються у вигляді детермінованих функцій. Тому такі об'єкти називають детермінованими. Погано організовані чи *дифузні* об'єкти є статистичні моделі. p align="justify"> Методи дослідження з використанням таких моделей не вимагають детального вивчення механізму процесів і явищ, що протікають в об'єкті.

Прикладом пасивного експерименту може бути аналіз роботи схеми, яка не має входів, лише виходи, і вплинути на її роботу неможливо.

Хорошим прикладом пасивного експерименту з дифузним об'єктом є вимірювання метеорологічних параметрів (температури, швидкості вітру тощо) при природних катаклізмах.

Активний експеримент дозволяє швидше та ефективніше вирішувати завдання дослідження, але складніший, вимагає великих матеріальних витрат і може перешкодити нормальному ходу технологічного процесу. Іноді немає можливості проведення активного експерименту (наприклад, щодо явищ природи). Проте, враховуючи переваги активного експерименту, тоді, коли це можливо, перевагу надають йому.

При активному експерименті фактори мають бути керованими та незалежними.

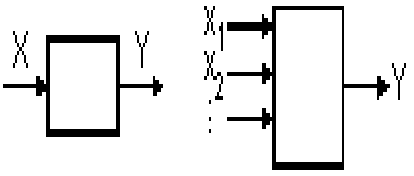
Активний експеримент передбачає можливість на перебіг процесу та вибору кожному досвіді рівнів чинників. При плануванні активного експерименту вирішується завдання раціонального вибору факторів, що істотно впливають на об'єкт дослідження, та визначення відповідної кількості дослідів, що проводяться. Збільшення числа включених у розгляд факторів призводить до різкого зростання кількості дослідів, зменшення - до суттєвого збільшення похибки досвіду. Чинник вважається заданим лише тоді, коли за його виборі вказується його область визначення – сукупність значень, які може набувати даний фактор. В експерименті використовується обмежена частина області визначення, що задається зазвичай у вигляді дискретної множини рівнів. Вибрані фактори повинні бути однозначно керованими та операційними, тобто піддаються регулюванню з підтримкою на заданому рівні протягом усього досвіду за дотримання

послідовності необхідних для цього дій. Повинна бути призначена також точність виміру факторів у вибраному діапазоні виміру.

Сукупності факторів мають відповідати вимогам сумісності та незалежності. Дотримання першого вимоги означає, що це комбінації чинників здійсненні і безпечні, другого - можливість встановлення чинника будь-якому рівні незалежно від рівнів інших чинників.

3.3. Однофакторний, багатофакторний та повний факторний експеримент

Різні види експериментів схематично представлені на рис.1.5.



а б у

Мал. 1.5

Однофакторний пасивний експеримент проводиться шляхом виконання ***n*** пар вимірювань дискретні моменти часу єдиного вхідного параметра ***x*** і відповідних значень вихідного параметра ***y*** (рис.1.5,а). Аналітична залежність між цими параметрами внаслідок випадкового характеру впливів, що обурює, розглядається у вигляді залежності математичного очікування ***y*** від значення ***x*** , що має назву регресійної. Метою однофакторного пасивного експерименту є побудова *регресійної моделі* - встановлення залежності ***y = f (x)*** .

Багатофакторний пасивний експеримент проводиться при контролі значень кількох вхідних параметрів x ***i*** (1.5,б) і його метою є встановлення залежності вихідного параметра від двох або більше змінних $_{y}$ ***= F (x1 , x2 , …)*** .

Повний факторний експеримент передбачає можливість керувати об'єктом по одному або декільком незалежним каналам (див. рис.1.5, в).

У загальному випадку схема експерименту може бути представлена у вигляді, представленому на рис.1.5, в. У схемі використовуються такі групи параметрів:

1. *Керуючі* (вхідні x_i)

2. *параметри стану* (вихідні ***y***)

3. *обурювальні впливи* (W_i)

При багатофакторному та повному факторному експерименті вихідних параметрів може бути декілька. Приклад такого пасивного багатофакторного експерименту буде розглянуто далі.

Керуючі параметри x_i є незалежними змінними, які можна змінювати для управління вихідними параметрами. Керуючі параметри називають *факторами* . Якщо ***i = 1*** (один керуючий параметр), то однофакторний експеримент. Багатофакторний експеримент відповідає кінцевому числу параметрів, що управляють. Повний факторний експеримент відповідає наявності впливів, що обурюють, в багатофакторному експерименті.

Діапазон зміни факторів x_i або число значень, яке вони можуть набувати називається *рівнем фактора* .

Повний факторний експеримент характеризується тим, що при фіксованих впливах W, що обурюють **,** W_i мінімальна кількість рівнів кожного фактора дорівнює двом. В цьому випадку, зафіксувавши всі фактори x_i крім одного, необхідно провести два виміри, що відповідають двом рівням цього фактора. Послідовно здійснюючи таку процедуру для кожного з факторів x_i отримаємо необхідне число ***N*** дослідів у повному факторному експерименті для реалізації всіх можливих поєднань рівнів факторів $N = 2^k$, де ***k*** – число факторів.

Наведемо класичний найпростіший приклад планування експерименту. Нехай нам необхідно зважити на терезах три тіла різної маси ***A , B , C*** за умови, що нульове положення ваг не відрегульоване. При складанні плану експерименту прийнято будувати *матрицю планування.* У таблиці 1.1 наведено план першого плану зважування. "1" і "-1" відповідають наявності або відсутності об'єкта на терезах.

Таблиця 1.1

№ досвіду	A	B	C	Результати зважування
1	-1	-1	-1	y_0
2	+1	-1	-1	y_1
3	-1	+1	-1	y_2
4	-1	-1	+1	y_3

Експеримент складається із чотирьох дослідів. При першому досвіді знімаються показання порожніх терезів і виставляється їхнє нульове положення, потім окремо зважується кожен з об'єктів. Розрахунок ваги та похибки вимірювань σ^2 кожного з тіл проводиться за такими формулами:

$$A = y_1 - y_0;$$

$$B = y_2 - y_0;$$

$$C = y_3 - y_0; \qquad (1.3.1)$$

$$\sigma^2[A] = \sigma^2[B] = \sigma^2[C] = 2\sigma^2[y].$$

Оскільки похибки незалежних вимірів складаються, а вага кожного об'єкта отримано внаслідок двох вимірів, похибка становить $2\sigma^2$.

Оптимально провести експеримент за схемою, показаною в Таблиці 1.2. У цьому випадку зважується окремо кожен з об'єктів та всі об'єкти разом. Безпосередній вимір похибки y_0 не проводять.

У цьому випадку виграш при проведенні експерименту полягає в тому, що маса кожного з об'єктів обчислюється за формулами (1.3.2), а дисперсія результатів виявляється вдвічі меншою. Цей результат виходить за рахунок того, що при другому плані експерименту усунення нуля вимірювальної апаратури (ваги) виключено.

Таблиця 1.2

№ досвіду	A	B	C	Результати зважування
1	+1	-1	-1	y_1
2	-1	+1	-1	y_2
3	-1	-1	+1	y_3
4	+1	+1	+1	y_4

$$A = \frac{(y_1 - y_2 - y_3 + y_4)}{2};$$

$$B = \frac{(-y_1 + y_2 - y_3 + y_4)}{2};$$

$$C = \frac{(-y_1 - y_2 + y_3 + y_4)}{2}; \qquad (1.3.2)$$

$$\sigma^2[A] = \sigma^2\left[\frac{(y_1 - y_2 - y_3 + y_4)}{2}\right] = \frac{4\sigma^2[y]}{4} = \sigma^2[y];$$

$$\sigma^2[B] = \sigma^2[C] = \sigma^2[y].$$

Цей приклад на простий випадок показує можливий виграш від зміни плану експерименту, т. е. планування експерименту дозволяє або зменшити кількість вимірювань, або збільшити їх точність.

Лекція № 2. Планування та проведення пасивного експерименту

1. Отримання апріорної інформації про об'єкт дослідження
2. Методи будови моделей об'єкта дослідження з використанням статистичних показників
3. Методи перевірки адекватності моделей

2.1. Кореляційні зв'язки та факторний аналіз даних при пасивному експерименті

При пасивному багатофакторному експерименті результатом є матриця експериментальних даних, що складається з набору n стовпців. Кожен із стовпців відповідає набору експериментальних значень окремого фактора. Якщо кожен фактор з номером k може набувати m значень $x_{1k}, x_{2k}, \ldots x_{mk}$, то матриця експериментальних даних матиме розмірність $n \times m$. Іноді фактори називають *ознаками* .

Виникає завдання встановлення статистичних зв'язків між рядами (стовпцями) даних. Завдання такого типу виникають при вимірі параметрів природних об'єктів (метеорологія, гідрологія тощо), де результати експерименту залежать від великої кількості параметрів, які значною мірою мають випадковий характер. Крім того, визначення статистичних зв'язків широко використовується у гуманітарних завданнях (економіці, соціології, психології тощо).

p align="justify"> Методи факторного аналізу дозволяють з деяким наближенням вирішити одну з найбільш поширених завдань наукового дослідження - задачу побудови тієї чи іншої схеми класифікації, тобто. компактного опису явища з урахуванням обробки великих інформаційних масивів.

Дослідження системи ознак, проведене на основі факторного аналізу, у ряді випадків дозволяє розкрити логічну структуру складного явища, відокремити взаємозалежні від незалежних ознак, перевірити чи висунути гіпотезу про взаємозв'язки у складній системі ознак.

Нехай є матриця ознак $\mathbf{Y}$ з елементами y_{ij}. Матриця вихідних ознак $\mathbf{Y}$ нормується в такий спосіб і перетворюється на матрицю $\mathbf{Z}$ з елементами

$$z_{ij} = \frac{y_{ij} - \bar{y}_i}{S_i}, \quad (6.1.1)$$

де $\bar{y}_i$ - Середнє значення показника в стовпці i; S_i - середньоквадратичне відхилення (СКО)

$$S_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2} \quad (6.1.2)$$

Для нормованої матриці $\mathbf{Z}$ виконуються такі умови:

$$\frac{1}{n}\sum_{j=1}^{n} z_{ij} = 0$$

$$\frac{1}{n-1}\sum_{j=1}^{n} z_{ij}^2 = 1 \quad (6.1.3)$$

Кореляційна матриця $\mathbf{R}$, що містить коефіцієнти кореляції r_{ij} між стовпцями матриці $\mathbf{Z}$, обчислюється за формулою

$$\frac{1}{n-1}\mathbf{Z}\mathbf{Z}^T = \mathbf{R} \quad (6.1.4)$$

Матриця $\mathbf{R}$ - Симетрична матриця.

Метою факторного аналізу є представлення величини z_{ij} у вигляді лінійної комбінації кількох змінних (чинників):

$$z_{ij} = a_{i1}p_{1j} + a_{i2}p_{2j} + \ldots + a_{ir}p_{rj}, \quad (6.1.5)$$

де p_{ij} – значення факторів; a_{ij} - Постійні коефіцієнти, які треба визначити.

Елементи a_{ij} матриці $\mathbf{A}$ називаються *факторними навантаженнями* .

Формула (6.1.5) у матричному вигляді переписується як

$$\mathbf{Ж} = \mathbf{Б} \cdot \mathbf{С}. \quad (6.1.6)$$

Підставивши (6.1.6) у (6.1.4), отримаємо

$$\mathbf{R} = \mathbf{A}\frac{1}{n-1}\mathbf{P} \cdot \mathbf{P}^T\mathbf{A}^T. \quad (6.1.7)$$

Можна стверджувати, що

$$\frac{1}{n-1}\mathbf{P} \cdot \mathbf{P}^T = \mathbf{C}, \quad (6.1.8)$$

де $\mathbf{C}$ є кореляційною матрицею.

Тоді з (6.1.7) випливає, що

$$\mathbf{R} = \mathbf{A} \cdot \mathbf{C} \cdot \mathbf{A}^T. \quad (6.1.9)$$

Якщо фактори, які визначаються матрицею $\mathbf{P}$ статистично незалежні, тобто $\mathbf{C} = \mathbf{E}$ - одинична матриця, то

$$\mathbf{R} = \mathbf{A} \cdot \mathbf{A}^T. \quad (6.1.10)$$

Якщо кількість факторів p_{ij} відповідає кількості показників у вихідній матриці z_{ij}, то для СКО справедливо

$$S_i^2 = \sum_{j=1}^{n} a_{ij}^2 = 1, \quad (6.1.11)$$

тобто для нормованої вихідної матриці сума квадратів всіх факторних навантажень одного показника дорівнює дисперсії нормованих його величин.

При факторному аналізі найчастіше використовують так званий *метод головних компонентів* . Для цього кореляційна матриця $\mathbf{R}$ перетворюється на діагональну матрицю $\mathbf{Л}$ власних значень λ_i. Для цього будується характеристичний багаточлен

$$\left|\lambda \cdot \mathbf{E} - \mathbf{R}\right| = \begin{vmatrix} \lambda - 1 & -r_{21} & \ldots & -r_{1p} \\ -r_{21} & \lambda - 1 & \ldots & -r_{2p} \\ \vdots & \vdots & & \vdots \\ -r_{p1} & -r_{p2} & \ldots & \lambda - 1 \end{vmatrix}. \quad (6.1.12)$$

На основі системи рівнянь (6.1.12) шукається коріння (власні значення) $\lambda_1, \ldots \lambda_p$, при цьому

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \text{ та } \lambda_1 + \lambda_2 + \ldots \lambda_p = 1. \quad (6.1.13)$$

За власними векторами будується нормована матриця перетворення V з елементами v_i:

$$(\lambda_i \mathbf{E} - \mathbf{R})v_i = 0. \qquad (6.1.14)$$

Матриця факторних навантажень $\mathbf{A}$ обчислюється за формулою

$$\mathbf{A} = \mathbf{V} \cdot \mathbf{Л}^{\frac{1}{2}},$$

де

$$\mathbf{Л}^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_p} \end{pmatrix}.$$

Таким чином, навантаження i-ой головної компоненти на j-у змінну обчислюється за формулою

$$p_i = a_{ji}\sqrt{\lambda_j}.$$

Якщо кількість факторів p_i збігається з кількістю вихідних ознак

Z_j, то $\sum_{j=1}^{k} a_{ji}^2 = \lambda_i$,

тобто, λ_i визначає СКО i- ого ознаки.

Матрицю основних компонентів можна отримати за формулою

$$\mathbf{P} = \mathbf{A}^{-1} \cdot \mathbf{Z} = \mathbf{Л}^{-\frac{1}{2}} \cdot \mathbf{V}^T \cdot \mathbf{Z},$$

де $\mathbf{Z}$ – матриця нормованих значень вихідних ознак.

Таким чином, факторні навантаження визначають кореляційний зв'язок між вихідними ознаками та головними компонентами (факторами).

З умови (6.1.13) слід, що можна вибрати кількість головних компонент i набагато менше кількості вихідних ознак n. Наприклад, три ознаки p_1, p_2, p_3 можуть давати 90% вкладу сумарну дисперсію. І тут виявляється, що з опису даних досить оцінити поведінка цих трьох чинників.

Таблиця 6.1 Таблиця 6.2

	ФФР1	ФФР2	ФФР3
EESR	**0,946**	-0,128	0,184
GAZA	**0,823**	-0,419	-0,046
IRG2	-0,076	0,210	**-0,878**
KMAZ	**0,823**	0,404	-0,242
LKON	**0,950**	-0,081	0,194
MGTS	**0,812**	0,088	0,123

ФФР1	ФФР2	ФФР3	ФФР4	ФФР5
0,428	-0,146	0,082	0,419	**0,763**
0,320	-0,089	0,075	-0,048	**0,928**
0,019	0,107	**-0,977**	-0,072	-0,12
0,730	0,312	-0,268	0,326	0,348
0,478	-0,069	0,150	0,383	**0,741**
0,714	-0,063	0,036	0,414	0,444

MSNG	**0,891**	0,282	0,185	0,676	0,184	0,082	0,460	0,449
NKEL	**0,820**	0,507	-0,013	0,321	0,362	-0,109	0,634	0,499
RTKM	**0,899**	0,156	0,296	0,392	0,046	0,137	0,611	0,626
SNGS	0,544	0,532	0,462	0,307	0,015	0,087	**0,930**	0,048
SPTL	0,094	**0,801**	0,244	0,046	**0,973**	-0,103	0,055	-0,133
TATN	**0,935**	-0,044	0,051	**0,761**	-0,055	0,038	0,242	0,546
S, %	61,7	14,2	11,0	24,6	10,5	29,3	21,0	29,1

Наприклад наведемо результати факторного аналізу котирувань акцій дванадцяти найбільших компаній на російському фондовому ринку. Аналізувалися дані за період 2000 – 2002 рр. (Табл.6.1). Для отримання матриці факторних навантажень, наведених у таблицях 6.1 і 6.2, вихідними даними були випадкові тимчасові зміни котирувань акцій. У лівому стовпці позначено 12 вихідних ознак (SNGS – «Сургутнафтогаз», EESR – «РАО ЄС Росії» тощо) S – процентний внесок кожного фактора у загальну дисперсію. При виділенні основних компонентів виявилося, що основними є три фактори фондового ринку (ФФР1, ФФР2 і ФФР3). Їх сумарний вклад становив 61,7+14,2+11,0=86,9 %.

Перший найбільш вагомий фактор ФФР1 має високий кореляційний зв'язок з більшістю обраних емітентів. Цей чинник у результаті аналізу даних необхідно інтерпретувати. Він може визначатися ціною на нафту, банківськими ставками, обмінним курсом національної валюти тощо. (або комбінацією цих макроекономічних показників). Важливим є те, що за даними табл.6.1 з певним ступенем достовірності можна стверджувати, що ціни на виділені жирним шрифтом акції пов'язані між собою та фактором ФФР1.

У табл.6.2 показані результати факторного аналізу тих самих вихідних даних, але розкладених по п'яти факторів. Відповідно до табл.6.2 сумарний внесок у загальну дисперсію зріс із 86,9 до 94 %. При порівнянні даних виявляється, що відбувається перерозподіл факторних навантажень. Найбільш вагомий фактор ФФР1 табл.6.1 поділяється на два фактори. Акції SNGS, навантаження яких було рівномірно розподілено на три фактори, випадають в окремий найбільш вагомий фактор.

Прийнято виділяти кореляційні зв'язки за коефіцієнта кореляції $|R| > 0.7$. С обліком формули (3.2.5) $R \approx \sqrt{1-(2\gamma)^2}$, де γ - Відносна похибка. При $R = 0.7$ похибці $\gamma < 0.5$, тобто це межа, при якому випадкова похибка можна порівняти з вимірюваною величиною.

Зв'язок факторного аналізу з *регресійним аналізом* на основі МНК (див. Розділ 3 цього посібника) полягає в наступному. Факторні навантаження (коефіцієнти розкладання a_{ij} (6.1.5)) лише приблизно відповідають регресійним залежностям, побудованим методом найменших квадратів. Для однофакторного зв'язку двох ознак і X коефіцієнт відповідає нахилу прямої a, що проходить (середньої) між лінійною регресією і X на X.

4.2. Основи планування багатофакторного експерименту

Як зазначалося у першій лекції, у випадку об'єкт дослідження можна як структурної схеми, показаної на рис.6.1.

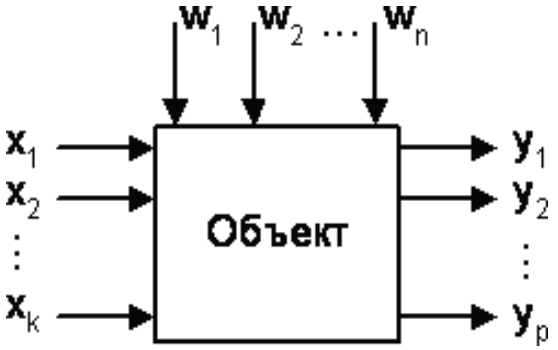


Рис.6.1

Подання об'єкта у вигляді такої схеми ґрунтується на принципі «чорної скриньки». Маємо такі групи параметрів:

1) управляючі (вхідні) x_i, які називаються *факторами* ;

2) вихідні параметри y_i, які називаються *параметрами стану* ;

3) w_i - *обурюючі впливи* .

Передбачається, що впливи, що обурюють, не піддаються контролю і або є випадковими, або змінюються в часі.

Кожен фактор x_i має область визначення, яка має бути встановлена до проведення експерименту.

Комбінацію чинників можна як крапку в багатовимірному просторі, характеризує стан системи.

На практиці метою багатофакторного експерименту є встановлення залежності

$$y = f(x_1, x_2, \dots x_k), \quad (6.2.1)$$

описує поведінка об'єкта. Найчастіше функція (6.2.1) будується як полінома

$$y = a_0 + a_1 x_1 + a_2 x_2 \quad (6.2.2)$$

або

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_{11} x_1^2 + a_{22} x_2^2 + a_{12} x_1 x_2. \quad (6.2.3)$$

Метою експерименту може бути, наприклад, побудова залежності (6.2.1) за мінімальної кількості вимірювань значень керуючих параметрів x_i.

На першому етапі планування експерименту необхідно вибрати область визначення факторів x_i. Вибір цієї області проводиться з апріорної інформації. Значення x_i називаються *рівнями керуючого параметра* .

Якщо вибрано лінійну модель (6.2.2), то для побудови апроксимуючої функції достатньо вибрати *основний рівень* та *інтервал варіювання* керуючого параметра x_i.

Для лінійної моделі інтервал варіювання можна визначити як

$$I = \frac{x_{max} - x_{min}}{2},$$

а основний (нульовий) рівень – як середнє значення

$$x_0 = \frac{x_{max} + x_{min}}{2}.$$

Для спрощення планування експерименту прийнято замість реальних рівнів x_i використовувати *кодовані* значення факторів. Для факторів з безперервною областю визначення це можна зробити за допомогою наступного перетворення

$$x_j = \frac{\tilde{x}_j - x_{j0}}{I_j},$$

де $\tilde{x}_j$ – натуральне значення фактора; I_j - інтервал варіювання; x_{j0} - Основний рівень; x_j - Кодоване значення. У результаті x_j набуває значень на межах , на основному рівні $x_j = 0$. Основна проблема полягає у виборі області варіювання, оскільки це завдання є неформалізованим.

Розглянемо *повний* факторний експеримент з прикладу лінійної моделі (6.2.2). Якщо число чинників k, то проведення повного факторного експерименту потрібно $N = 2^k$ дослідів, де 2 - число рівнів, якого досить побудови лінійної моделі.

Умову проведення цього експерименту можна зафіксувати у матриці планування (табл.5.3).

Таблиця 6.3

Номер досвіду	x_1	x_2	
1	-1	-1	y_1
2	+1	-1	y_2
3	-1	+1	y_3
4	+1	+1	y_4

Таким чином, для двох факторів побудова матриці планування є елементарною. Для більшої кількості чинників потрібно розробити правила побудови таких матриць. Наприклад, з появою чинника x_3 в табл.6.3 відбудуться такі зміни (табл.6.4): з появою нового стовпця кожна комбінація рівнів вихідної таблиці проявиться двічі.

Таблиця 6.4

Номер досвіду	x_1	x_2	x_3	
1	-1	-1	+1	y_1
2	+1	-1	+1	y_2

3	-1	+1	+1	y_3
4	+1	+1	+1	y_4
5	-1	-1	-1	y_5
6	+1	-1	-1	y_6
7	-1	+1	-1	y_7
8	+1	+1	-1	y_8

Не єдиний спосіб розширення матриці планування. Використовують також перемноження стовпців, правило чергування символів.

Дуже важливі загальні властивості матриці планування:

1) симетричність матриці щодо центру експерименту: $x_i = 0$. Тоді $\sum_{i=1}^{N} x_{ji} = 0$.

2) умова нормування $\sum_{i=1}^{N} x_{ij}^2 = N$, тобто сума квадратів елементів кожного стовпця дорівнює числу дослідів.

Перші дві властивості відносяться до побудови окремих стовпців матриці

3) сукупність стовпців має таку властивість $\sum_{i=1}^{N} x_{ij} x_{in} = 0$, де $j \neq n$.

4) *Ротатабельність* . Це означає, що точки (значення факторів) у матриці планування підбираються так, що точність передбачення вихідного параметра має бути однаковою на рівних відстанях від центру експерименту (нульового рівня) і не залежати від напрямку.

Планування експерименту першого порядку для двох змінних. План експерименту першого порядку для двох змінних показано на рис.6.2. Тобто потрібна функція $y = f(x_1, x_2)$описується модельно у вигляді площини

$$y = a_0 + a_1 x_1 + a_2 x_2 \qquad (6.2.4)$$

або гіперболоїда

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2. \qquad (6.2.5)$$

Розташування цієї моделі у просторі показано на рис.6.2 поверхнею, що проходить через точки 1 – 2 – 3 – 4.

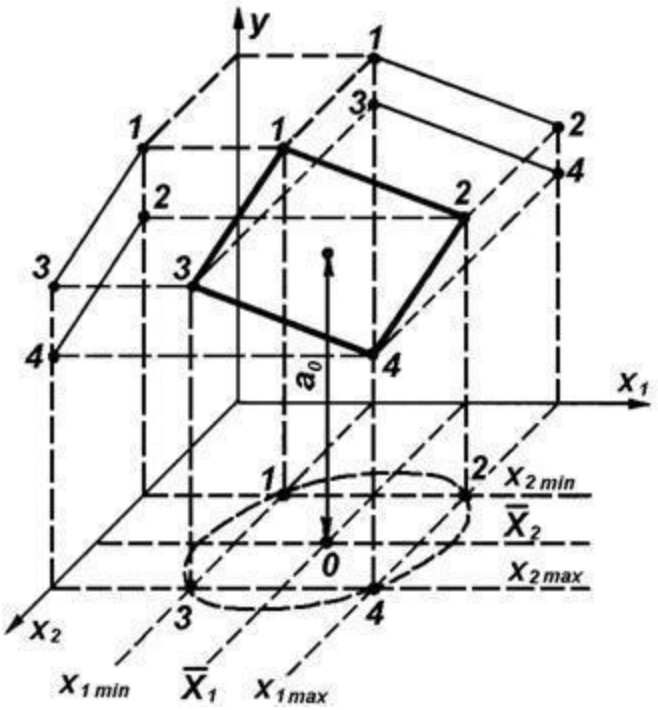


Рис.6.2

Необхідні рівні для повного факторного експерименту розташовані у площині (x_1, x_2). Для моделі як гіперболоїда цей план є гранично економним. Для побудови гіперболоїду необхідно визначити чотири коефіцієнти моделі (6.2.5). Це можна зробити, вирішуючи систему із чотирьох рівнянь. Отже, потрібні всі чотири досвіди. Теоретично планування експерименту використовується термін *насиченості* .

Якщо розглядати модель (6.2.4) у вигляді площини, то план експерименту є ненасиченим (*надлишковим*), тому що необхідно визначити тільки три коефіцієнти a_0, a_1і a_2. У випадку моделі (6.2.5) (насичений експеримент) рішення системи єдине, і поверхня гіперболоїда пройде через всі чотири експериментальні значення y_i. Наслідком цього є те, що насичений експеримент не дозволяє усереднити випадкові похибки і не дає відомості про їхній розмір.

Для ненасиченого плану (6.2.4) надмірна кількість дослідів дозволяє зробити усереднення та оцінити розміри похибки. Провівши площину через точки 1, 2 і 3, можна оцінити похибку, визначивши, на якій відстані від площини знаходиться точка 4. нахилу прямий 3 – 4. a_1Аналогічно коефіцієнт x_1можна x_2 визначити a_2з нахилу прямих 1 – 3 і 2 – 4.

Оскільки отримані таким чином значення a_1можуть a_2відрізнятися, ненасичений експеримент дозволяє провести їх усереднення і оцінити похибку.

Якщо рівняння площини подати у вигляді

$$y = a_0 + a_1(x_1 - \bar{x}_1) + a_2(x_2 - \bar{x}_2), \quad (6.2.6)$$

де $\bar{x}_1 = \frac{x_{1\min} + x_{1\max}}{2}$; $\bar{x}_2 = \frac{x_{2\min} + x_{2\max}}{2}$, то ми переносимо початок координат на точку з координатами $(\bar{x}_1, \bar{x}_2)$. Тоді коефіцієнт a_0знаходиться усереднення всіх чотирьох значень y_iяк висота центру площини 1 - 2 - 3 - 4.

Процес перенесення початок координат в центр простору факторів з координатами $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)$дуже важливий при обробці даних будь-яких експериментів, що описуються моделлю у вигляді гіперплощини, так як дозволяє отримати більш стійке усереднене значення для a_0.

Найважливішим фактором є те, що в результаті такого усереднення побудована площина задовольняє всі чотири значення y_i лише в середньому. У будь-якій точці може бути знайдена похибка відхилення експериментальних даних щодо моделі, і за цими чотирма відхиленнями можна обчислити СКО.

Таким чином, один із чотирьох дослідів є надлишковим і може бути виключений. Але тоді план експерименту стає неротатабельним, тобто нерівноточним у всіх напрямках. Якщо виключена точка 4 на рис.6.2, то в напрямку 3 - 2 в площині факторів буде забезпечена більша точність, ніж у напрямку 1 - 0. У цьому випадку для відновлення ротатабельності точки 1, 2 і 3 в площині факторів повинні бути рівновіддалені як один від одного, так і від центру в стороні 0 моделі (6.2.4), експеримент, що містить кінцеве число дослідів дозволяє отримати тільки оцінки для коефіцієнтів a_0, a_1 і a_2. Підставивши в рівняння моделі (6.2.4) відомі значення факторів x_{ij} та результати дослідів y_i отримаємо систему лінійних рівнянь алгебри для визначення a_i. Якщо a_2 кількість цих рівнянь більше трьох, то значення оцінок a_0 і a_1 можуть бути отримані за допомогою МНК:

$$a_j = \left(\sum_{i=1}^{N} x_{ij} y_i \right) \Big/ N \quad , (6.2.7)$$

де N – кількість дослідів. Тут враховано, що x_{ij} набувають значення -1, +1.

Для обчислення коефіцієнтів лінійної моделі за формулою (6.2.7) отримаємо:

$$a_1 = \frac{[(-1)y_1 + (+1)y_2 + (-1)y_3 + (+1)y_4]}{4},$$

$$a_2 = \frac{[(-1)y_1 + (-1)y_2 + (+1)y_3 + (+1)y_4]}{4}$$

Таким чином, для обчислення a_1 можна a_2 використовувати (6.2.8). Для визначення a_0 у формулі (6.2.4) знайдемо середнє значення $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, що дорівнює $\bar{y} = a_0 + a_1\bar{x}_1 + a_2\bar{x}_2$, де $\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i}$, $\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2i}$.

У разі симетричності матриці планування $\bar{x}_1 = \bar{x}_2 = 0$, звідки $\bar{y} = a_0$. Щоб коефіцієнт моделі обчислювався за єдиною формулою (6.2.7), у матриці планування вводять фіктивну змінну x_0, яка набуває значення 1 у всіх дослідах і відповідає коефіцієнту a_0. Коефіцієнт при незалежних змінних x_i свідчить про силу впливу чинників: що більше значення має коефіцієнт a_i, то більше впливає відповідний чинник. У цьому сенсі результат планування експерименту алогічний факторного аналізу. Для пасивних експериментів факторний аналіз може використовуватися як апріорні дані при плануванні.

Плануючи експеримент, прагнуть отримати лінійну модель, однак у вибраних інтервалах варіювання апріорі не відомо, що лінійна модель адекватно описує поведінку системи.

Нелінійність пов'язана зі змішаною взаємодією. Формула (6.2.5) завжди може бути оцінена за повним факторним експериментом. Для повного факторного експерименту $N = 2^2$ матриця планування з урахуванням ефекту взаємодії наведено у табл.6.5.

Таблиця 6.5

Номер досвіду	x_0	x_1	x_2	x_1x_2	
1	+1	-1	-1	+1	y_1
2	+1	+1	-1	-1	y_2
3	+1	-1	+1	-1	y_3
4	+1	+1	+1	+1	y_4

У цьому випадку коефіцієнт a_{12}також може бути обчислений за формулою (6.2.7):

. (6.2.7)

Стовпці x_1, x_2що задають планування експерименту – за ними визначають результати досвіду; стовпці x_0, x_1x_2служать лише розрахунку.

Зі зростанням числа факторів кількість можливих взаємодій зростає. Наприклад, для факторного експерименту $N = 2^3$крім x_0, x_1, x_2, x_3у матриці планування з'являються стовпці x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$. Усього в матриці планування виявляється вісім стовпців, отже необхідно визначати вісім коефіцієнтів. Усі вісім коефіцієнтів необхідно визначати у тому випадку, якщо враховувати змішану взаємодію. Якщо модель задається як гіперплощини (лінійна модель), досить визначити чотири коефіцієнта: x_0, x_1, x_2, x_3. Повний факторний експеримент виявляється надмірним і у експериментатора виникає вибір:

1. Побудувати гіперплощину за чотирма експериментами, інші чотири досвіду використовуватиме оцінки похибки.

2. Провести експеримент, що складається із 4-х дослідів, тобто реалізувати економний план експерименту.

Таким чином, на відміну від моделі гіперболоїду, яка потребує визначення 2^kневідомих коефіцієнтів, модель гіперплощини, містить $k+1$коефіцієнт і вимагає відповідної кількості дослідів, тобто повний факторний план (ПФП) для моделі гіперплощини надмірний.

Для побудови гіперплощини, отже, досить використовувати лише деяку частину ПФП. Цю частину теорії планування експерименту називають *дробової* реплікою чи дробовим факторним планом (ДФП). Якщо дроблення ПФП виробляється послідовним розподілом числа дослідів на 2, то репліку називають *регулярною* . Число pпослідовного поділу називають *дробовістю* репліки.

Число дослідів регулярного ДФП дорівнює $n = 2^{k-p}$. При $p = 1$ДФП називають напівреплікою (або 1/2 репліки), при $p = 2$- 1/4 репліки і т.д.

Відповідну кількість дослідів та параметрів планування наведено у таблиці 6.6.

Таблиця 6.6

Число факторів, k	Число коеф. моделі, $k+1$	Число досліді в ПФП	Вигляд плану	Число досліді в плану	Надмірність
2	3	4	ПФП	4	1
3	4	8	Напівреплiка	4	0
4	5	16	Напівреплiка	8	3
5	6	32	Четвертьреплiка	8	2
6	7	64	1/8 репліка	8	1
7	8	128	1 / 16 репліка	8	0
8	9	256	1 / 16 репліка	16	7

Для складання планів-таблиць регулярних дробових реплік часто використовують зване правило двійкового коду. Воно говорить, що для моделі у вигляді гіперболоїда знаки "+" і "-" в стовпцях плану повинні чергуватись за правилом чергування двійкових чисел у розряді двійкового коду, тобто в стовпці X_1- через 1, в стовпці X_2- через 2, в стовпці X_3- через 4, в стовпці X_k- через 2^{k-1}.

Проведення експериментів та обробка результатів. Так як експеримент містить елемент невизначеності внаслідок обмеженості оброблюваних даних, то розстановка повторних або паралельних дослідів не дає результатів, що повністю збігаються. Отримана похибка (*відтворюваності*) оцінюється стандартними методами усереднення, тобто

$$\bar{y} = \frac{1}{n}\sum_{q=1}^{n} y_q ,$$

де n- Число паралельних дослідів.

У цьому випадку дисперсія дорівнює

$$\sigma^2 = \frac{1}{n-1}\sum_{q=1}^{n} (y_q - \bar{y})^2 . \qquad (6.2.8)$$

Якщо врахувати, що матриця планування складається із серії дослідів, то оцінка дисперсії всього експерименту виходить у результаті усереднення дисперсії всіх дослідів. І тут говорять про дисперсії відтворюваності жодного досвіду, а експерименту загалом. Така дисперсія дорівнює

$$\sigma^2(y) = \frac{\left[\sum_{i=1}^{N}\sum_{q=1}^{n} (y_{iq} - \bar{y})^2\right]}{N(n-1)} , \qquad (6.2.9)$$

де N – число різних дослідів (кількість елементів у матриці планування); n- Число повторних дослідів.

Формула (6.2.9) справедлива, якщо дотримується рівність числа повторних дослідів у всіх точках експериментальних матриці планування. Насправді у різних точках буває виконано різне число дослідів. В цьому випадку для оцінки дисперсії відтворюваності користуються *середньозваженим* значенням

$$\sigma^2(y) = \frac{\sum_{i=1}^{N} f_i \sigma_i^2}{\sum_{i=1}^{N} f_i}, \quad (6.2.10)$$

де σ_i^2 - Оцінка дисперсії i-го досвіду, f_i- Число ступенів свободи в i-ому досвіді; це число розраховують як $f_i = n_i - 1$(число паралельних дослідів мінус 1).

4.3. Планування експерименту за оптимальних умов

Знаходження оптимальних умов досліджуваного об'єкта – найважливіше практичне завдання. Найчастіше при багатофакторному експерименті потрібно знайти значення факторів x_iтакі, при яких відгук системи набуває значення y_{max}або y_{min}. Таким чином будується цільова функція відгуку

$$y = y(x_1, x_2, \ldots, x_k), \quad (6.3.1)$$

і завдання оптимізації зводиться до знаходження $x_{1опт}, x_{2опт}, \ldots, x_{kопт}$, що забезпечують екстремум функції мети

$$y(x_{1опт}, x_{2опт}, \ldots, x_{kопт}) = y_{min}(y_{max}). \quad (6.3.2)$$

Крім того, на значення факторів накладаються додаткові обмеження

$$y_i(x_1, x_2, \ldots, x_k) \{\geq, =, \leq\} R_i \text{де } i = 1 \ldots r.$$

Таким чином, завданням оптимізації є знаходження екстремуму функції відгуку за умови, що сама функція апріорі невідома. Таке завдання може бути вирішене багатьма способами:

1. Шляхом повного факторного експерименту будується нелінійна модель функції відгуку і потім ця функція має екстремум. Така модель може виявитися складною і вимагати великої кількості дослідів, тому що вимоги знаходження її екстремуму можуть змусити проводити повний факторний експеримент у широкому діапазоні варіювання та при великій кількості дослідів.

2. Більше практично прийнятним виявляється “покроковий” підхід до вирішення завдання перебування екстремуму. У цьому випадку експеримент проводиться в обмеженій ділянці. Знаходиться напрям зростання функції відгуку (при знаходженні максимуму) або напрямок падіння функції відгуку (при знаходженні мінімуму). Потім експеримент проводиться у наступній області тощо. Таким чином, здійснюється послідовний пошук екстремуму функції відгуку. У цьому випадку задача оптимізації може бути вирішена без повного опису функції відгуку у всій галузі варіювання факторів.

При послідовному знаходженні екстремуму yдосягнення максимуму за найменшу кількість кроків відбувається при послідовному русі у напрямку найбільшого зростання (зменшення) функції відгуку, тобто у напрямку градієнта

$$grad\ y(\vec{x}) = \frac{\partial y}{\partial x_1}\vec{x}_1 + \frac{\partial y}{\partial x_2}\vec{x}_2 + \ldots + \frac{\partial y}{\partial x_k}\vec{x}_k, \quad (6.3.3)$$

де $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$– орти відповідних координатних осей.

При досягненні екстремуму вираз (6.3.3) змінює знак протилежний. При здійсненні покрокового руху методом градієнта змінюються одночасно значення всіх факторів. Приріст змінних $x_1, x_2, \ldots, x_k$вибирається пропорційно відповідної складової градієнта, тобто алгоритм має вигляд:

$$x_j^{(L+1)} = x_j^{(L)} + h^{(L+1)} \frac{\partial y(x^{(L)}) / \partial x_j}{\sum_{j=1}^{k} \left(\frac{\partial y(x^{(L)})}{\partial x_j} \right)^2}, \quad (6.3.4)$$

де $x_j^{(L+1)}$ – значення змінної після L-го та $(L+1)$-го кроків; $h^{(L+1)}$ - Розмір кроку.

При знаходженні екстремуму градієнтним методом необхідно на кожному кроці вимірювань будувати гіперплощину для знаходження градієнта, тобто проводити не менш ніж $(k+1)$ досвід. Найбільшого практичного застосування отримав різновид градієнтного спуску – метод *крутого сходження* (швидкого спуску), названий на ім'я автора Бокса-Уилсона.

У цьому методі градієнт функції відгуку визначається лише в початковій точці і надалі рух здійснюється в цьому вибраному напрямку, але обчислення градієнта на кожному кроці не провадиться. Покроковий рух здійснюється до потрапляння функції до приватного оптимуму (екстремум функції у вибраному напрямку – крива 2 на рис.6.3). У точці приватного екстремуму знаходиться знову градієнт і визначається напрямок подальшого руху тощо.

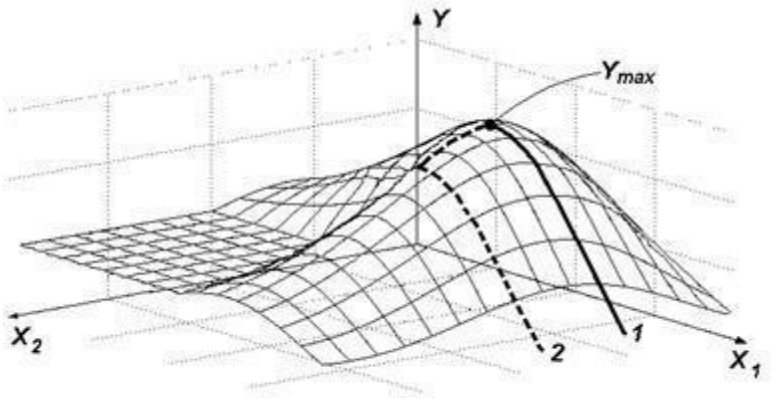


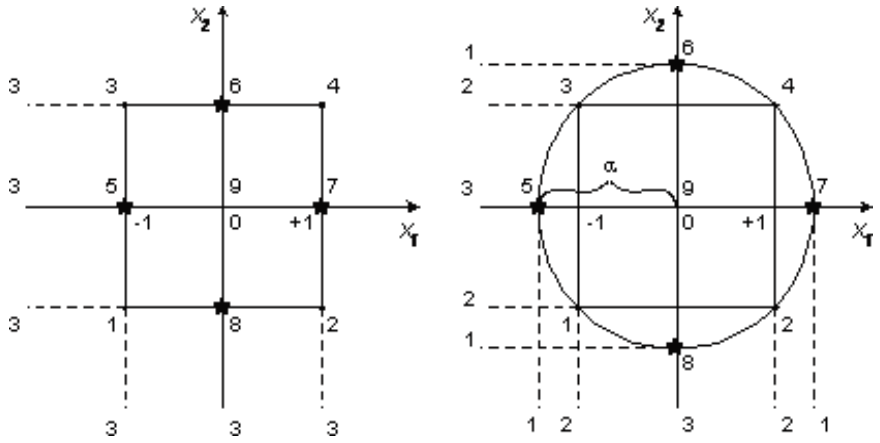
Рис.6.3

В результаті покрокового руху обома методами експериментатор потрапляє в квазістаціонарну область, близьку до точки оптимуму. Ця область апріорі не може бути описана гіперплощиною і вимагає опис у вигляді нелінійної моделі (гіперболоїда, параболоїда тощо).

При $k = 2$ загальний вигляд функції відгуку другого порядку буде таким:

. (6.3.5)

Для визначення такої поверхні фактори X_1 повинні X_2 варіюватися не на двох, а мінімум на трьох рівнях. Експеримент у цьому випадку потребує не менше шести дослідів для двох факторів (за кількістю коефіцієнтів у рівнянні (6.3.5)).



а б

Рис.6.4

На малюнку 6.4 показано вибір додаткових точок для такого експерименту. До дослідів 1, 2, 3, 4 до побудови площини додаються досліди 5, 6, 7, 8 і, обов'язково, точка 9 в центрі квадрата.

Якщо додаткові точки розташовуються на сторонах квадрата (рис.6.4, а), напрям з вершин через центр визначає кривизну поверхні. План представлений малюнку 6.4,а надлишковий (дев'ять дослідів визначення шести коефіцієнтів), але неротатабелен. Для забезпечення ротатабельності точки 5, 6, 7, 8 необхідно видалити від центру тяжіння на відстань α, названу *зоряним плечем* . Зіркове плече має в $\sqrt{2}$ рази більшу довжину, ніж сторона квадрата.

Загальне правило побудови ротатабельних планів 2-го порядку для довільної кількості факторів:

За ядро плану приймають ПФП для k факторів, до них додають зоряних точок на відстані α від центру плану. Зіркове плече α визначається за формулою

$$\alpha = 2^{\frac{k-p}{2}},$$

де k - Число факторів; p - Дрібність репліки.

З іншого боку, планується n_0 дослідів у центрі плану (точка 9 на рис.6.4).

4.4. Планування експерименту щодо визначення динамічних характеристик об'єкта

Вище ми розглядали функцію відгуку об'єкта і планування експерименту у разі, якщо $y = f(x_1, x_2, \ldots, x_n)$ характеристики об'єкта не залежать від часу.

Якщо об'єкт динамічний, тобто. вхідні чинники і параметри об'єкта залежить від часу, то завдання ідентифікації об'єкта (побудова функції відгуку) багато чому аналогічна електротехнічним завданням.

Для лінійного об'єкта $y = ax + b$; якщо об'єкт вважається динамічним $y = y(t), x = x(t), a = a(t), b = b(t)$.

Аналогія з електротехнічними завданнями у тому, що з наявності реактивних елементів (індуктивність, ємність електричних ланцюгів) вихідні сигнали $y = y(t)$ реагують на вхідні $x = x(t)$ з певної тимчасової затримкою.

Як було показано вище, теоретично реєструючих приладів, характеристики лінійних електричних чотириполюсників можуть у тимчасовій області описуватися функцією відгуку або, в частотній області, - смугою пропускання. Ці властивості об'єкта пов'язані Фур'є-перетворенням. Функція відгуку є реакцією динамічної системи на імпульсні (у вигляді δ функції або θ функції) впливу.

У випадку динамічні об'єкти описуються нелінійними інтегральними чи диференціальними рівняннями. Як правило, на околицях робочих режимів проводиться лінеаризація об'єктів.

Методи ідентифікації динамічних об'єктів щодо добре розроблені для лінійних об'єктів.

Для простоти розглянемо одномірний об'єкт, що має один вхід $x(t)$ та один вихід $y(t)$. $x(t)$ та $y(t)$ пов'язані деяким диференціальним рівнянням. Визначити це рівняння за експериментальними даними виявляється можливим лише у деяких найпростіших випадках.

Імпульсна характеристика і перехідна функція об'єкта можуть бути експериментально отримані як реакція об'єкта на вплив у вигляді стрибка (δ-функції).

Частотні характеристики можна визначити, подаючи на вхід гармонійний вплив постійної амплітуди та змінної частоти (визначити АЧХ).

Більшість складних об'єктів характерно наявність випадкових обурень і завдання ідентифікації вимагає статистичних методів визначення динамічних характеристик. Найпростіший приклад – на вхід електронної схеми з власними шумами подається гармонійний сигнал, амплітуда та фаза якого на виході виявляються випадково промодульованими. За характеристиками цієї модуляції можна визначити динамічні характеристики об'єкта (випадкові). Такий метод ідентифікації вимагає значного часу вимірювань і не завжди застосовується на практиці.

Найбільшого поширення для ідентифікації динамічних об'єктів набули кореляційні методи, що використовуються отримання імпульсних характеристик об'єктів.

Імпульсна характеристика $g(t)$ стаціонарного лінійного об'єкта за умови некорелювання внутрішніх обурень із вхідним сигналом визначається інтегральним рівнянням Вінера-Хопфа:

$$K_{xy}(\tau) = \int_0^{\infty} K_{xx}(t-\tau) g(t) dt \quad , (6.4.1)$$

де - $K_{xx}(\tau)$ автокореляційна функція вхідного сигналу $x(t)$

$K_{xy}(\tau)$ - взаємна кореляційна функція вхідного та вихідного сигналу.

Щоб за допомогою рівняння (6.4.1) ідентифікувати об'єкт необхідно вирішити це рівняння щодо $g(t)$ експериментальних залежностей $K_{xx}(\tau)$ і $K_{xy}(\tau)$.

Найпростіше рішення такого завдання виходить у тому випадку, якщо на вхід об'єкта подається випадковий сигнал у вигляді «білого» шуму або короткий імпульс у вигляді δ-функції ($K_{xx}(\tau) = \delta(\tau)$), тоді рівняння Вінера-Хопфа набуде вигляду:

$$K_{xy}(\tau) = \int_0^{\infty} \delta(t-\tau) g(t) dt = g(\tau) \quad . \qquad (6.4.2)$$

На практиці створити некорельований білий шум або нескінченно короткий імпульс із нескінченно широким спектром неможливо. Тому доводиться використати реальні сигнали з максимально широким спектром. Це створює певну (іноді значну) похибку ідентифікації об'єктів.

Рішення інтегрального рівняння (6.4.1) у всіх інших випадках є *некоректно поставленим завданням* . Некоректна постановка завдання у разі перш за все означає неоднозначність рішення для функції $g(t)$.

Завдання полягає у знаходженні підінтегральної функції за відомим інтегралом. Якщо вважати, що інтеграл - площа під кривою, що описується підінтегральним виразом, має місце нескінченне число кривих, що дають однакове значення інтеграла. Математичне визначення некоректно поставленої задачі в даному

випадку означає, що нескінченно малі похибки у визначенні $K_{xx}(\tau)$ та $K_{xy}(\tau)$ призводять до кінцевих похибок в оцінці $g(t)$.

Для усунення некоректності завдання розвинені звані *методи регуляризації* . Як правило, вони засновані на використанні апріорної інформації (*a priori (Лат)* - з попереднього). У найпростішому випадку функція $g(t)$ параметризується. Зокрема - може бути задано у вигляді ряду з коефіцієнтами, які необхідно визначити.

Розглянемо такий параметричний метод, що грунтується на розкладанні функції $g(t)$ в ряд за системою відомих ортогональних функцій:

$$g(t) = \sum_{i=0}^{N} C_i \varphi_i(t) \quad , (6.4.3)$$

де

$\varphi_i(t)$ - Базисна система відомих функцій.

-Коефіцієнти розкладання.

При використанні (6.4.3) метою експерименту визначення невідомих коефіцієнтів C_i. У цьому випадку вихідний сигнал запишеться у вигляді:

$$y_M(t) = \int_0^{\infty} x(t-\tau) \cdot \sum_{i=0}^{N} C_i \varphi_i(\tau) d\tau \qquad (6.4.4)$$

Для отримання оцінок коефіцієнтів скористаємося критерієм мінімуму середньоквадратичної похибки ε^2:

$$\varepsilon^2 = M\left\{\left[y(t) - y_M(t)\right]^2\right\} = M\left\{\left[y(t) - \int_0^{\infty} x(t-\tau) \sum_{i=0}^{N} C_i \varphi_i(\tau) d\tau\right]^2\right\}$$

(6.4.5)

Введемо векторні позначення $C = \{C_0, C_1, \ldots, C_N\}$

$$Z_i(t) = \int_0^{\infty} \varphi_i(\tau) x(t-\tau) d\tau \quad . \qquad (6.4.6)$$

Функцію $Z_i(t)$ називають *вихідними реакціями* фільтра з імпульсними характеристиками $\varphi_i(t)$ вхідний сигнал *x (t)*.

З урахуванням (6.4.6) запишемо (6.4.5) у вигляді:

$$\varepsilon^2 = M\left\{\left[y(t) - \sum_{i=0}^{N} C_i Z_i(t)\right]^2\right\}. \qquad (6.4.7)$$

Завдання полягає в мінімізації ε^2. У цьому випадку вектор *можна* знайти з (N +1) рівнянь:

$$\frac{\partial \varepsilon^2}{\partial C_j} = 0 \qquad j = 1,2,\ldots,N.$$

Продиференціювавши (6.4.6) C_j і прирівнявши похідну до нуля, отримаємо систему рівнянь для знаходження оцінок коефіцієнтів $\hat{C}_i$.

$$M\left\{\left[y(t) - \sum_{i=0}^{N} \hat{C}_i Z_i(t)\right] Z_j(t)\right\} = 0 \quad . \qquad (6.4.8)$$

Перепишемо (6.4.8) у такому вигляді:

$$M\{y(t)Z_i(t)\} = \sum_{i=0}^{N} \hat{C}_i M\{Z_i(t)Z_j(t)\} \quad . \qquad (6.4.9)$$

Моменти $M\{y(t)Z_i(t)\}$і $M\{Z_i(t)Z_j(t)\}$є початкові значення взаємних кореляційних функцій відповідних сигналів $y(t), Z_i(t), Z_j(t)$.

Введемо позначення: $K_{yj} = M\{y(t)Z_j(t)\}; \quad K_{ij} = M\{Z_i(t)Z_j(t)\}$. (6.4.10)

Тоді (6.4.9) можна записати у вигляді:

$$K_{yj} = \sum_{i=0}^{N} \hat{C}_i K_{ij} \qquad j = 0,1, \ldots, N \quad . (6.4.11)$$

Щоб система рівнянь (11) мала рішення, необхідно, щоб матриця K_{yj} була діагональною.

$$K_{yj} = \begin{cases} K, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \qquad (6.4.12)$$

Крім того, базисна система функцій $\varphi_i(t)$має бути ортогональна. Цю умову можна записати в наступному вигляді:

$$\int_0^{\infty} \varphi_i(t)\varphi_j(t)dt = \begin{cases} A, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \qquad (6.4.13)$$

При ортогональній базисній системі функцій ортогональність вихідних сигналів фільтра $Z_i(t)$ забезпечується в тому випадку, коли спектральна щільність вхідного шуму *x (t)* була постійна, тобто рішення рівняння Вінера-Хопфа (1) для практичних завдань завжди можливе лише приблизно.

З вищесказаного можна дійти невтішного висновку, що вибір форм вхідних сигналів у завданнях ідентифікації динамічних об'єктів завжди є ключову проблему. Це необхідно враховувати під час планування активного експерименту з динамічними об'єктами.

Тема 3. Використання математичних моделей для синтезу систем регулювання

Лекція №5. Методи статистичного аналізу.

1. Регресійний та кореляційний аналіз.
2. Канонічний аналіз.
3. Факторний аналіз.
4. Дисперсійний аналіз.
5. Коваріаційний аналіз.

Вирішення практичних завдань математичними методами послідовно здійснюється шляхом математичного формулювання задачі (розробки математичної моделі), вибору методу проведення дослідження отриманої математичної моделі, аналізу отриманого математичного результату.

Математичне формулювання завдання зазвичай представляється як чисел, геометричних образів, функцій, систем рівнянь тощо. п. Опис об'єкта (яви) може бути представлено з допомогою безперервної чи дискретної, детермінованої чи стохастичної та інші математичними формами.

Математична модель являє собою систему математичних співвідношень-формул, функцій, рівнянь, систем рівнянь, що описують ті чи інші сторони об'єкта, що вивчається, явища, процесу

Першим етапом математичного моделювання є постановка задачі, визначення об'єкта та цілей дослідження, завдання критеріїв (ознак) вивчення об'єктів та управління ними. Неправильна чи неповна постановка завдання може звести нанівець результати всіх наступних етапів.

Дуже важливим на цьому етапі є встановлення меж області впливу об'єкта, що вивчається Межі області впливу об'єкта визначаються областю значимої взаємодії із зовнішніми об'єктами. Ця область може бути визначена на основі наступних ознак: межі області охоплюють ті елементи, вплив яких на досліджуваний об'єкт не дорівнює нулю; за цими межами дія досліджуваного об'єкта зовнішні об'єкти прагне нулю. Облік області впливу об'єкта при математичному моделюванні дозволяє включити в цю модель всі істотні фактори і розглядати систему, що моделюється, як замкнуту, тобто, з відомим ступенем наближення, незалежним від зовнішнього середовища Останнє значно спрощує математичне дослідження

Наступним етапом моделювання є вибір типу математичної моделі Вибір типу математичної моделі є найважливішим моментом, що визначає напрямок всього дослідження. Зазвичай послідовно будується кілька моделей Порівняння результатів дослідження з реальністю дозволяє встановити найкращу їх

На етапі вибору типу математичної моделі за допомогою аналізу даних пошукового експерименту встановлюються: лінійність чи нелінійність, динамічність чи статичність, стаціонарність чи нестаціонарність, а також ступінь детермінованості досліджуваного об'єкта чи процесу

Лінійність встановлюється за характером статичної характеристики досліджуваного об'єкта. Під *статичною* характеристикою об'єкта розуміється зв'язок між величиною зовнішнього на об'єкт (величиною вхідного сигналу) і максимальною величиною його на зовнішній вплив (максимальної амплітудою вихідний характеристики системи). Під *вихідний* характеристикою системи розуміється зміна вихідного сигналу системи у часі. Якщо статична характеристика досліджуваного об'єкта виявляється лінійною, то моделювання цього об'єкта здійснюється з використанням лінійних функцій.

Нелінійність статичної характеристики та наявність запізнення у реагуванні об'єкта на зовнішній вплив є яскравими ознаками нелінійності об'єкта. У цьому випадку для його моделювання має бути прийнято нелінійну математичну модель.

Застосування лінійної математичної моделі значно спрощує її подальший аналіз, оскільки така модель дозволяє користуватися принципом суперпозиції. *Принцип суперпозиції* стверджує, що коли лінійну систему впливають кілька вхідних сигналів, кожен з них фільтрується системою так, ніби ніякі інші сигнали неї не діють. Загальний вихідний сигнал лінійної системи за принципом суперпозиції утворюється в результаті підсумовування її реакції на кожен вхідний сигнал

Встановлення динамічності або статичності здійснюється за поведінкою досліджуваних показників об'єкта в часі. Стосовно детермінованої системи можна говорити про статичність чи динамічність за характером її вихідної характеристики. Якщо середнє арифметичне значення вихідного сигналу за різними відрізками часу не виходить за допустимі межі, що визначаються точністю методики вимірювання досліджуваного показника, це свідчить про статичності об'єкта. Стосовно ймовірнісних систем їхня статичність встановлюється за мінливістю рівня її відносної організації. Якщо мінливість цього рівня не перевищує допустимі межі, система визначається як статична.

Дуже важливим є вибір відрізків часу, на яких встановлюється статичність або динамічність об'єкта. Якщо об'єкт на малих відрізках часу виявився статичним, то при збільшенні цих відрізків результат не зміниться Якщо статичність встановлена для великих відрізків часу, то при їх зменшенні результат може змінитися і статичність об'єкта може перейти в динамічність.

При виборі типу (класу) моделі ймовірнісного об'єкта важливим є встановлення його стаціонарності. Зазвичай про стаціонарність або нестаціонарність імовірнісних об'єктів судять щодо зміни в часі параметрів законів розподілу випадкових величин. Найчастіше для цього використовують середнє арифметичне випадкової величини $M(\tau_i)$ та середнє квадратичне відхилення випадкових величин σ_i (i =1, 2, ..., *n)* від середнього арифметичного та середнього квадратичного відхилення в часі.

З ряду середніх арифметичних $M(\tau_1)$, $M(\tau_2)$,…, $M\{\tau_i)$ вибирається мінімальне значення $M(\tau min)$ і будуються інтервали з межами

$$M(\tau_{min})+\Delta x, M(\tau min)-\Delta x$$

Δx -точність методики вимірювання досліджуваного показника .

Якщо значення $M(\tau_i)$ укладається в цей інтервал, то об'єкт визначається як стаціонарний по середньому арифметичному $M(\tau)$.

Аналогічно визначається стаціонарність за середнім квадратичним відхиленням.

Граничні значення а при встановленні стаціонарності визначаються за формулами:

$$\sigma_i = \sqrt{D_1}/n, \sigma_i = \sqrt{D_2}/n \quad \text{або } D_1 = \sum(x_i - (M(min) + \Delta x))^2)1/(n-1);$$

$$D_2 = \sum(x_i - (M(\tau_{min}) + \Delta x))^2)1/(n-1);$$

тут *n* - кількість спостережень.

Якщо всі значення укладаються в інтервал $1 \ldots {}_{2,\text{то}}$ об'єкт вважається стаціонарним. В іншому випадку об'єкт визначається як імовірнісний нестаціонарний , навіть якщо величина середнього арифметичного *M* не змінюється в часі.

Встановлення загальних показників об'єкта дозволяє вибрати математичний апарат, з якого будується математична модель. Вибір математичного апарату може

бути здійснений відповідно до схеми, представленої на рис. 6.2. Як очевидно з цієї схеми, вибір математичного апарату перестав бути однозначним і жорстким.

Так, для детермінованих об'єктів може використовуватися апарат лінійної та нелінійної алгебри, теорії диференціальних та інтегральних рівнянь, теорії автоматичного регулювання.

Адекватним математичним апаратом для моделювання ймовірнісних об'єктів є теорія детермінованих та випадкових автоматів з детермінованими та випадковими середовищами, теорія випадкових процесів, теорія марківських процесів, евристичне програмування, методи теорії інформації, методи теорії управління та оптимальні моделі.

При описі квазідетермінованих (імовірнісно-детермінованих) об'єктів може використовуватися теорія диференціальних рівнянь з коефіцієнтами, що підпорядковуються певним законам.

Мета та завдання, які ставляться при математичному моделюванні, відіграють важливу роль при виборі типу (класу) моделі. Практичні завдання вимагають простого математичного апарату, а фундаментальні — складнішого, допускають проходження ієрархії математичних моделей, починаючи від суто функціональних і закінчуючи моделями, що використовують твердо встановлені закономірності та структурні параметри.

Не менше значення на вибір моделі надає аналіз інформаційного масиву, отриманого як результат аналітичного огляду результатів досліджень інших авторів або пошукового експерименту. Розподіл масиву на залежні та незалежні фактори, на вхідні та вихідні змінні, попередній пошук взаємозв'язку між різними даними вибірки дозволяє визначити адекватний математичний апарат.

Аналіз інформаційного масиву дозволяє встановити безперервність або дискретність досліджуваного показника та об'єкта в цілому.

У безперервних об'єктах всі сигнали являють собою безперервні функції часу. У дискретних об'єктах всі сигнали квантуються за часом та амплітудою. Якщо сигнали квантуються тільки за часом, тобто подаються у вигляді імпульсів з рівною амплітудою, **такі** об'єкти називають дискретно-безперервними.

Встановлення безперервності об'єкта дозволяє використовувати для його моделювання диференціальні рівняння. У свою чергу, дискретність об'єкта зумовлює використання математичного моделювання апарату теорії автоматів.

Крім вищевикладеного на встановлення типу (класу) математичної моделі може надати істотний вплив необхідність певного відображення гіпотези.

Облік цілей і завдань математичного моделювання, характеру гіпотези та аналізу інформаційного масиву дозволяє конкретизувати модель, тобто в обраному типі (класі) моделей визначити їхній вид. Вибір виду математичної моделі у цьому класі є третім етапом математичного моделювання. Даний етап пов'язаний із завданням областей визначення досліджуваних параметрів об'єкта, тобто значення, які є допустимими, та встановлення залежностей між ними. Для кількісних (числових) параметрів залежності задаються як системи рівнянь (алгебраїчних чи диференціальних), для якісних — використовуються табличні способи завдань функцій. Якщо параметри описуються суперечливими залежностями, визначаються їх вагові коефіцієнти, виражені у частках одиниці, балах. Тим самим суперечливі залежності перетворюються на імовірнісні

Для опису складних об'єктів з великою кількістю параметрів можливе розбиття об'єкта на елементи (підсистеми), встановлення ієрархії елементів та описи зв'язків між ними на різних рівнях ієрархії.

Особливе місце на етапі вибору виду математичної моделі займає опис перетворення вхідних сигналів у вихідні характеристики об'єкта.

Якщо попередньому етапі було встановлено, що об'єкт є статичним, то побудова функціональної моделі здійснюється з допомогою алгебраїчних рівнянь. При цьому крім найпростіших залежностей алгебри використовуються регресійні моделі і системи алгебраїчних рівнянь.

Якщо заздалегідь відомий характер зміни досліджуваного показника, то кількість можливих структур моделей алгебри різко скорочується і перевага віддається тій структурі, яка виражає найбільш загальну закономірність або загальновідомий закон. Якщо характер зміни досліджуваного показника невідомий, то ставиться пошуковий експеримент. Перевага надається тій математичній формулі, яка дає найкращий збіг з даними пошукового експерименту.

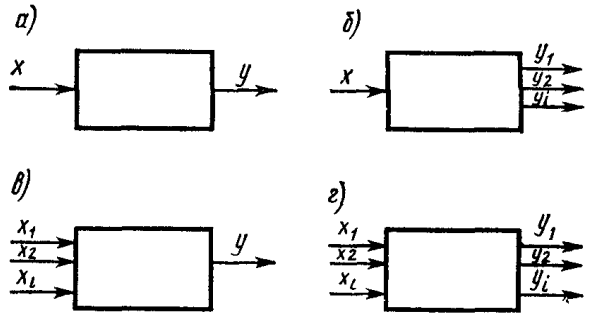
Результати пошукового експерименту та апріорний інформаційний масив дозволяють встановити схему взаємодії об'єкта із зовнішнім середовищем за співвідношенням вхідних та вихідних величин. У принципі можливе встановлення чотирьох схем взаємодії:

одномірно-одномірна схема (рис. 6.3 а) — на об'єкт впливає лише один фактор, а його

поведінка розглядається за одним показником (один вихідний сигнал);

одновимірно-багатомірна схема (рис. 6.3, *б)* - на об'єкт впливає один фактор, а його поведінка оцінюється за декількома показниками;

багатовимірно-одномірна схема (рис.63, *в)* - на об'єкт впливає кілька факторів, а його поведінка оцінюється за одним показником;



Мал. 6 3. Схеми взаємодії об'єкта із зовнішнім середовищем

багатовимірно-багатомірна схема (рис. 6.3, г) - на об'єкт впливає безліч факторів і його поведінка оцінюється по множині показників.

При одномірно-одномірній взаємодії статичного стаціонарного детермінованого об'єкта із зовнішнім середовищем постійна вхідна дія зв'язується з постійним вихідним сигналом через постійний коефіцієнт. Якщо цей об'єкт є нестаціонарним, то зазначений зв'язок описується різними функціями $y = f(x)$. Найчастіше ця функція описується поліномом.

У разі виявлення багатовимірно-одномірної схеми статичний стаціонарний детермінований об'єкт описується такою моделлю:

при рівнозначності зовнішніх впливів:

$$y = a \sum x_i;$$

при нерівнозначності зовнішніх впливів:

$$y = \sum a_i x_i;$$

де a_i - Постійний коефіцієнт; m - Число зовнішніх впливів (факторів).

Для статичного нестаціонарного об'єкта (при тій же схемі взаємодії) часто використовується модель у вигляді повного статечного полінома:

$$y = a_0 + \sum a_i x_i + \sum\sum a_{ij} x_i x_j + \sum\sum\sum a_{ijv} x_i x_j x_v + \ldots$$

де m_1, m_2 - Число парних і потрійних поєднань факторів ($m_1 == C_m^2$; $m_2 = C_m^3$).

При одновимірно-багатомірній схемі статичний стаціонарний та нестаціонарний об'єкт описується аналогічно одномірно-одномірній схемі взаємодії статичного стаціонарного об'єкта із зовнішнім середовищем. При цьому визначаються окремо математичні моделі вхідної дії з кожним вихідним сигналом. Вихідні сигнали вважаються незалежними.

Багатомірно-багатомірна взаємодія зводиться до багатовимірно-одномірної та математична модель об'єкта приймається аналогічною до викладеної вище. Для нестаціонарної одновимірно-одномірної (багатомірної) взаємодії алгебраїчні функції можуть являти собою розв'язання диференціальних рівнянь. При цьому необхідно розглядати похідні математичного очікування змінного фактора. Наприклад, експоненційна залежність може бути рішенням наступного диференціального рівняння:

$$dy / dx + a\, y = a\, y^m \text{ (при } x = 0)$$

де y - максимальне значення математичного очікування.

Вибір виду моделі динамічного об'єкта зводиться до складання диференціальних рівнянь. Модель динамічного об'єкта може бути побудована і в класі алгебраїчних функцій. Однак такий підхід є обмеженим, тому що не дозволяє в математичному описі врахувати вплив вхідних впливів на динаміку виходу без істотної перебудови самих функцій алгебри (структури і коефіцієнтів).

Тому за повнотою моделі надається перевага математичним моделям, побудованим у класі диференціальних рівнянь.

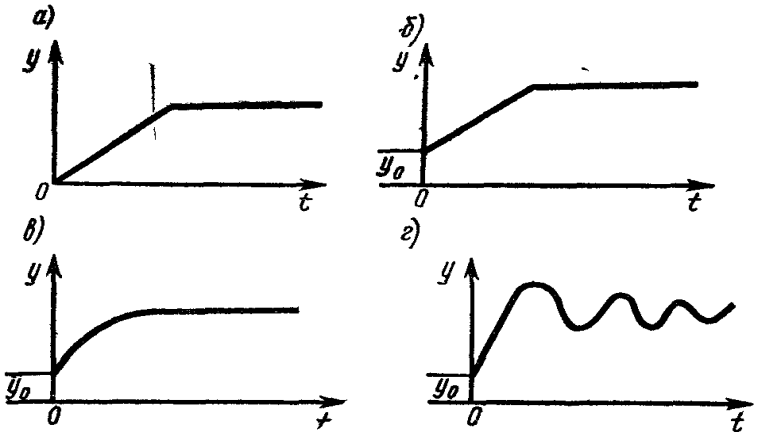
Якщо цікаві дослідника змінні є лише функціями часу, то моделювання застосовуються прості диференціальні рівняння. Якщо ж ці змінні є також функціями просторових координат, то для опису таких об'єктів недостатньо звичайних і слід скористатися складнішими диференціальними рівняннями в приватних похідних.

Методологія моделювання динамічних систем у класі диференціальних рівнянь суттєво залежить від схеми взаємодії об'єкта із середовищем та ступенем знання входу та виходу об'єкта.

Розглянемо випадок, коли вхід та вихід об'єкта відомі.

При одномірно-одномірному та одномірно-багатомірному взаємодії детермінованого об'єкта з середовищем структура диференціального рівняння визначається за видом вихідної характеристики об'єкта для типового вхідного впливу (наприклад, ступінчастого).

Однією з найпростіших вихідних характеристик об'єкта є лінійна (рис. 6.4 а). Така зміна виходу визначається рішенням диференційного



Рис» 6 4. Вихідні характеристики детермінованого об'єкта при ступінчастому зовнішньому впливі рівняння

$$dy / dt = kx , \quad y\ 0 = 0,$$

де $k > 0$-коефіцієнт розмірності та пропорційності; y_0 - Початкове значення вихідного сигналу; t – час.

Якщо *yo≠O,* то вихідна характеристика об'єкта відповідає рис. 6.4, *б.* Проте диференціальне рівняння залишається незмінним.

Більш складний вид реакції об'єкта на ступінчасту вхідну дію (рис. 6.4 в) може бути описаний повним неоднорідним диференціальним рівнянням першого порядку:

$$dy/dt + a_0 y = kx, \ y(0) = Yo ,$$

де а - коефіцієнт диференціального рівняння.

Реакція об'єкта, що відповідає рис. 6.4 *г* дозволяє використовувати як математичну модель об'єкта диференціальне рівняння другого порядку:

$$d^2 y/dt^2 + a_1 dy/dt + a\ 0_y = kx, \ (0) - = y\ 0,.$$

Розглянуті види математичних моделей відповідають постійному вхідному впливу (x = const). Якщо вхідні дії є деякими функціями часу, зокрема алгебраїчними, то наведених диференціальних рівняннях змінюються праві частини $x = f(t)$.

При багатовимірно-одномірному взаємодії (у разі детермінованого об'єкта) динамічні моделі також шукають у класі диференціальних рівнянь. При цьому допускається, що вхідні фактори є незалежними щодо їх дії на об'єкт. Усі чинники наводяться до суми коефіцієнтами чутливості у правій частині диференціального рівняння. Диференціальне рівняння підбирається на вигляд вихідний характеристики об'єкта.

Якщо припущення про незалежність дії факторів неправомірне, то попередньо встановлюється вплив кожного з факторів на вихідний показник об'єкта та за видом вихідних характеристик підбираються відповідні диференціальні рівняння. При одночасному дії чинників вважається, що вихідна характеристика об'єкта є сумою рішень незалежних диференціальних рівнянь, відповідних кожному фактору.

Для багатовимірно-багатомірної взаємодії побудова динамічних моделей може бути зведена до багатовимірно-одномірної схеми взаємодії.

За відсутності апріорної інформації про вхід і вихід об'єкта диференціальні рівняння, що моделюють динаміку об'єкта, складаються на основі припущень або знань про властивості та структуру об'єкта.

Універсального методу складання диференціальних рівнянь немає, можна використовувати деякі загальні підходи до складання рівнянь першого порядку.

Геометричні або фізичні завдання зазвичай призводять до одного з трьох видів рівнянь:

1) диференціальні рівняння у диференціалах;

2) диференціальні рівняння у похідних;

3) найпростіші інтегральні рівняння з подальшим перетворенням їх у диференціальні рівняння.

Розглянемо, як складаються рівняння кожного з наведених видів окремо.

1. *Рівняння у диференціалах* . При складанні диференціальних рівнянь першого порядку зручно застосовувати метод диференціалів. Сутність його у тому, що з умови завдань складаються наближені співвідношення між диференціалами. Для цього малі збільшення величин замінюються їх диференціалами, нерівномірно протікають фізичні процеси протягом малого проміжку часу dt розглядаються як рівномірні.

Ці припущення та заміни не відбиваються на остаточних результатах внаслідок того, що заміна прирощень диференціалами зводиться до відкидання нескінченно малих вищих порядків дрібниці. Так як відношення диференціалів функції та аргументу є межею відношення їх прирощень, то в міру того, як прирости прагнуть нуля, прийняті припущення виконуються з великою точністю. Диференціальні рівняння, що виходять при цьому, виявляються точними, якщо вони однорідні і лінійні щодо диференціалів.

Розглянемо геометричний приклад застосування методу диференціалів.

Нехай перед дослідником стоїть завдання визначення поверхні обертання, по якій потрібно відшліфувати дзеркало рефлектора, щоб світлові промені, що виходять з однієї точки після відображення в дзеркалі, перетиналися в іншій точці.

По суті, дане завдання зводиться до знаходження рівняння перерізу поверхні, що шукається меридіанною площиною, що проходить через точку *PZ* , в якій міститься джерело світла, і точку *F* , в якій перетинаються відбиті промені; (рис. 6.5).

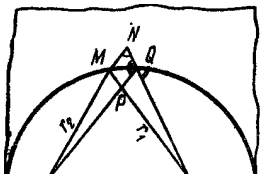


Рис. 6 5. Расчетная схема

Пусть *MQ* - мала дуга цього перерізу. Будемо її вважати прямо лінійним відрізком. Опишемо з точок *F* 1 і F *2 як* з центрів, дуги *MN* і *MP* кіл радіусами *F 1 M = r 1* і F *M = r 2* . Ці дуги також вважатимемо прямолінійними відрізками.

Трикутники MQN і *MQP* — прямокутні (∠MNQ і ∠MPQ — прямі) із загальною гіпертенузою MQ .

Користуючись відомим законом оптики про рівність кутів падіння та відображення, а також властивістю рівності вертикальних кутів, знаходимо, що ∠ *MQN = ∠ MQP* і ΔM *QN = ΔMQP ,* Звідси випливає, що *QN = QP* . Оскільки QN = Δri , а QP =Δ r 2 , то, замінюючи збільшення радіусів-векторів r 1 і r 2 їх диференціалами, маємо

$$dr\ 1 + dr\ 2 = 0.$$

Диференційне рівняння складено. Воно легко інтегрується. Для цього перепишемо його так:

d (r 1+ r 2) = 0, звідки знаходимо загальний інтеграл

$$r\ 1 + r\ 2 = C$$

Отже , переріз поверхні, що шукається, меридіанною площиною є еліпсом. Отже, дзеркало рефлектора треба відшліфувати поверхнею еліпсоїда обертання.

2. Рівняння у похідних. Для складання диференціальних рівнянь використовується видозмінений спосіб диференціалів, який називають шляхом похідних.

Сутність методу похідних у тому, що з умови завдання складаються наближені співвідношення між швидкостями зміни функції та аргументу. У цьому часто

використовується геометрична інтерпретація швидкості — кутовий коефіцієнт дотичної. Відсутність нескінченно малих у методі похідних здається, оскільки швидкість зміни досліджуваної величини сама з'явилася з розгляду нескінченно малих елементів.

При дослідженні зростання числа публікацій у науці виходять із припущення, що швидкість зростання dy / dt пропорційна досягнутому рівню у числа публікацій. Це тотожно твердженню, що відносна швидкість зростання 1/ y * dy / dt = const .

Зазначене припущення дозволяє скласти диференційне рівняння у формі

$$dy/dt=ky, \quad \text{або} \quad 1/y*dy/dt=k\ (k>0),$$

де ke = const , що характеризує в середньому відгуки на публікації в тій чи іншій галузі знання.

Вирішення цього диференціального рівняння має вигляд

$$y = ae^{kt},$$

де а - Постійна, що характеризує деяке початкове число публікацій.

3. Найпростіші інтегральні рівняння. Під час розгляду роботи сил, обсягів тіл, площ криволінійних поверхонь їх можна описати з допомогою певного інтеграла чи інтегральних формул. Якщо при такому описі невідомі функції потрапляють під знак інтеграла, то формальний запис, що отримується, називається інтегральним рівнянням.

Подальше диференціювання інтегрального рівняння перетворює їх у диференціальне.

Для ілюстрації методу побудови інтегральних рівнянь з подальшим перетворенням їх у диференціальні розглянемо таку задачу.

Нехай потрібно знайти закон прямолінійного руху матеріальної точки маси *m,* якщо відомо, що робота чинної на точку сили пропорційна часу *t ,* що пройшов від початку руху. Початковий шлях і початкова швидкість рівні відповідно до So і *vo* .

З курсу механіки відомо, що у разі прямолі нейного переміщення точки, коли напрямки сили та швидкості збігаються, робота A = $\int_{s_0}^{s} F(u)du = kt$, де $F(u)$ - сила, що діє на точку. За умовою завдання

$$A = kt .$$

Порівнюючи обидва вирази для *A,* знаходимо $\int_{s_0}^{s} F(u)du = kt$

Диференціюванням по s отримуємо F (s) = k dt / ds , а так як dt / ds = *v* - швидкість руху і dt / ds = 1 / (ds / dt) = 1 / v то $F(s)$ = k / v .

З іншого боку, з другого закону Ньютона випливає, що

$$F(s) = mdv / dt .$$

За

Сравнивая оба выражения для *F(s),* составляем дифференциальное уравнение

$$mdv/dt= k/v.$$

Общее решение этого уравнения представляется в виде

$$mv^2/2=kt+C_1,.$$

початкової умови *v* = *Vo* при t = 0 знаходимо, що

$$C1 = mv_0^2 /2.$$

$$v = \sqrt{\frac{2k}{m}t + v_0^2}$$

$$C_2 = s_0 - \frac{mv_0^3}{3k}$$

Отже,

Замінюючи *v* на ds / dt та інтегруючи, знаходимо

$$s = \frac{m}{3k}\left(\frac{2k}{m}t + v_0^2\right)^{3/2} + C_2$$

При t =0 s = So , отже,

Таким чином, закон руху матеріальної точки набуває вигляду

$$s = \frac{m}{3k}\left(\frac{2k}{m}t + v_0^2\right)^{3/2} + s_0 - \frac{mv_0^3}{3k}$$

$$C_2 = s_0 - \frac{mv_0^3}{3k}$$

При складанні диференціальних рівнянь регульованих об'єктів необхідно передусім визначити умови отримання рівноважного режиму роботи об'єкта, тобто. рівняння статичної рівноваги

У багатьох випадках рівняння статичної рівноваги виявляється загальним для різних об'єктів дослідження. Наприклад, при поступальному русі досліджуваний об'єкт перебуватиме у стані статичного рівноваги (рух буде рівномірним) тільки в тому випадку, коли рушійні сили F $_д$ дорівнюють силам опору *Fc* . Рівняння статичної рівноваги набуває вигляду

F д- Fc = 0.

У тому випадку, коли об'єкт дослідження (колінний вал двигуна внутрішнього згоряння, ротори електродвигуна і генератора, турбіни і т.д.) здійснює обертальний рух, то умовою статичної рівноваги є рівність крутного моменту $M_д$ моменту опору *Mc,* тобто M д- *Mc = 0.*

При виконанні цієї умови обертання валу буде рівномірним.

$$V_{пр} - V_{расх} = 0$$

Для регульованого резервуара, у якому необхідно підтримувати постійний рівень, умовою статичної рівноваги є рівняння

де Vnp -прихід рідини в резервуар; V витрата-витрата рідини.

$$G_{пр} - G_{расх} = 0$$

Аналогічно виглядає рівняння статичної рівноваги регулятора ресивера з газом або парою

де Gnp , Gpacx -маси приходить і йде з ресивера газу.

Якщо регульованим об'єктом є обсяг, в якому повинна підтримуватися постійна

$$Q_{пр} - Q_{расх} = 0$$

температура, то умова статичної рівноваги набуває вигляду

де Qnp - кількість теплоти, що надходить в об'єм за одиницю часу; Qpacx - кількість теплоти, що йде з об'єму в одиницю часу.

Перехідний процес у досліджуваному об'єкті може з'явитися тільки в тому випадку, якщо буде порушена статична рівновага. Поява збільшення одного з членів рівняння статичної рівноваги неминуче спричинить збільшення і другого члена рівняння, причому такі збільшення, як правило, не рівні між собою. В результаті умова статичної

$$F_c + \Delta F_c \neq F_c + \Delta F_c$$

рівноваги порушується і тоді при поступальному русі

при обертальному русі

при заповненні резервуару рідиною

при заповненні ресивера газом

$$M_c + \Delta M_c \neq M_c + \Delta M_c$$

при порушенні теплового режиму

$$G_{пр} + \Delta G_{пр} \neq G_{расх} + \Delta G_{расх}$$

$$Q_{пр} + \Delta Q_{пр} \neq Q_{расх} + \Delta Q_{расх}$$

Отримані нерівності можуть бути спрощені, якщо врахувати в них умову статичної рівноваги:

$$\Delta F_{д} \neq \Delta F_{с}; \quad \Delta M_{д} \neq \Delta M_{с}; \quad \Delta V_{пр} \neq \Delta V_{расх}; \quad \Delta G_{пр} \neq \Delta G_{расх}; \quad \Delta Q_{пр} \neq \Delta Q_{расх}.$$

Таким чином, при порушенні умов статичної рівноваги в об'єкті, що досліджується, виникає надлишок або недолік рушійних сил (або моментів), надходження рідини, газу, надлишок або недолік теплоти. Цей надлишок чи недолік викликає зміну характеру роботи об'єкта.

Подальше перетворення отриманих нерівностей здійснюється шляхом залучення загальновідомих залежностей, принципів, законів чи аналізу вихідних характеристик об'єкта, одержаних у пошуковому експерименті. При цьому можуть використовуватися феноменологічні закони (такі як закони Гука, Фур'є), напівемпіричні співвідношення (наприклад, додаткове співвідношення Ньютона в теорії удару) і суто емпіричні співвідношення.

Порушення статичної рівноваги об'єкта, що досліджується, що має поступальний рух, призведе до прискореного або уповільненого руху. Оскільки об'єкт, що рухається, має масу *m* і прискорення *a,* то на підставі принципу Даламбера рівняння динамічної рівноваги (рівняння руху) може бути записане у вигляді

$$ma = \Delta F_{д} - \Delta F_{с},$$

$$m\frac{dv}{dt} = \Delta F_{д} - \Delta F_{с}$$

або,

де *v* - Швидкість руху; *t* - час,

Для об'єкта, що має обертальний рух, рівняння динамічної рівноваги записується на підставі принципу Даламбера. Якщо J - наведений до осі валу двигуна момент інерції

$$a = \frac{dv}{dt}$$

$$J\frac{d\omega}{dt} = \Delta M_{д} - \Delta M_{с}$$

деталей, що обертаються або зворотно-поступально рухаються, і w - кутова швидкість обертання валу, то рівняння руху отримує вигляд

При порушенні статичної рівноваги роботи резервуара в ньому відбувається накопичення (в сенсі алгебри) рідини і, як наслідок, зміна її рівня.

якщо dV — зміна об'єму рідини в резервуарі за час *dt* , то

$$y = f(x_1, x_2, \ldots, x_n)$$

або

$$\frac{dV}{dt} = \Delta V_{\text{пр}} - \Delta V_{\text{расх}}$$

При порушенні рівноваги в роботі ресивера кількість газу в ньому змінюється на величину dG за елементарний інтервал часу d *t тому*

$$dG = \left(\Delta G_{\text{пр}} - \Delta G_{\text{расх}}\right)dt$$

звідки

$$\frac{dG}{dt} = \Delta G_{\text{пр}} - \Delta G_{\text{расх}}$$

Порушення теплового режиму деякого об'єму призводить до зміни кількості теплоти, зосередженої в обраному об'ємі, на dQ з а інтервал часу *dt* .

$$dQ = \left(\Delta Q_{\text{пр}} - \Delta Q_{\text{расх}}\right)dt$$

$$\frac{dQ}{dt} = \Delta Q_{\text{пр}} - \Delta Q_{\text{расх}}$$

або

Порівняння отриманих диференціальних рівнянь для різних об'єктів, що регулюються, показує їх ідентичність.

Їх подальше перетворення можливе на основі знання фізики загальних властивостей газу, теплопередачі тощо. Наприклад, відомо, що стан ідеальних газів описується рівнянням Клапейрона

pV = GTR

де *p*- тиск газу в ресивері: V - обсяг ресивера;

G - кількість газу, що знаходиться в ресивері; *T* - абсолютна температура газу; R -газова постійна. З рівняння стану випливає

G = pV / RT

$$\frac{V}{RT}\frac{dp}{dt} = \Delta G_{\text{пр}} - \Delta G_{\text{расх}}$$

Тоді рівняння динамічної рівноваги ресивера перетворюється на вигляд

Опис прирощень у правій частині розглянутих рівнянь також вимагає попереднього знання зв'язків із властивостями об'єкта.

Введення у рівняння динамічної рівноваги залежностей, що описують збільшення, часто призводить до підвищення порядку диференціальних рівнянь. Однак за деякого спрощення порядок диференціальних рівнянь вдається знизити. Таким спрощенням у більшості випадків є нехтування інерційністю об'єкта або лінеаризація прирощень. Для лінеаризації останніх часто використовують розкладання функції до ряду Маклорена.

Нехай, наприклад, момент, що крутить, об'єкта (валу двигуна), що має обертальний

$$M_{\text{д}} = f(\omega, h).$$

рух, залежить від кутової швидкості обертання w і положення *h* органу управління паливного насоса двигуна, тобто.

Для визначення збільшення обертального моменту двигуна в залежності від збільшення w і *h* отриману функцію слід розкласти в ряд Маклорена:

$$M_д + \Delta M_д = M_д + \frac{\partial M_д}{\partial \omega} d\omega + \frac{\partial^2 M_д}{\partial \omega} \frac{d\omega^2}{2!} + \text{[illegible]} + \frac{\partial M_д}{\partial h} dh + \frac{\partial^2 M_д}{\partial h^2} \frac{\Delta h^2}{2!} + \text{[illegible]}$$

Якщо в розкладанні замінити нескінченно малі величини dω і dh величинами Δω і

$$M_д + \Delta M_д = M_д + \frac{\partial M_д}{\partial \omega} \Delta\omega + \frac{\partial^2 M_д}{\partial \omega} \frac{d\omega^2}{2!} + \text{[illegible]} + \frac{\partial M_д}{\partial h} \Delta h + \frac{\partial^2 M_д}{\partial h^2} \frac{\Delta h^2}{2!} + \text{[illegible]}$$

Δh кінцевими, але досить малими, то ряд матиме вигляд

Похідні в цьому розкладанні повинні підраховуватися в положенні рівноваги режиму роботи двигуна , за якого виконується умова статичної рівноваги.

При малых конечных приращениях Δω и Δh и неразрывности функции $M_д = f(\omega, h)$ можно отбросить без внесения существенной ошибки все члены ряда с Δω и Δh в степенях выше первой, т. е. практически произвести замену действительной функции $M_д = f(\omega, h)$ ее касательной в точке равновесного режима (такая замена обычно называется линеаризацией). Після лінеаризації збільшення крутного моменту набуває вигляду

$$\Delta M_д = \frac{\partial M_д}{\partial \omega} \Delta\omega + \frac{\partial M_д}{\partial h} \Delta h$$

Нехай момент опору є лише функцією кутової швидкості

$$M_c = f(\omega)$$

Тоді після розкладання цієї функції до ряду Маклорена та лінеаризації отримаємо

$$\Delta M_c = \frac{dM_c}{d\omega} \Delta\omega$$

Після підстановки перетворених прирощень до рівняння динамічної рівноваги

$$J \frac{d\omega}{dt} + \left(\frac{\partial M_c}{\partial \omega} - \frac{\partial M_д}{\partial \omega} \right) \Delta\omega = \frac{\partial M_д}{\partial h} \Delta h$$

об'єкта отримуємо

Будь-які диференціальні рівняння - це модель цілого класу явищ, тобто сукупність явищ, що характеризуються однаковими процесами. При інтегруванні рівнянь одержують велику кількість рішень, що задовольняють вихідного диференціального рівняння. Щоб одержати з безлічі можливих рішень одне, що задовольняє процесу, що розглядається, необхідно задати додаткові умови диференціального рівняння. Вони повинні чітко виділити явище, що вивчається, з усього класу явищ.

Умови, які розкривають всі особливості даного рівняння, називаються умовами однозначності та характеризуються такими ознаками: геометрією системи (форма та розміри тіла) фізичними властивостями тіла (тепловроводність, вологопровідність, пружність тощо) початковими умовами, тобто станом системи у початковий момент; граничними умовами, тобто умовами взаємодії системи на кордонах з навколишнім середовищем. Початкові та граничні умови називають *крайовими.*

Розглянуті приклади вибору виду моделі об'єкта мають відношення лише до таких об'єктів, які можна розглядати як детерміновані.

Розглянемо методи вибору виду математичних моделей для імовірнісних об'єктів. Як і раніше, для

статичних об'єктів під стаціонарністю входу розумітимемо постійне його значення. Якщо вхідний сигнал приймає кілька значень, його вважатимемо не стаціонарним.

Нехай є одномірно-одномірна схема взаємодії об'єкта із зовнішнім середовищем. Якщо вплив на вході об'єкта постійно у часі, то як математичну модель статичного імовірнісного об'єкта може виступати певний закон розподілу вихідної величини. Якщо вхідний вплив може приймати різні значення і кожному значенню відповідає ряд значень вихідний величини об'єкта, то як модель імовірнісного об'єкта приймається набір законів розподілу вихідної величини для всіх значень вхідного впливу.

При моделюванні ймовірнісних об'єктів, крім законів розподілу вхідних і вихідних величин, істотний зв'язок між ними. Тому до складу моделі включають коефіцієнти взаємної кореляції та функції

$Hm = f(x); R = f(x); y_{cp} = f(x); \sigma = f(x),$

де *x*- вхідний вплив; *Hm* -максимальна ентропія вихідних характеристик; R -відносна організація вихідних характеристик; y_{cp} - середнє значення вихідний величини; σ-середньоквадратичне відхилення вихідних величин.

Максимальна ентропія вихідних характеристик оцінюється за формулою

$$Hm = \log_2 n$$

де n - Число станів об'єкта.

$$n = \frac{y_{\max} - y_{\min}}{\Delta y}$$

Для оцінки кількості станів об'єкта використовується формула

де У m ax, *Ут і п* - максимальне і мінімальне значення вихідної величини; Δу-точність вимірювання вихідних величин.

Відносна організація вихідних показників оцінюється за формулою Ферстера:

$$R = 1 - \frac{H}{H_m}$$

Де

$$H = -\sum_{i=1}^{N} \frac{m_i}{N} \log_2 \frac{m_i}{N}$$

Тут mi - Число появи *yi* значення вихідний характеристики; *N*- повне число спостережень вихідних характеристик

При багатовимірно-одномірній схемі взаємодії статичного імовірнісного об'єкта із зовнішнім середовищем завдання математичного моделювання зводиться до одномірно-одномірної схеми для кожного поєднання постійних вхідних впливів

Нестаціонарний випадок відрізняється тим, що кожна вхідна дія може набувати кількох значень . При цьому для кожного конкретного поєднання завдання аналізу зв'язку між входами і виходом може вирішуватися аналогічно задачі для багатовимірно-одномірної стаціонарної схеми.

Оцінка ступеня зв'язку виходу з входами проводиться шляхом зіставлення статичних параметрів та обчислення коефіцієнтів взаємної кореляції

Моделювання об'єкта при одновимірно-багатомірній та багатовимірно-багатомірній схемах взаємодії проводиться аналогічно вищевикладеному.

Розглянемо далі моделювання динамічних режимів імовірнісних об'єктів

При одномірно-одномірній схемі взаємодії об'єкта із зовнішнім середовищем і багаторазовому надходженні на його вхід однієї і тієї ж функції часу *x (t)* можливі два випадки-

на виході об'єкта спостерігається стаціонарний випадковий процес;

на виході об'єкта спостерігаєте випадковий процес.

У першому випадку як математична модель вихідної величини об'єкта приймається закон розподілу значень вихідний величини, що має одні й самі параметри для всіх

$$H_m(i,\Delta t) = f(x)$$
$$R(i,\Delta t) = f(x), \quad i = 1,2,3,\ldots$$

зрізів за часом Зрізи за часом приймаються з інтервалом t . Модель доповнюється залежностями

Ці залежності можуть бути представлені функцією алгебри або диференціальними рівняннями. У другому випадку нестаціонарний вихід об'єкта) як математичну модель об'єкта приймаються функціональні зв'язки

де m $_y$ *(i ,* Δ t)—математичне очікування розподілу вихідних величин *y* по

$$m_y(i,\Delta t) = f(x)$$
$$\sigma_y(i,\Delta t) = f(x), \quad i = 1,2,3,\ldots ,$$

дискретних відрізках часу з кроком Δ t ;σ(i , Δ t)—середньоквадратичне відхилення .

Ці зв'язки також описуються за допомогою алгебраїчних або диференціальних рівнянь.

Якщо на вхід об'єкта багаторазово надходять різні функції часу *Xi (t),* то їх розглядають як випадкові реалізації випадкового процесу. Для кожного перерізу в часі (i , Δt) визначаються інформаційні та імовірнісні характеристики $H^x{}_m$, ***Rx*** **,** *mx* , σx . Аналогічно, для кожної сукупності виходу для тих самих перерізів за часом (i , Δ t) визначаються інформаційні та ймовірнісні характеристики

$$H_m^{y_i}, R_{yi}, m_{yi}, \sigma_{yi}$$

Як математичну модель об'єкта приймаються функціональні зв'язки

$$H_m^{y_i}(i,\Delta t) = f_{1i}\left[H_m^x(i,\Delta t)\right]$$
$$R_{yi}(i,\Delta t) = f_{2i}\left[R_x(i,\Delta t)\right]$$
$$m_{yi}(i,\Delta t) = f_{3i}\left[m_x(i,\Delta t)\right]$$
$$\sigma_{yi}(i,\Delta t) = f_{4i}\left[\sigma_x(i,\Delta t)\right], \quad (i = 1,2,\ldots,k)$$

Іноді всі масиви реалізації вихідної величини сприймаються як єдиний масив. У цьому випадку в якості математичної моделі об'єкта приймаються функціональні зв'язки між узагальненими параметрами виходу та параметрами входу.

При багатовимірно-одномірній схемі взаємодії імовірнісного об'єкта з середовищем

$$H_m^y = f_{11}\left(H_m^{x_1}\right); \quad m_y = f_{31}(m_{x1});$$
$$\square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square$$
$$H_m^y = f_{1m}\left(H_m^{x_1}\right); \quad m_y = f_{3m}(m_{xn});$$

$$R_y = f_{21}(R_{x1}); \quad \sigma_y = f_{11}(\sigma_{x1});$$
$$\square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square$$
$$R_y = f_{2m}(R_{xm}); \quad \sigma_y = f_{2m}(\sigma_{xn});$$

у нестаціонарному режимі як його математичну модель приймаються функціональні залежності:

При одновимірно-багатомірній схемі взаємодії об'єкта в нестаціонарному режимі як математична модель об'єкта приймаються функціональні залежності

При багатовимірно-багатомірній схемі взаємодії в нестаціонарному режимі встановлюються функціональні залежності інформаційних та ймовірнісних характеристик для кожного входу та виходу. При цьому можна вважати, що будь-який вихід залежить від усіх входів і відповідно вибрати вид математичних моделей.

Процес вибору математичної моделі об'єкта закінчується попереднім контролем. У цьому здійснюються такі види контролю: размерностей; порядків; характеру залежностей; екстремальних ситуацій; граничних умов; математичної замкнутості; фізичного змісту; стійкість моделі.

$$H_m^{y_i}(i, \Delta t) = f_{1i}\left[H_m^x\right]$$
$$R_{yi}(i, \Delta t) = f_{2i}\left[R_x\right]$$
$$M_{yi}(i, \Delta t) = f_{3i}\left[m_x\right]$$
$$\sigma_{yi} = f_{4i}\left[\sigma_x\right], \quad (i = 1,2,\square\ ,n)$$
$$H_m^{y_i}, R_{yi}, m_{yi}, \sigma_{yi}$$

Контроль розмірностей зводиться до перевірки виконання правила, за яким прирівнюватися і складатися можуть лише величини однакової розмірності.

Контроль порядків спрямовано спрощення моделі. При цьому визначаються порядки величин, що складаються, і явно малозначні складові відкидаються.

Контроль характеру залежностей зводиться до перевірки напряму та швидкості зміни одних величин при зміні інших. Напрями і швидкість, які з математичної моделі, повинні відповідати фізичному змісту завдання.

Контроль екстремальних ситуацій зводиться до перевірки наочного рішення при наближенні параметрів моделі до нуля або нескінченності.

Контроль граничних умов полягає в тому, що перевіряється відповідність математичної моделі граничним умовам, що випливають із змісту завдання. При цьому перевіряється, чи дійсно граничні умови поставлені та враховані при побудові функції, що шукається, і що ця функція насправді задовольняє таким умовам.

Контроль математичної замкнутості зводиться до перевірки, що математична модель дає однозначне рішення.

Контроль фізичного сенсу зводиться до перевірки фізичного змісту проміжних співвідношень, що використовуються при побудові математичної моделі.

Контроль стійкості моделі полягає у перевірці того, що варіювання вихідних даних у рамках наявних даних про реальний об'єкт не призведе до істотної зміни рішення.

Лекція №6. Аналогове моделювання систем управління.

Побудова функціональних залежностей за експериментальними даними

6.1. Побудова функціональної залежності при однофакторному експерименті

Нехай при однофакторному експерименті є вибірка, що описує зміни вхідних параметрів, та набір вихідних величин (рис.3.1). Необхідно побудувати залежність $y = F(x)$.

Рис.3.1

Для аналізу експериментальних даних існує дуже багато способів завдання цієї залежності аналітичними та чисельними методами. Ми відзначимо лише найпоширеніші з них:

1. Подальша обробка може проводитись за безпосереднього чисельного використання масиву значень (x_i, y_i).

2. У разі коли кількість вимірювань i не надто велика і розкид значень (x_i, y_i) малий, залежність $y = F(x)$ може бути побудована шляхом *інтерполяції* (апроксимації) через всі експериментальні точки. У цьому випадку проводиться залежність $y = F(x)$ через усі точки з координатами (x_i, y_i). Найпростіший варіант проведення такої залежності полягає в побудові полінома (статечного ряду).

Нехай

$$F(x_0) = y_0,\ F(x_1) = y_1, \ldots F(x_n) = y_n. \qquad (3.1.1)$$

Інтерполіруюча функція

$$P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n.$$

Багаточлен $P_n(x)$ має $n+1$ член.

Вимагаючи виконання умови (3.1.1), отримаємо систему із $(n+1)$ рівнянь із $(n+1)$ невідомими:

$$\sum_{k=0}^{n} a_k x_i^{n-k} = y_i \qquad (3.1.2)$$

де кожному $i = 0, \ldots n$ відповідає своє рівняння.

Замість розв'язання системи рівнянь (3.1.2) на практиці використовуються зручніші та менш трудомісткі способи, зокрема:

- €€€€€€€€€€ інтерполювання *багаточленом Лагранжа* ;
- €€€€€€€€€€ інтерполювання *багаточленом Ньютона* .

Інтерполяційні формули Ньютона особливо зручні у разі рівновіддалених вузлів ($x_{i+1} - x_i$ однаково всім i) . Якщо i велика (велика кількість вузлів), інтерполяційний многочлен має високий рівень і виявляється незручним для обчислень.

3. За надто високого ступеня полінома проблеми можна уникнути, розбивши відрізок інтерполяції на кілька частин з побудовою для кожної їх свого інтерполяційного многочлена. Таке інтерполювання має серйозний недолік: у точках стику інтерполяційних багаточленів виявляється розривною перша похідна. На малюнку 3.2 показаний найпростіший спосіб такої інтерполяції експериментальної залежності - з'єднання сусідніх точок прямими (багаточлен ступеня $n = 1$) .

4. Якщо потрібно, щоб залежність мала безперервні похідні, користуються сплайнами.

Сплайн (від англ. spline - Рейка) - функція, що є алгебраїчним багаточленом на кожному відрізку $[x_i, x_{i+1}]$, і безперервна у всій області разом зі своїми похідними. Найчастіше користуються сплайнами третього ступеня. Відповідну залежність показано на рис.3.2 курсивом.

Рис.3.2.

5. При однофакторному експерименті, коли є результати багаторазових вимірювань із випадковою похибкою (див. параграф 2.2 цього посібника), проведення залежності через усі експериментальні точки безглуздо. Найчастіше у разі для побудови функціональної залежності користуються шляхом найменших квадратів (МНК).

Побудова функціональної залежності з допомогою методу найменших квадратів. Даний метод використовується тоді, коли число точок $\vec{n}$(вузлів) велике і побудова плавної залежності

$$y = F(x), \quad (3.1.3)$$

проходить через усі точки (x_i, y_i)неможливо через велике розкидання значень.

Функція (3.1.3) називається *рівнянням регресії на* x. Нехай наближена функція описує $y = F(x)$ залежить від трьох параметрів $y = F(x,a,b,c)$. Ця функція не проходитиме через всі точки з координатами (x_i, y_i), тоді можна знайти суму квадратів різниць

$$\sum_{i=1}^{n} [y_i - F(x_i,a,b,c)]^2 = \Phi(a,b,c) \quad . \quad (3.1.4)$$

Завдання зводиться до пошуку мінімуму $\Phi(a,b,c)$, тобто. до вирішення системи рівнянь

$$\begin{cases} \dfrac{\partial \Phi}{\partial a} = 0, \\ \dfrac{\partial \Phi}{\partial b} = 0, \\ \dfrac{\partial \Phi}{\partial c} = 0. \end{cases}$$

А саме

$$\begin{cases} \sum_{i=1}^{n}\left[y_i - F(x_i,a,b,c)\right]\cdot F'_a(x,a,b,c) = 0, \\ \sum_{i=1}^{n}\left[y_i - F(x_i,a,b,c)\right]\cdot F'_b(x,a,b,c) = 0, \\ \sum_{i=1}^{n}\left[y_i - F(x_i,a,b,c)\right]\cdot F'_c(x,a,b,c) = 0. \end{cases} \quad (3.1.5)$$

Вирішивши систему (3.1.5) щодо параметрів a,b,c, знаходимо конкретний вид шуканої функції.

Наближаюча (Наближена) функція може мати будь-який вид: лінійна залежність, парабола, синусоїда і т.д. Найчастіше використовуються алгебраїчні багаточлени не вище третього порядку. У більшості випадків аналізується лінійна регресія, коли

$$y = F(x) = ax + b. \quad (3.1.6)$$

Головна особливість *регресійного аналізу* полягає в тому, що регресія не x відповідає регресії x на (див. рис.3.3).

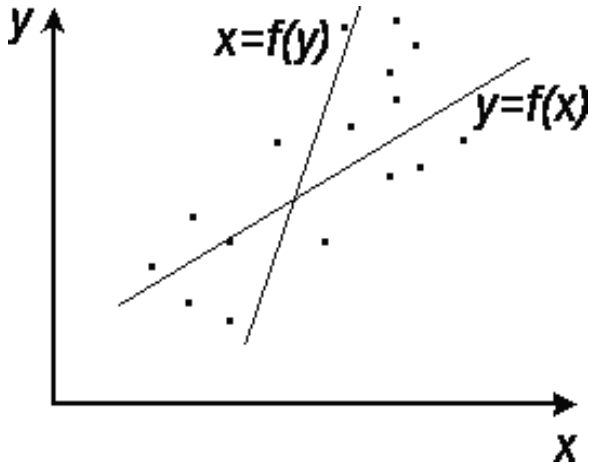


Рис.3.3.

Пояснимо цю властивість регресійних залежностей. Нехай формула регресії має вигляд (3.1.6), наведемо її зворотну функцію:

$$x = f(y) = \frac{y}{a} - \frac{b}{a}. \quad (3.1.7)$$

$\frac{b}{}$

Звернемо увагу, що у (3.1.7) вільний член a залежить від коефіцієнта нахилу a прямої залежності (3.1.6). При побудові регресії пряма проходить приблизно через середину області, що охоплює експериментальні точки і її нахил визначається ставленням розкиду значень по осях x і (перетин функцій $y = f(x)$ і $x = f(y)$ знаходиться в середині області експериментальних значень). Таким чином, регресія *x (y)* , побудована за експериментальними даними, не співпадатиме з (3.1.7) через наявність вільного члена.

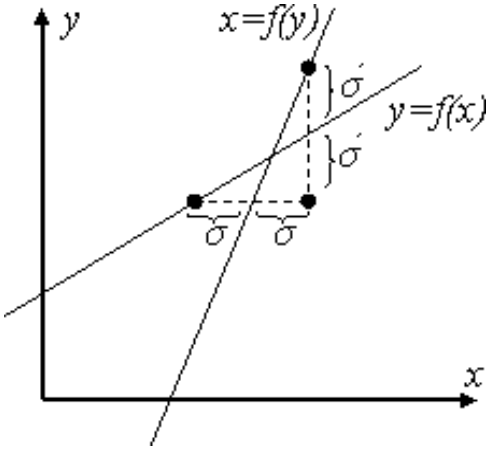


Рис.3.4

Графічно це пояснюється рис. 3.4, де за трьома експериментальними точками побудовані регресії *y (x)* і *x (y)* , які не збігаються. Для мінімізації СКО трьох експериментальних точок від прямої залежність повинна проходити через одну з них і в середині між двома іншими точками. Як очевидно з рис.3.4, лінійні регресії, побудовані з цих міркувань припиняються у центрі області експериментальних значень і мають різний нахил.

6.2. Швидкі методи побудови функціональних залежностей

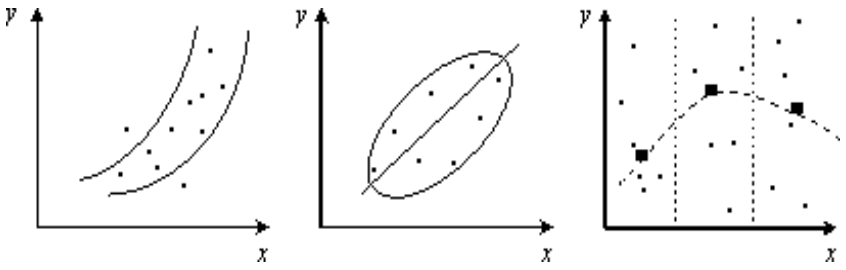
Завдання вибору виду функціональної залежності - завдання *неформализируемая* , оскільки одна й та експериментальна залежність може бути описана різними аналітичними виразами приблизно з однаковою точністю. Наприклад, U -подібна крива може бути описана як параболою, так і шматком синусоїди.

Основна вимога до математичної моделі – компактність та зручність використання, тому найчастіше користуються алгебраїчними багаточленами, експоненціальними та тригонометричними функціями. Інша вимога - інтерпретованість . Наприклад, якщо експериментальна залежність описує зміну амплітуди загасаючих коливань, функціональна залежність може бути побудована у вигляді $\frac{1}{ax}$ або e^{-ax}. У цьому випадку, зі знання природи залежності (теоретичної моделі загасаючих коливань), буде обрано експоненційну залежність e^{-ax}.

Похибка у виборі функціональної залежності називається *похибкою адекватності* моделі. Для її усунення слід розглядати теоретичну модель описуваного явища чи процесу.

Швидкі методи встановлення графічного вигляду однофакторних залежностей. Найпростіший експрес-метод статистичної обробки – *метод контуру* (рис.3.5, а, б).

Його суть – обведення експериментальних точок плавними кордонами. Вимога плавності має на увазі, що деякі точки можуть опинитися поза контуром (рис.3.5, а). Метод контуру можна використовувати тоді, коли розкид експериментальних точок невеликий (рис.3.5, б).



а б у

Рис.3.5

На малюнку 3.5 показано побудову експериментальної залежності суворішим експрес-методом, - *методом медіанних центрів* . Для цього область експериментальних даних розбивається вертикальними лініями на кілька областей (у даному випадку – три області), у кожній з яких знаходиться рівна кількість експериментальних точок. Медіанними центрами кожної з цих областей по координаті X є точки, праворуч і

ліворуч від яких знаходиться рівна кількість експериментальних відліків. Знайшовши таким чином координати x_i медіанних центрів, аналогічним чином у кожній області знаходять їх вертикальні координати y_i вище і нижче яких була б рівна кількість точок. Потім за точками з координатами x_i, y_i будується плавна експериментальна крива. Необхідно пам'ятати, що координати ( ) медіанних центрів не збігаються із середніми значеннями експериментальних даних.

Зв'язок коефіцієнта лінійної регресії, коефіцієнта кореляції та відносної похибки. Нехай за результатами однофакторного експерименту будується лінійна регресія $y = f(x) = ax + b$, тоді із системи (3.1.5) випливає:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad b = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \qquad (3.2.1)$$

З іншого боку коефіцієнт кореляції, що характеризує зв'язок між x_i і y_i, за визначенням

$$R = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} =$$

$$= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}}. \qquad (3.2.2)$$

Зіставляючи (3.2.1) і (3.2.2), знайдемо зв'язок між коефіцієнтом регресії a та коефіцієнтом кореляції R:

$$a = \left(\frac{\sigma_y}{\sigma_x}\right)^2 R, \qquad (3.2.3)$$

де σ_y, σ_x - середньоквадратичні відхилення x_i та y_i.

Таким чином, коефіцієнт кореляції пов'язаний з розкидом значень по осях x і визначає можливий ступінь відхилення лінії регресійної залежності по нахилу. Нехай величина σ_x фіксована,

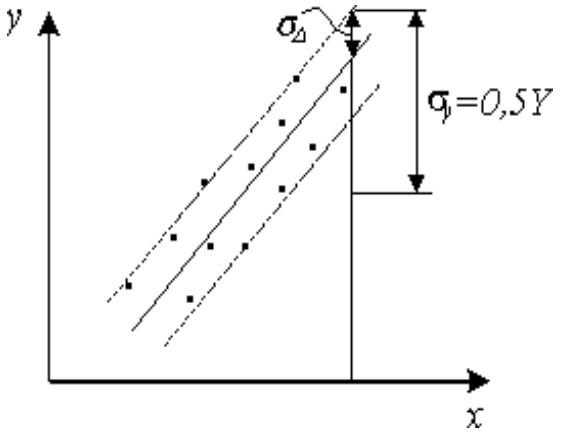


Рис.3.6

тоді можливе відхилення осі від середнього значення $\bar{y}$ становить $a\sigma_x + \sigma_\Delta$, де σ_Δ- середньоквадратичне відхилення від лінії регресії (див. рис.3.6). У зв'язку з цим, враховуючи (3.2.3), коефіцієнт кореляції дуже часто визначають як

$$R = \frac{1}{\sqrt{1+\left(\frac{\sigma_\Delta}{\sigma_y}\right)^2}}, \quad (3.2.4)$$

де σ_Δ- ширина смуги похибок по ; σ_y-Розкид значень , який визначається діапазоном зміни величини Y.

Оскільки в практичних випадках $\sigma_\Delta << Y$, то формулу (3.2.4) з урахуванням наближеного розкладання до першого члена до ряду Тейлора приводять до вигляду

$$R = \frac{1}{\sqrt{1+\left(\frac{\sigma_\Delta}{\sigma_y}\right)^2}} \approx \sqrt{1-\left(\frac{\sigma_\Delta}{\sigma_y}\right)^2} = \sqrt{1-(2\gamma)^2} \quad . \quad (3.2.5)$$

де $\gamma = \frac{\sigma_\Delta}{Y}$ – наведена похибка. Таким чином, у більшості практичних випадків зв'язок між коефіцієнтом кореляції та наведеною похибкою може бути встановлений за допомогою найпростішої наближеної формули (3.2.5).

Швидка оцінка коефіцієнта кореляції вихідних даних. Швидку оцінку коефіцієнта кореляції та похибки вихідних даних можна провести також методом медіанних центрів (рис.3.7).

Розіб'ємо поле експериментальних точок вертикальною рисою на дві рівні за кількістю точок області (8×8крапок). У лівій та правій частинах знайдемо медіанні центри. Проведена через ці медіанні центри, позначені зірочкою, пряма a– регресія на X. Тепер розіб'ємо експериментальну область на рівну кількість точок по вертикалі горизонтальною межею і, після знаходження відповідних медіанних центрів, отримаємо пряму b- регресію X на . Прямі a і b збігаються тільки в тому випадку, коли коефіцієнт кореляції між X_i і Y_i дорівнює одиниці, тобто $R = 1$.

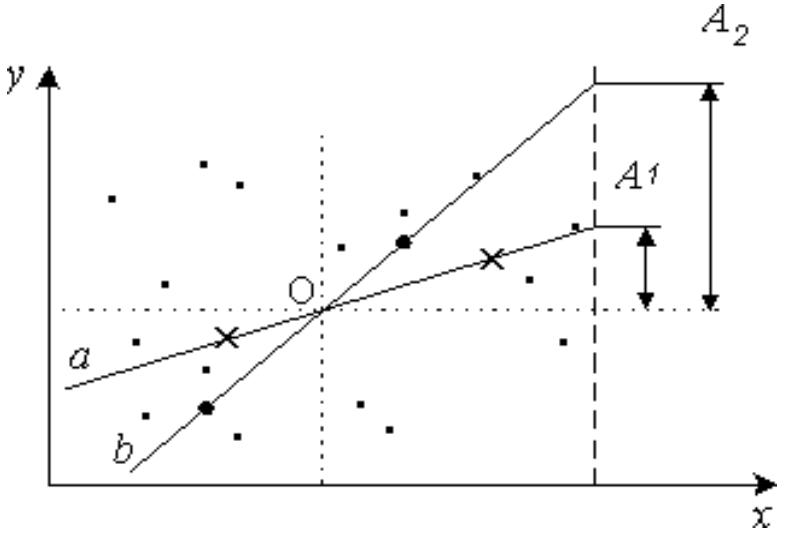


Рис.3.7

За відмінністю прямих a можна b з урахуванням (3.2.3) оцінити коефіцієнт кореляції:

$$R = \sqrt{\frac{A_1}{A_2}}, \quad (3.2.6)$$

де $\frac{A_1}{A_2}$ визначається ставленням кутів їхнього нахилу. Для швидкої оцінки відносної похибки підставимо величину R(3.2.6) у звернену формулу (3.2.5):

$$\gamma = \frac{\sqrt{1 - R^2}}{2}. \quad (3.2.7)$$

Таким чином, швидка оцінка коефіцієнта кореляції та значення відносної похибки ґрунтується на тому, що прямі aта bобов'язково проходять через точку перетину кордонів О. При цьому, чим вищий розкид експериментальних даних (невитягнута область), тим більше буде кут між прямими aі b.

При побудові регресійних залежностей методом медіанних центрів необхідно пам'ятати, що отримані лінії регресії в загальному випадку відрізняються від відповідних залежностей, отриманих за допомогою МНК. Їх відмінності зменшуватимуться зі збільшенням кількості експериментальних точок, якщо розкид експериментальних даних підпорядковується нормальному закону розподілу.

6.3 Згладжування експериментальних часових рядів

Однією з найбільш значущих завдань є побудова плавних тимчасових залежностей функції $x(t)$, якщо ця функція має випадкову складову. Дискретну залежність $x_i(t_i)$називають *тимчасовим рядом* чи, у випадку, дискретним *випадковим процесом* . Загальні властивості та методи опису випадкових процесів ми розглянемо у п'ятому розділі цього посібника. Порівняно з експериментальними залежностями $y = f(x)$, які ми вважали результатом однофакторного експерименту та розглянули вище, особливості експериментальних часових рядів полягають у наступному:

1. Насправді зазвичай доводиться аналізувати тимчасові ряди з досить великою кількістю відліків (щонайменше кілька десятків).

2. Відліки, як правило, проводяться через рівні проміжки часу (рівновіддалені вузли залежності при $t_{i+1} - t_i = const$)

3. Залежність $x_i(t_i)$свідомо немонотонна і найчастіше обмежена. Тому, якщо вибрати кінцевий інтервал Δx, то для будь-якого xзі збільшенням довжини часового ряду кількість значень x_i, що потрапляють в інтервал, $[x, x + \Delta x]$буде збільшуватися.

4. Наближувальну функцію, що апроксимує часовий ряд по всій його довжині, як правило, неможливо описати аналітично.

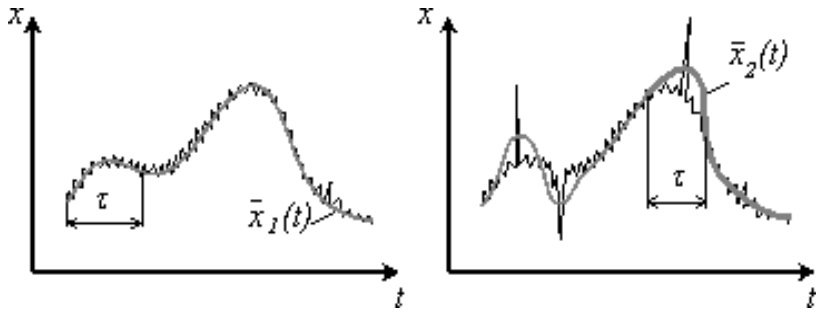
Особливості часових рядів добре пояснює найпростіший приклад. Проведемо однофакторний експеримент, реєструючи значення вхідних x_iта вихідних y_iпараметрів одночасно, в моменти часу t_iчерез рівні проміжки $t_{i+1} - t_i = const$. Тоді функціональна залежність $y_i = f(x_i)$, що описує результати експерименту, визначається параметрично, через два часових ряди $x_i(t_i)$, $y_i(t_i)$.

У теорії обробки часових рядів існує безліч способів їх згладжування: фільтрація з використанням перетворення Фур'є, кускова апроксимація багаточленами та ін.

1) метод ковзного середнього;

2) медіанне згладжування.

Розглянемо дві випадкові залежності, показані на рис.3.7. Залежності мають однакову регулярну складову $\bar{x}(t)$, а випадкові складові залежностей відрізняються: залежність рис.3.7,б крім дрібних випадкових флуктуацій, має рідкісні викиди досить великої амплітуди.



а б

Рис.3.8

Метою згладжування є отримання плавної залежності $\bar{x}(t)$. Метод ковзного середнього передбачає вибір *вікна усереднення* τ, і кожного t_i розраховується середнє значення цьому інтервалі:

$$\bar{x}_1(t_i) = \frac{1}{\tau}\int_{t_i-\frac{\tau}{2}}^{t_i+\frac{\tau}{2}} x(t)dt \quad . \qquad (3.3.1)$$

Цей метод дозволяє згладити випадкову складову залежності, тобто позбутися високочастотних флуктуацій. При цьому фільтрація високих частот залежить від довжини інтервалу усереднення τ.

Якщо застосувати метод ковзного середнього залежно від рис.3.7,б, який має значні, але рідкісні викиди, то отримана згладжена залежність $\bar{x}_2(t)$ різко відрізняється від $\bar{x}_1(t)$. Викиди за рахунок їх високої амплітуди сильно впливають на середнє значення, як тільки потрапляють в інтервал усереднення, тобто кожен викид спотворює $\bar{x}(t)$ на інтервалі довжиною τ. У випадку, коли викиди мають рідкісний та випадковий характер, для виділення регулярної складової метод ковзного середнього неприйнятний.

За наявності рідкісних викидів (рис.3.7,б) зручніше застосовувати метод медіанного згладжування, у якому на «ковзному» інтервалі τ для отримання $\bar{x}(t)$ використовується не середнє значення функції, а медіана.

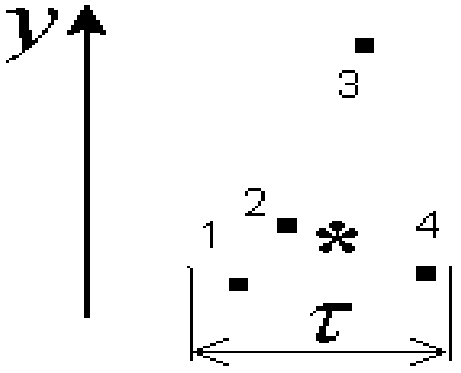


Рис.3.9

Нехай на інтервал τ потрапило чотири значення часового ряду $y_i(t_i)$ і одне із значень y_i сильно відрізняється від інших (див. рис.3.9, точка 3). Побудова медіани передбачає, що медіанний центр, позначений зірочкою, перебуватиме в області точок 1, 2 і 4, тому що вище і нижче його повинні знаходитися по дві точки. Таким чином, викид у точці 3 згладженої залежності буде усунений.

Необхідно мати на увазі, що залежності $\bar{x}_1(t)$ і $\bar{x}_2(t)$ отримані методом ковзного середнього і методом медіанного згладжування відрізняються. Їх відмінність зростає зі збільшенням частоти виникнення аномальних викидів. Якщо викиди виникають у аналізованої часової залежності досить часто, всі вони можуть розглядатися як невід'ємна характеристика флуктуаційної складової та їх усунення при згладжуванні спотворює адекватний опис випадкового ряду. Для оцінки частоти викидів можуть використовуватись різні критерії, наприклад, середня частота виникнення викидів на вибраному часовому інтервалі або відношення їх загальної кількості до довжини часового ряду. Крім того, постає питання: флуктуації якої амплітуди вважати викидами? Докладніше викиди випадкових процесів будуть розглянуті у п'ятому розділі.

Лекція № 7. Імітаційне моделювання та його роль у дослідженні систем управління.

7.1. Використання сучасних прикладних математичних пакетів математичного аналізу.

НЕПАРАМЕТРИЧНІ АЛГОРИТМИ УПРАВЛІННЯ

Тема 4 . Адаптивні алгоритми

Лекція № 8. Синтез систем автоматичного управління з структурою, що перебудовується.

1. Постановка завдання управління
2. Типова система регулювання
3. Адаптивна система автоматичного регулювання
4. Системи автоматичного регулювання із структурною адаптацією

8.1. Постановка завдання управління

Суть питання зводиться до вибору такого управління u , при якому вихідне значення y об'єкта управління збігалося б із значенням s або їх різниця лежала б в допустимих межах при зміні зовнішнього впливу f і a .

Обурення f називається координатним, а обурення a – параметричним. Під впливом зовнішніх обурень, інформації про які часто недостатньо, взаємозв'язок між входом і виходом об'єкта стає неоднозначним і невизначеним, що ускладнює вирішення завдання.

Координатне обурення є невідомою величиною з боку навантаження на об'єкт управління, що проявляється у вигляді неконтрольованих довільних змін технологічних параметрів та характеру зміни в часі може бути імпульсною та повільною мінливою. Параметричне обурення є невідомою величиною з деякої обмеженої множини, в результаті дії якої відбувається повільна зміна параметрів об'єкта керування.

Слід зазначити принципову різницю між цими двома типами збурень. Розглянемо випадок, коли об'єкт управління Wоб(p) з вхідним сигналом u і виходом y діють обидва типи впливів, що обурюють. Тоді вихідна координата об'єкта прийме вигляд

$$y = W_{об}(p, a)\, g = W_{об}(p, a)\,(f + u) = W_{про}(p, a)\, f + W_{про}(p, a)\, u.$$

Тепер наочно видно якісну відмінність впливу збурень f та a на виході об'єкта. Координатне обурення f вносить адитивний та незалежний від входу u внесок у реакцію об'єкта, рівний W про (p, a) f. Параметричне ж обурення a змінює лише вигляд або параметри W про (p, a) і не має незалежного від u та f впливу на вихід об'єкту.

Таким чином, обурення f формує «лінійний» вплив довкілля на регульовану координату, а обурення a – «нелінійний» її вплив.

8.3 Адаптивна система автоматичного регулювання

Загальноприйнятий порядок синтезу систем управління полягає в наступному:
- Задається математична модель об'єкта (на практиці це зазвичай модель, одержана на основі експериментальної перехідної характеристики об'єкта);
- Приймається критерій оптимальності системи управління;
– за моделлю об'єкта визначаються структура та чисельні значення параметрів алгоритму функціонування контролера (регулятора), що задовольняють прийнятому критерію оптимальності.

Вважається, що якщо модель досить близька до реального об'єкта, а обраний метод синтезу та розрахунки виконані бездоганно, то спроектована система запрацює без будь-якої суттєвої доведення під час пуску. Проте досвід свідчить у тому, що такий оптимістичний прогноз, зазвичай, не виправдовується. Пояснюється це двома причинами:

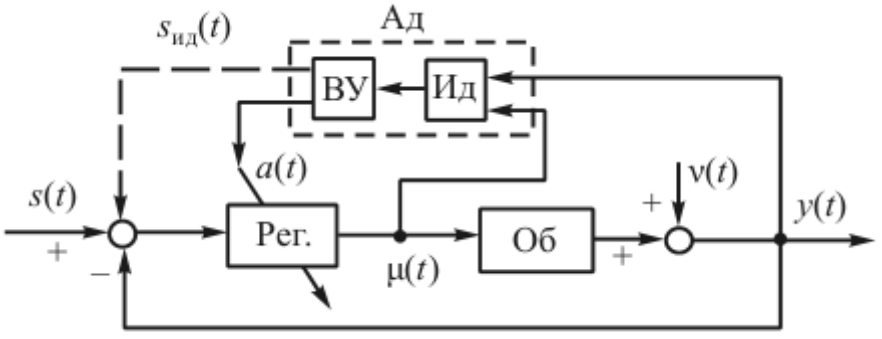
– системним характером завдання отримання математичної моделі об'єкта; це означає, що з формулювання критерію наближення останньої необхідно мати алгоритм функціонування контролера, визначення якого, власне, і потрібна ця модель [15];
- Практичною неможливістю обліку відхилення прийнятої в розрахунках динамічної моделі контролера від реальної (наявність широтно-імпульсного перетворення сигналу на виході контролера, зони нечутливості, люфтів у механічних зчленуваннях виконавчого механізму тощо).

Вихід із ситуації полягає в тому, що системи управління навіть із відносно стабільними об'єктами повинні проектуватися як адаптивні (з автоматизованим налаштуванням). Ефективність таких систем визначається тим, що вони оперують усією системою загалом, причому за відповідного вибору режиму ідентифікації можна здійснювати автоматичну лінеаризацію нелінійності в значному для кожної конкретної системи діапазон частот і відхилень сигналів [1].

У функції адаптації не входить підстроювання параметрів регуляторів до властивостей об'єкта, що швидко змінюються, викликаним контрольованими обуреннями, насамперед – змінами навантаження об'єкта. У цьому випадку має застосовуватися звичайна корекція налаштування регуляторів за заздалегідь заданими законами, що реалізуються у відповідних коригувальних блоках. Однак у функції адаптації входить налаштування цих коригувальних блоків. Взагалі можливості теорії автоматичного управління (як і будь-якої іншої теорії) обмежені деякими межами. За надто швидких змін властивостей об'єкта та пов'язаних з цією появою нелінійних ефектів Важлива можливість адаптації систем управління досить складними в динамічному відношенні об'єктами виявляється дуже проблематичною.

Структура адаптивної системи управління може бути представлена такою, як показано на рис. 5.6. До контуру регулювання, що включає об'єкт Про та регулятор Рег, приєднується адаптуюче пристрій Пекло, на вхід якого подаються вхідний μ (t) і вихідний у (t) сигнали об'єкта. В ідентифікуючому пристрої Ід за отриманими сигналів оцінюється модель об'єкта, а обчислювальному пристрої ВУ визначаються оптимальні параметри налаштування регулятора, які потім встановлюються за допомогою адаптуючого впливу а(t). Причому для реалізації адаптуючого пристрою пекло використовується один з відомих в даний час методів адаптації. У роботі застосовується метод адаптації, який використовує сигнальне гармонійний ідентифікуючий вплив (метод Циглера-Нікольса) [18]. Перевагою такого методу є можливість обґрунтованого застосування методів математичної статистики у процесі проведення ітераційної процедури руху до оптимуму.

Практична значимість цієї обставини полягає у можливості зменшення амплітуди впливів до прийнятного рівня та, незважаючи на це, отримання задовільних оцінок параметрів вихідних коливань завдяки збільшенню тривалості адаптації.



Мал. 5.6. Структура адаптивної системи

Спостереження за станом об'єкта у процесі нормального функціонування без запровадження додаткових пошукових складових не приводить до успіху. Пояснюється це тим, що оскільки об'єкт знаходиться у складі системи, то і оперувати слід із впливами, що є вхідними сигналами всієї системи; при цьому вхідний сигнал слід вибрати таким чином, щоб канал, що ідентифікується системи залежав тільки від одного невідомого оператора об'єкта.

Існуюча проблема зводиться не до того, щоб створити систему адаптації, що функціонує без викликаних ідентифікуючими впливами додаткових відхилень регульованої величини, а до тому, щоб зробити ці відхилення досить малими, прийнятними для практики Це досягається декомпозицією процедури пошуку з використанням алгоритмів налаштування нижнього рівня спеціально розроблених неекстремальних критеріїв.

Така процедура заснована на використанні активних частотних методів ідентифікації об'єктів та розрахунку оптимального налаштування регулятора.

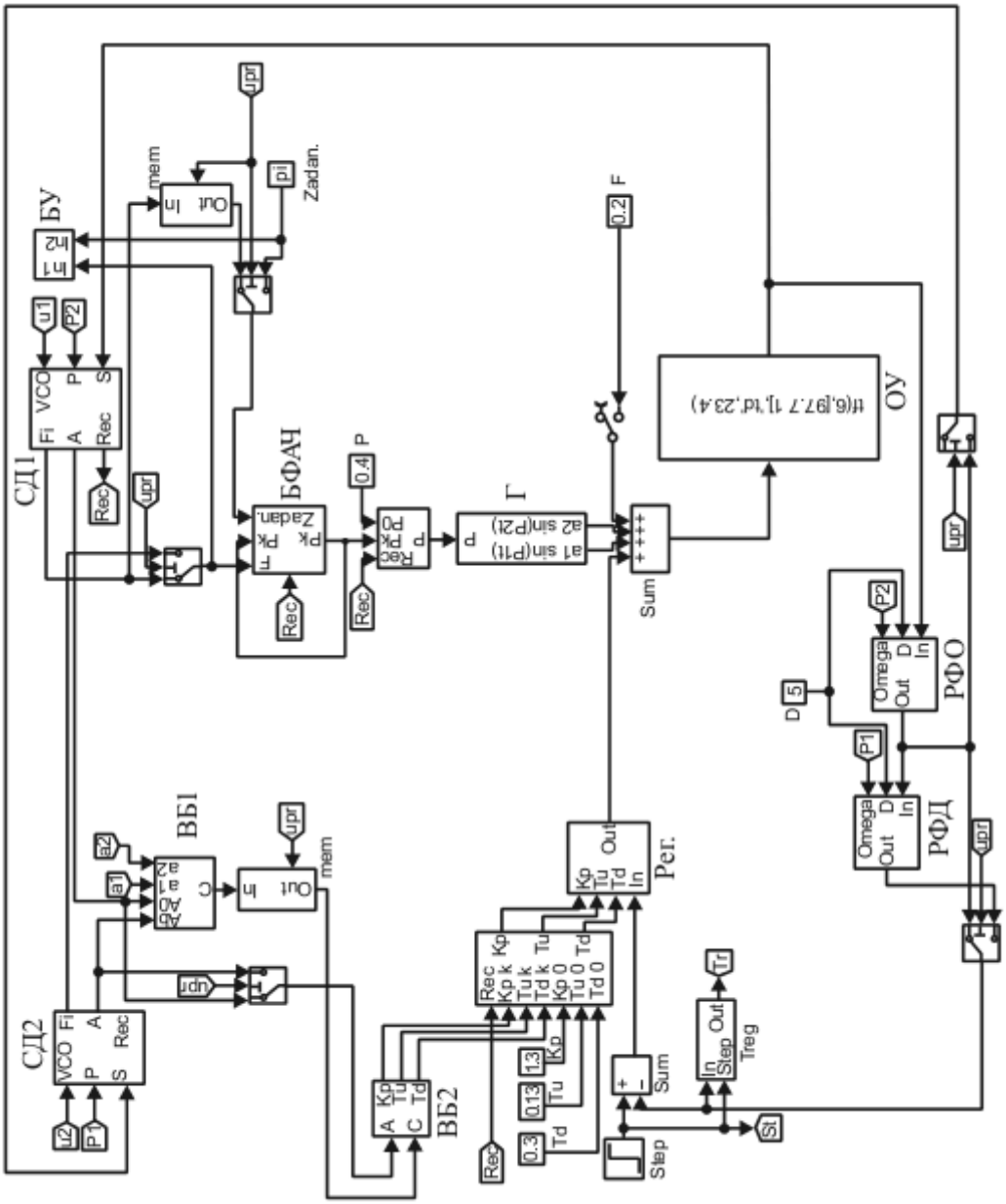
При формуванні процесу оцінки моделі об'єкта структура адаптивної системи керування, наведена на рис. 5.6, має бути доповнена ще одним ідентифікуючим впливом, який повинен надавати адаптуючий пристрій Пекло на систему з метою ідентифікації об'єкта. Величина $s_{ід}(t)$ показана пунктирною лінією в вигляді сигнального впливу, що подається на задатчик регулятора. Такий ідентифікуючий вплив не обов'язково має бути сигнальним. Воно може бути алгоритмічним, параметричним, структурним.

Найчастіше автоматичне налаштування здійснюється шляхом включення до каналу сигналу помилки двопозиційного реле з малим вихідним сигналом. Потім за параметрами автоколивань, що виникають у замкнутій системі, визначаються необхідні налаштування регулятора. Причому при такому способі самоналаштування відбувається припинення процесу регулювання об'єкта на час налаштування, спостерігається висока чутливість до шумів у каналі вимірювання, виникає небезпека зриву автоколивань при дії збурень.

Є також алгоритм налаштування регулятора у замкнутому контурі шляхом подачі на вхід системи пробного синусоїдального сигналу. Для цього алгоритму потрібен досить великий час налаштування (біля 8-10 періодів коливань на резонансній частоті замкнутої системи.

У роботі використовується метод Циглера-Нікольса з частотним поділом каналів управління та самоналаштування, що досягається включенням двох режекторних цифрових фільтрів у зворотний зв'язок контуру регулювання.

На рис. 5.7 наведено структурну схему адаптивної системи управління. Основний контур складається з регулятора Р, що настроюється, власне об'єкта управління ОУ та двох режекторних фільтрів (основного РФО та додаткового РФД). Додатковий режекторний фільтр за допомогою перемикача П1 включається лише на першому етапі чи періодично визначення необхідних методом Циглера–Никольса налаштувань. Блоки синхронного детектування ЦД 1, ЦД 2 визначають значення амплітуд Аб , А0 і фаз Фб , Ф0 пробних складових у вихідних сигналах основного режекторного фільтра у1 та об'єкта управління у . Визначення заданого фазового зсуву здійснюється за допомогою блоку фазового автопідстроювання частоти (БФАЧ).



Мал. 5.7. Структурна схема адаптивної системи

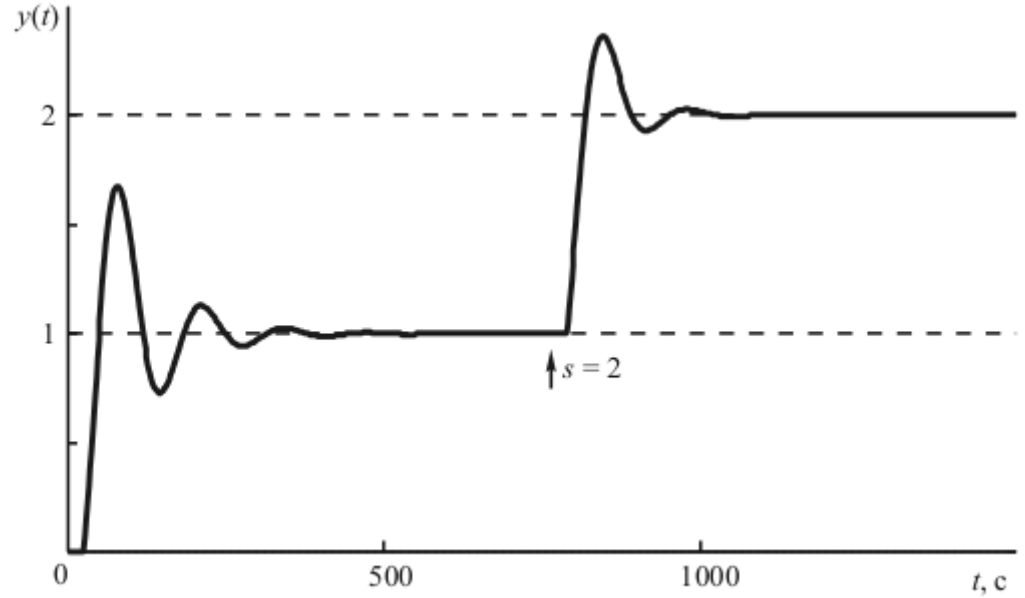
Порівнюючи значення фазового зсуву Ф 0 із заданим Ф з , блок коригує частоти пробних синусоїдальних коливань першого Р 1 і другого Р 2 генераторів. При цьому частота генератора Г 2 обрана вдвічі вище за частоту генератора Г 1 . Обчислювальні блоки ВБ 1, ВБ 2, ВБ 3 за сигналами з блоку управління БУ проводять обчислення поправочного коефіцієнта З параметрів налаштування регулятора і коефіцієнтів режекторних фільтрів. Значення коефіцієнта і фазового зсуву Ф 0 в об'єкті управління в момент розрахунку необхідних налаштувань зберігаються в пристроях ЗУ 1 , ЗУ 2 . Комутація вхідних сигналів відповідних блоків на різних етапах роботи системи відбувається за командами БО за допомогою перемикачів П3, П4, П5, П6. Останню операцію доцільно здійснювати із застосуванням багатофункціональних логічних модулів [7].

Процес адаптації складається з двох етапів: двочастотного, наприкінці якого обчислюються необхідні налаштування та коефіцієнт С; основного, в якому процес самоналаштування відбувається вже на одній підвищеній частоті пробних коливань з одним фільтром режектором РФО у контурі зворотний зв'язок. Саме визначення критичної частоти та відповідної їй амплітуди коливань для блоку ОУ+РФО дозволяє працювати з потрібними параметрами налаштувань другого етапу.

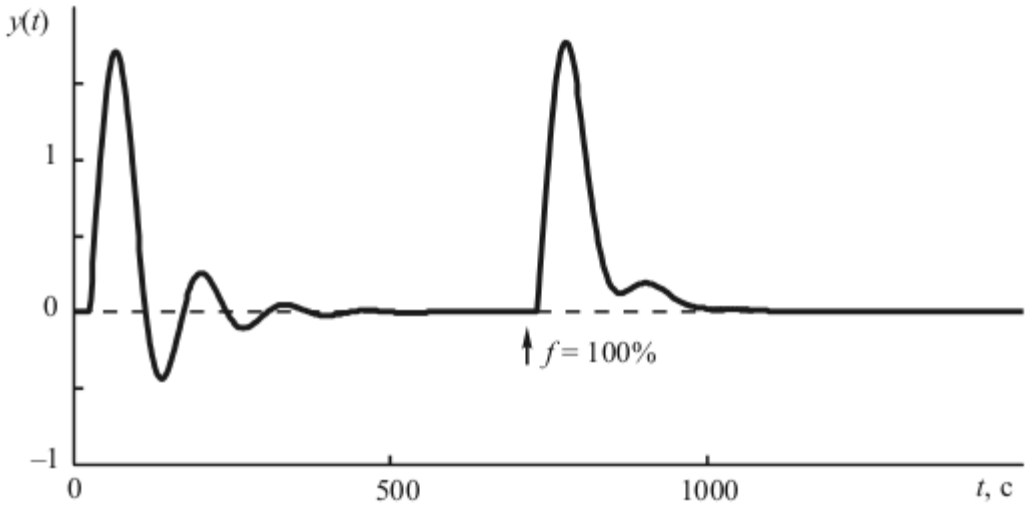
Здійснюється це за допомогою поправного коефіцієнта С, що встановлює зв'язок між необхідними за методом Циглера-Нікольса налаштуваннями на першому та другому етапах роботи системи [12].

На рис. 5.8, 5.9 представлені перехідні процеси аналізованої системи каналом задає і обурює впливу відповідно. Після процесу адаптації значно зменшилися коливання, максимальна динамічна помилка та час регулювання.

На рис. 5.10 представлені порівняльні динамічні характеристики створеного адаптивного регулятора з налаштуванням за методом Циглера-Нікольса (крива 1), традиційний ПІД-регулятор, що налаштовується вручну на кожному етапі роботи адаптивного регулятора (крива 2), традиційний ПІД-регулятор з фіксованим налаштуванням методу РАФЧХ на вихідне значення постійного часу об'єкта управління (крива 3).

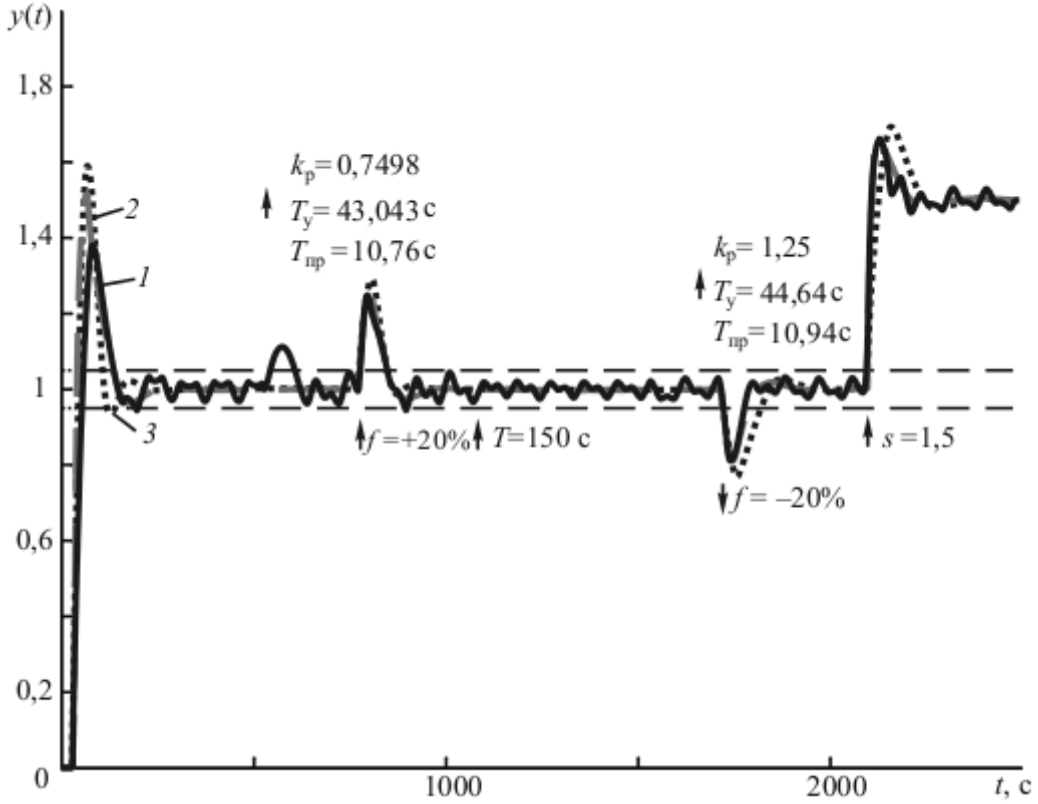


Мал. 5.8. Перехідний процес в адаптивній системі автоматичного регулювання по каналу впливу



Мал. 5.9. Перехідний процес в адаптивній системі автоматичного регулювання каналом обурення, що йде з боку регулюючого органу

Для функціонування системи достатньо в розроблений адаптивний ПІД-регулятор ввести налаштування, що забезпечують стійкість замкнутої системи, тому скористаємося знайденими в попередньому параграфі параметрами налаштування ПІ-регулятора та приймаємо коефіцієнт передачі kp = 0, 47 постійну часу подвоєння T y = 52, 6 з і постійну часу попередження T пр = 0 .



Мал. 5.10. Динамічні характеристики: 1 – адаптивний ПІД-регулятор; 2 – класичний ПІД-регулятор, що налаштовується вручну на кожному етапі роботи адаптивного регулятора; 3 – ПІД-регулятор, налаштований на початкове значення постійного часу об'єкта

Тоді передатна функція регулятора на початковий момент регулювання набуде вигляду

$$W_{\text{рег}}(p) = k_{\text{р}}\left(1 + \frac{1}{T_{\text{у}}p} + T_{\text{пр}}p\right) = 0,47 + \frac{0,47}{52,6p}.$$

Параметричний синтез ПІД-регулятора за методом РАФЧХ для об'єкта, що описується передавальною функцією (5.4) при заданих ψ = 0,95 і другий інтегральний критерій якості, дав такі результати: kp = 0,727; T у = 51,929 с; T пр = 5,77 с. Синтез методом Циглера-Нікольса дозволив визначити значення: kp = 0,72; T у = 42,35 с; T пр = 5,57 с.

Інтегральні критерії якості за обома методами приймають майже однакові значення:

$$I_2^{\text{РАФЧХ}} = \int_0^\infty |\varepsilon(t)|\,dt = 59,59; \quad I_2^{\text{Ц-Н}} = \int_0^\infty |\varepsilon(t)|\,dt = 62,43; \qquad (5.7)$$

$$I_3^{\text{РАФЧХ}} = \int_0^\infty \varepsilon(t)^2\,dt = 41; \quad I_3^{\text{Ц-Н}} = \int_0^\infty \varepsilon(t)^2\,dt = 39,71. \qquad (5.8)$$

Встановлено, що перехідний процес аналізованої системи автоматичного регулювання з регулятором, розрахованим за методом Циглера-Нікольса, має близькі до мінімуму інтегральні оцінками якості (5.7), (5.8), однак у тимчасовій області перехідний процес має менше перерегулювання та меншу максимальну динамічну помилку порівняно з процесами, налаштованими за аналітичним методом РАФЧХ, що, безперечно, краще для технологічних процесів.

Адаптивний ПІД-регулятор з автоматичним підстроюванням коефіцієнтів має близькі динамічні характеристики з ПІД-регулятором, налаштованим вручну, на кожному етапі адаптації, методу Циглера-Нікольса, і помітний виграш як регулювання порівняно з ПІД-регулятором, що має фіксовану налаштування при параметричному обуренні.

Лекція № 9. Системи автоматичного регулювання з структурою, що перебудовується

1. Формування логічного закону управління
2. Приклад синтезу системи без запізнення у контурі управління
3. Приклад синтезу системи із запізненням у контурі управління
4. Інтегральний регулятор з структурою, що перебудовується
5. Інтегральний дискретний регулятор з структурою, що перебудовується

9.1. Формування логічного закону управління

У загальному випадку закон управління для систем з перебудовується структурою (СПС) виглядає так:

$$u = \Psi \cdot \varepsilon \text{ , (5.12)}$$

де - сигнал помилки регулювання; Ψ – логічний закон, приймає певне значення залежно від типу об'єкта управління та його поточного стану.

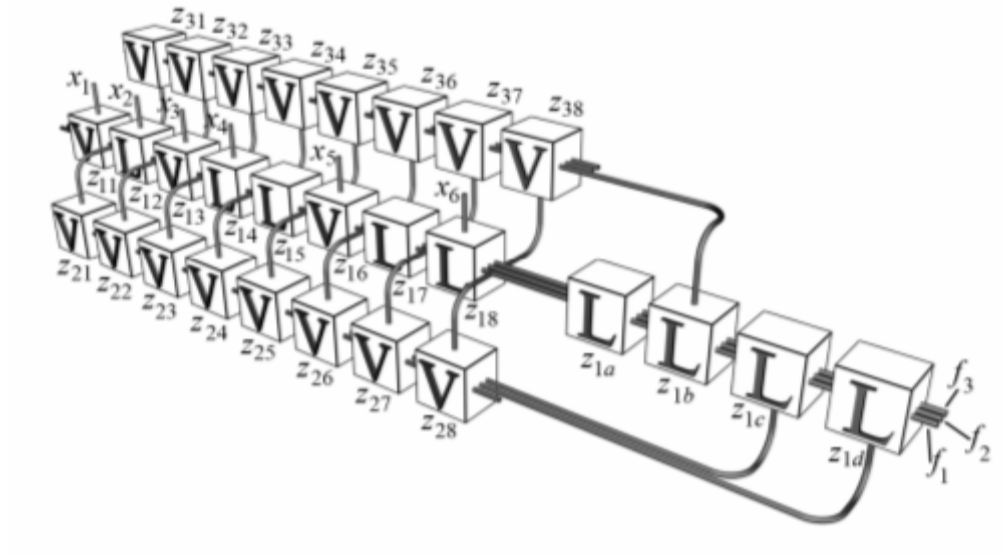
Пристрій управління, що реалізує (5.12), можна уявити вигляді заданого класу динамічних коригувальних ланок A_i (i = 1, 2, ..., n), логічного пристрою M та допоміжних блоків формування вхідних аргументів X та налаштування структури Z . Останні блоки служать для формування логічних змінних, що залежать від стану системи у фазовому просторі та налаштування логічного пристрою M на потрібний логічний закон управління (рис. 5.15).

У системах із змінною структурою необхідні динамічні властивості замкнутої системи забезпечуються належним вибором поверхні перемикання, вид якої визначається при синтезі [14]. У нашому випадку поверхня перемикання реалізується квазіізотропною середовищем, побудованим на основі МЛМ двох типів (див. 2.4.1 та 2.4.2 – L - і V-структури) [19, 20].

Мал. 5.15. Структурна схема пристрою керування

Квазіізотропне середовище є три межі просторової шестиспрямованої структури (рис. 5.16) [11]. Центральна (перша) грань є лінійною структурою з L - і V - осередків з

режимними бічними входами (yi , i = 1, 2,3) і поданими аргументами X j (j = 1, 2, ..., n, n = 6 - число аргументів) системи, що описують логічний закон управління, на функціональні входи x осередків. Друга і третя грані є також лінійними структурами з V - осередків з режимними входами. На функціональні входи x подані проміжні значення після сусідніх осередків центральної грані першого та другого каналів відповідно для другої та третьої граней.



Мал. 5.16. Квазіізотропне середовище

p align="justify"> Для різних типів технологічних об'єктів необхідно використовувати різні закони управління. Так, наприклад, для об'єктів з постійними параметрами, у разі використання системи із змінною структурою, закон управління описується виразом

⎧ α при ε s 1 > 0,
Ψ=⎨ (5.13)
⎩ β при ε s 1 < 0,

а для об'єкта з постійними параметрами, що володіє транспортним запізненням, має вигляд [5]:
⎧ α при ε s $_1$ > 0,
Ψ = ⎨ β при s $_2$ s $_1$ < 0, (5.14)
⎩ 0 при ε s $_2$ < 0.

Тут α та β – коефіцієнти передачі першої та другої лінійних структур відповідно (A1 = α, A1 = β, див. рис. 5.15); si = sgn(ε' + cіε), i = 1, 2 , – інформація про знак лінійної комбінації помилки та її похідну, що характеризує положення системи у фазовому просторі щодо прямих перемикань; сі – коефіцієнт нахилу прямих. Як видно, для функціонування системи не потрібно точного значення похідної від сигналу помилки, а достатньо лише інформації про знак її лінійної комбінації з величиною помилки, яку можна отримати порівняно простими технічними засобами, наприклад, описаними у [5, 6].

У булевій формі стратегію вибору динамічного закону управління для об'єктів без транспортного запізнення можна уявити наступним чином:

$$\begin{cases} f1 = X1, \\ f2 = X2, \\ f = 0, \\ 3 \end{cases}$$

(5.15)

а для об'єктів із запізненням – у вигляді

$$\begin{cases} f1 = (X1 \vee X4)\, X2\, X3, \\ f2 = X1\, X3\, X5, \\ f = XX(X \vee X), \\ 1\,6 \\ 2 \\ 5 \\ 3 \end{cases}$$

(5.16)

де X j (j = 1, 2, ...,6) - логічні аргументи, що відображають поведінку аналізованої системи у фазовому просторі, що виробляються блоком формування вхідних аргументів (рис. 5.15).

Система булевих формул (5.15) або (5.16) формується логічним пристроєм М (рис. 5.15), для цього кожен вхід Z ij -ї осередку квазіізотропного середовища (див. рис. 5.16) необхідно налаштувати згідно табл. 5.4 чи 5.5 відповідно.

Таблиця 5.4

Настроювальні коди квазіізотропного середовища для реалізації алгоритму керування об'єктом без транспортного запізнення

i	*j*											
	1	2	3	4	5	6	7	8	*a*	*b*	*c*	*d*
1	0011	001	0110	010	010	0110	010	010	010	010	010	010
2	0101	1111	0110	0110	0110	0110	0110	0110	—	—	—	—
3	0101	1111	0110	0110	0110	0110	0110	0110	—	—	—	—

Таблица 5.5

Настроечные коды квазиизотропной среды для реализации алгоритма управления объектом с транспортным запаздыванием

i	*j*											
	1	2	3	4	5	6	7	8	*a*	*b*	*c*	*d*
1	1001	111	0101	010	100	1001	000	110	000	010	000	010
2	0101	0110	0110	0110	0110	0100	0110	0101	—	—	—	—
3	0110	1001	0110	0110	0110	0110	0110	0110	—	—	—	—

Отже, два закони управління можна реалізувати на одній обчислювальному середовищі, використовуючи як її елементарні вузлові елементів МЛМ. Кількість реалізованих алгоритмів представленої квазіізотропного середовища не обмежується тільки двома, які ми використовуємо. Воно залежить від числа фіксованих структур, що реалізуються логічним пристроєм M , який, у свою чергу, залежить від узагальненого настроювального коду логічного пристрою та складає 2 r структур, де r - розрядність цього коду. У нашому випадку неважко підрахувати (за табл. 5.4 чи 5.5), що r = 103 . Крім того, при різних фіксованих структурах, а значить, і різних настроювальних кодах пристрою M останнє може реалізовувати одні й самі алгоритми, тому їх число $m \leq 2^r$.

Наведемо методику синтезу логічного пристрою M на вже розроблених або новостворених МЛМ [17].

1. Записуються алгоритми управління, що задовольняють вимогам до системи.
2. Вибрані алгоритми подаються у булевій формі.
3. Вибирається один алгоритм, отриманий у п. 2, що покриває Найбільший клас булевих формул.
4. Вибирається безліч МЛМ, на яких реалізовуватиметься алгоритм (це безліч може бути і поодиноким – випадок ізотропного середовища), і вид комбінаційних зв'язків між сусідніми МЛМ.
5. З набору фіксованих структур, що реалізуються обраними МЛМ, виділяються ті, що забезпечують реалізацію відповідної булевої формули у логічному законі управління на підставі п. 3.
6. На основі виділеного підмножини автоматних відображень вибраних МЛМ синтезується ізотропне або квазіізотропне середовище, що забезпечує реалізацію логічного закону управління (тобто. виконується процес послідовного вкладення кожного аргументу логічного закону в квазіізотропне середовище). Результатом є синтезоване квазіізотропне середовище та коди налаштування кожного МЛМ середовища на реалізацію логічного закону (див. п. 3; рис. 5.16 та табл. 5.5).
7. Налаштовуємо квазіізотропне середовище на реалізацію наступного алгоритму з п. 2, використовуючи її як «каркас» і змінюючи автоматні відображення кожної МЛМ, шляхом подачі на нього певного настроювального коду таким чином, що аргументи алгоритму, що реалізується, послідовно вкладаються в кваіізотропне середовище. Результатом є код налаштування квазіізотропного середовища на відповідний алгоритм (див. табл. 5.4).
8. Повторюємо п. 7 $k - 2$ рази, де k – число алгоритмів із п. 2.

Таким чином, кількість входів X логічного пристрою M залежить від складності реалізованого алгоритму, що задовольняє вимогам системи, тобто. від кількості аргументів, що використовуються при його описі у булевій формі. Кількість же настроювальних входів Z залежить від обраного варіанта комбінаційних зв'язків між МЛМ, а значить, від форми квазіізотропного середовища, кількості МЛМ у середовищі та кількості настроювальних входів кожного МЛМ.

Тема 5. Нечіткі алгоритми .

Лекція №10. Управління на базі нечіткої логіки.

1. Загальні особливості управління на базі теорії нечітких множин
2. Принцип роботи нечіткого регулятора. Алгоритми нечіткого виведення
3. Процес прийняття рішень у системі з одним вихідним та n вхідними параметрами

Проектування систем керування Fuzzy Logic Toolbox

С.Д.Штовба "Введення в теорію нечітких множин та нечітку логіку"

1.1 Основні терміни та визначення

Поняття нечіткої множини - це спроба математичної формалізації нечіткої інформації для побудови математичних моделей. В основі цього поняття лежить уявлення про те, що складові дана безліч елементи, що володіють загальним властивістю, можуть мати цю властивість у різному ступені і, отже, належати до даної множини з різним ступенем. При такому підході висловлювання типу “якийсь елемент належить даній множині” втрачають сенс, оскільки необхідно вказати “наскільки сильно” або з яким ступенем конкретний елемент задовольняє властивостями цієї множини.

Визначення 1. *Нечіткою множиною (Fuzzy set)* $\tilde{A}$ на універсальній множині U називається сукупність пар ($\mu_A(u), u$), де $\mu_A(u)$ - ступінь приналежності елемента $u \in U$ до нечіткої множини $\tilde{A}$. Ступінь приналежності – це число з діапазону [0, 1]. Чим вище ступінь належності, тим більшою мірою елемент універсальної множини відповідає властивостям нечіткої множини.

Визначення 2. *Функцією приладдя (membership function)* називається функція, яка дозволяє обчислити ступінь належності довільного елемента універсальної множини до нечіткої множини.

Якщо універсальна множина складається з кінцевої кількості елементів $U = \{u_1, u_2, \ldots, u_k\}$, тоді нечітка множина $\tilde{A}$ записується у вигляді $\tilde{A} = \sum_{i=1}^{k} \mu_A(u_i)/u_i$. У разі безперервної множини U використовують таке позначення $\tilde{A} = \int_U \mu_A(u)/u$.

Примітка: знаки Σ й $\int$ у формулах означають сукупність пар $\mu_A(u)$ і u.

Приклад 1. Уявити як нечіткого безлічі поняття “чоловік середнього зростання”.

Рішення: $\tilde{A}$ = 0/155+0.1/160+0.3/165+0.8/170+1/175+1/180+0.5/185+0/180.

Визначення 3. *Лінгвістичною змінною* (linguistic variable) називається змінна, значеннями якої можуть бути слова або словосполучення деякої природної чи штучної мови.

Визначення 4. *Терм-множиною (term set)* називається безліч всіх можливих значень лінгвістичної змінної.

Визначення 5. *Термом (term)* називається будь-який елемент терм-множини. Теоретично нечітких множин терм формалізується нечітким безліччю з допомогою функції власності.

Приклад 2. Розглянемо змінну " *швидкість автомобіля* ", яка оцінюється за шкалою " *низька* ", " *середня* ", " *висока* " та " *дуже висока* ".

У цьому прикладі лінгвістичної змінної є “ *швидкість автомобіля* ”, термами - лінгвістичні оцінки “ *низька* ”, “ *середня* ”, “ *висока* ” та “ *дуже висока* ”, які й становлять терм-множина.

Визначення 6. *Дефаззифікація (defuzzification)* називається процедура перетворення нечіткої множини в чітке число.

Теоретично нечітких множин процедура дефазифікації аналогічна знаходження характеристик становища (математичного очікування, моди, медіани) випадкових величин теоретично ймовірності. Найпростішим способом виконання процедури дефазифікації є вибір чіткого числа, що відповідає максимуму функції належності. Однак придатність цього обмежується лише одноекстремальними функціями власності. Для багатоекстремальних функцій приналежності Fuzzy Logic Toolbox запрограмовані такі методи дефазифікації:

Centroid – центр тяжкості;

Bisector – медіана;

LOM (Largest Of Maximums) – найбільший з максимумів;

SOM (Smallest Of Maximums) – найменший з максимумів;

Mom (Mean Of Maximums) – центр максимумів .

Визначення 7. *Дефаззифікація* нечіткої множини $\tilde{A} = \int_{[\underline{u},\bar{u}]} \mu_A(u)/u$ *за методом центру тяжкості* здійснюється за формулою $a = \frac{\int_{\underline{u}}^{\bar{u}} u \cdot \mu_A(u)\,du}{\int_{\underline{u}}^{\bar{u}} \mu_A(u)\,du}$.

Фізичним аналогом цієї формули є знаходження центру тяжкості плоскої фігури, обмеженої осями координат та графіком функції належності нечіткої множини. У разі

дискретної універсальної множини дефаззифікація нечіткої множини $\tilde{A} = \sum_{i=1}^{k} \mu_A(u_i)/u_i$ за методом центру тяжіння здійснюється за формулою $a = \frac{\sum_{i=1}^{k} u_i \cdot \mu_A(u_i)}{\sum_{i=1}^{k} \mu_A(u_i)}$.

Визначення 8. *Дефаззифікація* нечіткої множини $\tilde{A} = \int_{[\underline{u},\bar{u}]} \mu_A(u)/u$ *за методом медіани* полягає у знаходженні такого числа a, що $\int_{\underline{u}}^{a} \mu_A(u)\,du = \int_{a}^{\bar{u}} \mu_A(u)\,du$.

Геометричною інтерпретацією методу медіани є знаходження такої точки на осі абцис, що перпендикуляр, відновлений у цій точці, ділить площу під кривою функції належності на дві рівні частини. У разі дискретної універсальної множини дефаззифікація нечіткої множини $\tilde{A} = \sum_{i=1}^{k} \mu_A(u_i)/u_i$ за методом медіани здійснюється за формулою $a = \min_{\forall j} \left(u_j \right)$ $\sum_{i=1}^{j} \mu_A(u_i) \geq \frac{1}{2} \sum_{i=1}^{k} \mu_A(u_i)$.

Визначення 9. *Дефаззифікація* нечіткої множини $\tilde{A} = \int_{[\underline{u},\bar{u}]} \mu_A(u)/u$ *за методом центру максимумів* здійснюється за формулою:

$$a = \frac{\int_G u\,du}{\int_G du},$$

де G - безліч всіх елементів з інтервалу $[\underline{u}, \bar{u}]$, що мають максимальний ступінь належності нечіткої множини $\tilde{A}$.

У методі центру максимумів знаходиться середнє арифметичне елементів універсальної множини, що мають максимальні ступені приладдя. Якщо безліч таких елементів звичайно, то формула визначення 9 спрощується до наступного виду:

$$a = \frac{\sum_{u_j \in G} u_j}{|G|},$$

де $|G|$ - потужність множини G.

У дискретному випадку *дефаззифікація за методами* найбільшого з максимумів і найменшого з максимумів здійснюється за формулами $a = \max(G)$ і $a = \min(G)$ відповідно. З останніх трьох формули видно, що й функція власності має лише один максимум, його координата і є чітким аналогом нечіткого множини.

Приклад 3. Провести дефаззифікацію нечіткої множини “чоловік *середнього зростання* ” з прикладу 1 методом центру тяжкості.

Рішення: Застосовуючи формулу визначення 7, отримуємо:

$$a = \frac{0 \cdot 155 + 0.1 \cdot 160 + 0.3 \cdot 165 + 0.8 \cdot 170 + 1 \cdot 175 + 1 \cdot 180 + 0.5 \cdot 185 + 0 \cdot 190}{0 + 0.1 + 0.3 + 0.8 + 1 + 1 + 0.5 + 0} = 175.4$$

Визначення 10. *Нечіткою базою знань (fuzzy knowledge base)* про вплив чинників

значення параметра у називається сукупність логічних висловлювань типу:

ЯКЩО $\left(x_1 = a_1^{j1}\right) \text{И} \left(x_2 = a_2^{j1}\right) \text{И} \ldots \text{И} \left(x_n = a_n^{j1}\right)$

АБО $\left(x_1 = a_1^{j2}\right) \text{И} \left(x_2 = a_2^{j2}\right) \text{И} \ldots \text{И} \left(x_n = a_n^{j2}\right) \ldots$

АБО $\left(x_1 = a_1^{jk_j}\right) \text{И} \left(x_2 = a_2^{jk_j}\right) \text{И} \ldots \text{И} \left(x_n = a_n^{jk_j}\right)$,

ТО $y = d_j$, для всіх $j = \overline{1,m}$,

де a_i^{jp} - нечіткий терм, яким оцінюється змінна x_i у рядку з номером jp ($p = \overline{1,k_j}$);

k_j - кількість рядків-кон'юнкцій, в яких вихід у оцінюється нечітким термом d_j; $j = \overline{1,m}$

m - кількість термів, які використовуються для лінгвістичної оцінки вихідного параметра у.

За допомогою операцій $\cup$ (АБО) та $\cap$ (І) нечітку базу знань з визначення 10 перепишемо в більш компактному вигляді:

$$\bigcup_{p=1}^{k_j} \bigcap_{i=1}^{n} (x_i = a_i^{jp}) \longrightarrow y = d_j, \quad j = \overline{1,m}. \qquad (1)$$

Визначення 11. *Нечітким логічним висновком (fuzzy logic inference)* називається апроксимація залежності $y=f(x_1,x_2,...,x_n)$ за допомогою нечіткої бази знань та операцій над нечіткими множинами.

Нехай $\mu^{jp}(x_i)$ - $\mu^{d_j}(y)$ функція приналежності входу $\bowtie i$ нечіткому терму a_i^{jp} ,,,,, $i=\overline{1,n}$ т $j=\overline{1,m}$. Е $a_i^{jp}=\int_{\underline{x}_i}^{\overline{x}_i}\mu^{jp}(x_i)/x_i$.; $p=\overline{1,k_j}$ - функція приналежності виходу у нечіткого терму d_j ,, $j=\overline{1,m}$ тобто $d_j=\int_{\underline{y}}^{\overline{y}}\mu^{d_j}(y)/y$. Тоді ступінь належності конкретного вхідного вектора $X^*=\{x_1^*,x_2^*,...x_n^*\}$ нечітким терм d_j із бази знань (1) визначається наступною системою нечітких логічних рівнянь:

$\mu^{q_j}(X^*)=\bigvee_{p=\overline{1,k_j}}\bigwedge_{i=\overline{1,n}}\left[\mu^{jp}(x_i^*)\right], j=\overline{1,m}$, (2)

де $\vee(\wedge)$ - Операція максимуму (мінімуму).

Нечітка множина $\tilde{y}$, що відповідає вхідному вектору X^*, визначається наступним чином:

$$\tilde{y}=\bigcup_{j=\overline{1,m}}\int_{\underline{y}}^{\overline{y}}\min\left(\mu^{d_j}(X^*),\mu^{d_j}(y)\right)/y \quad , (3)$$

де $\bigcup$ - Операція об'єднання нечітких множин.

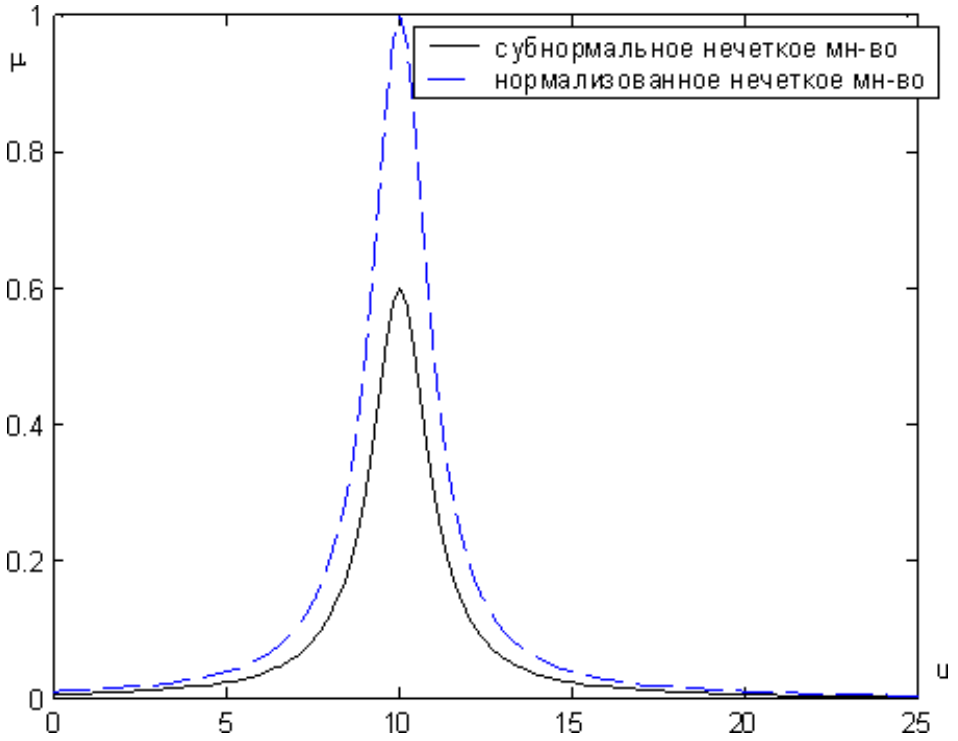
Чітке значення виходу у відповідне вхідному вектору X^* визначається в результаті дефазифікації нечіткого $\tilde{y}$.

1.2. Властивості нечітких множин

Визначення 12. *Висотою* нечіткої множини $\tilde{A}$ називається верхня межа його функції приналежності: $hgt(\tilde{A})=\sup_{u\in U}\mu_{\tilde{A}}(u)$. Для дискретної універсальної множини U супремум стає максимумом, а значить висотою нечіткої множини буде максимум ступенів приналежності його елементів

Визначення 13. Нечітка множина $\tilde{A}$ називається *нормальною,* якщо її висота дорівнює одиниці. Нечіткі множини не є нормальними називаються *субнормальними* .
Нормалізація -перетворення субнормального нечіткого множини $\tilde{A}'$ на нормальне $\tilde{A}$

визначається так: $\tilde{A} = \text{norm}(\tilde{A}') \Leftrightarrow \mu_A(u) = \frac{\mu_{A'}(u)}{\text{hgt}(\tilde{A}')}, \forall u \in U$. Як приклад на рис. 1 показана нормалізація нечіткої множини $\tilde{A}'$з функцією належності $\mu_{A'}(u) = \frac{0.6}{1+(10-u)^2}$.



Малюнок 1 - Нормалізація нечіткої множини

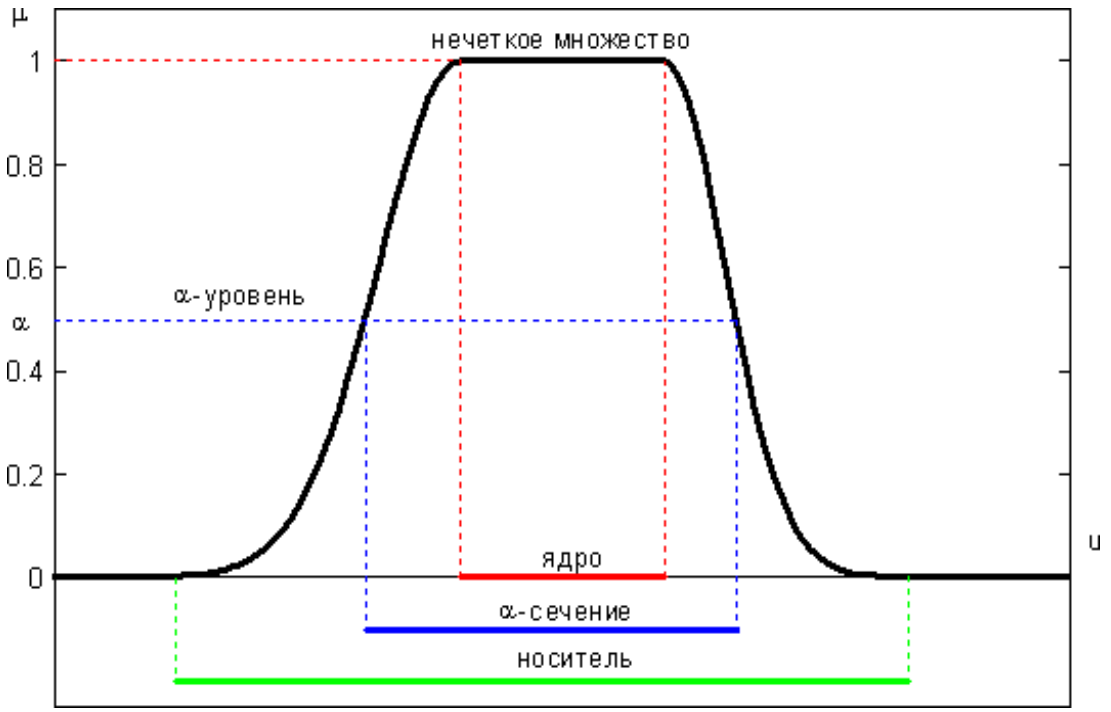
Визначення 14. *Носієм* нечіткої множини $\tilde{A}$називається чітке підмножина універсальної множини U, елементи якої мають ненульові ступеня приналежності: $\text{supp}(\tilde{A}) = \{u : \mu_A(u) > 0\}$.

Визначення 15. Нечітка множина називається *порожньою* , якщо її носій є порожньою множиною.

Визначення 16. *Ядром* нечіткої множини $\tilde{A}$називається чітке підмножина універсальної множини U, елементи якого мають рівні приналежності рівні одиниці: $\text{core}(\tilde{A}) = \{u : \mu_A(u) = 1\}$. Ядро субнормальної нечіткої множини порожнє.

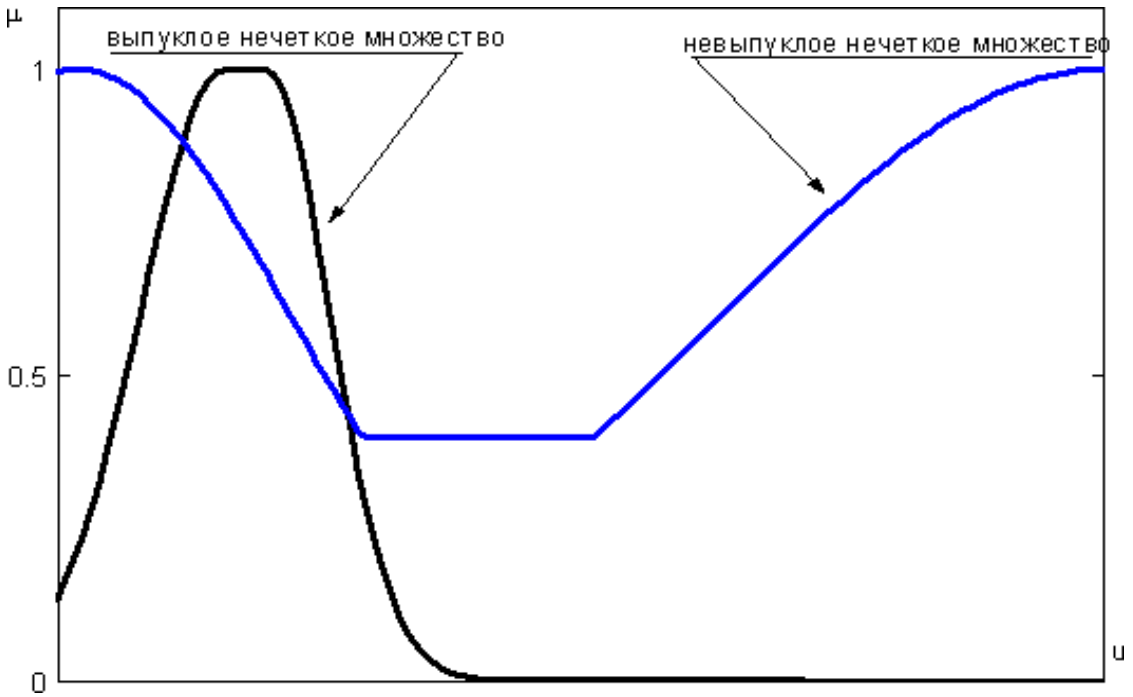
Визначення 17. α- *перетином (або безліччю* α*-рівня)* нечіткої множини $\tilde{A}$називається чітке підмножина універсальної множини U, елементи якого мають ступеня приналежності великі або рівні α: $A_\alpha = \{u : \mu_A(u) \geq \alpha\}$, $\alpha \in [0, 1]$. Значення αназивають α *-рівнем* . Носій (ядро) можна розглядати як переріз нечіткої множини на нульовому (одиничному) αрівні.

Мал. 2 ілюструє визначення носія, ядра, α- перерізу та α- рівня нечіткої множини.



Малюнок 2 - Ядро, носій та α- переріз нечіткої множини

Визначення 18. Нечітка множина $\tilde{A}$ називається *опуклим* якщо: $\mu_A(\lambda u_1 + (1-\lambda)u_2) \geq \min(\mu_A(u_1), \mu_A(u_2))$, $u_1, u_2 \in U$, $\lambda \in [0,1]$. Альтернативне визначення: нечітка множина буде *опуклим* , якщо всі його α- перерізи - опуклі множини. На рис. 3 наведені приклади опуклого та непуклого нечітких множин.



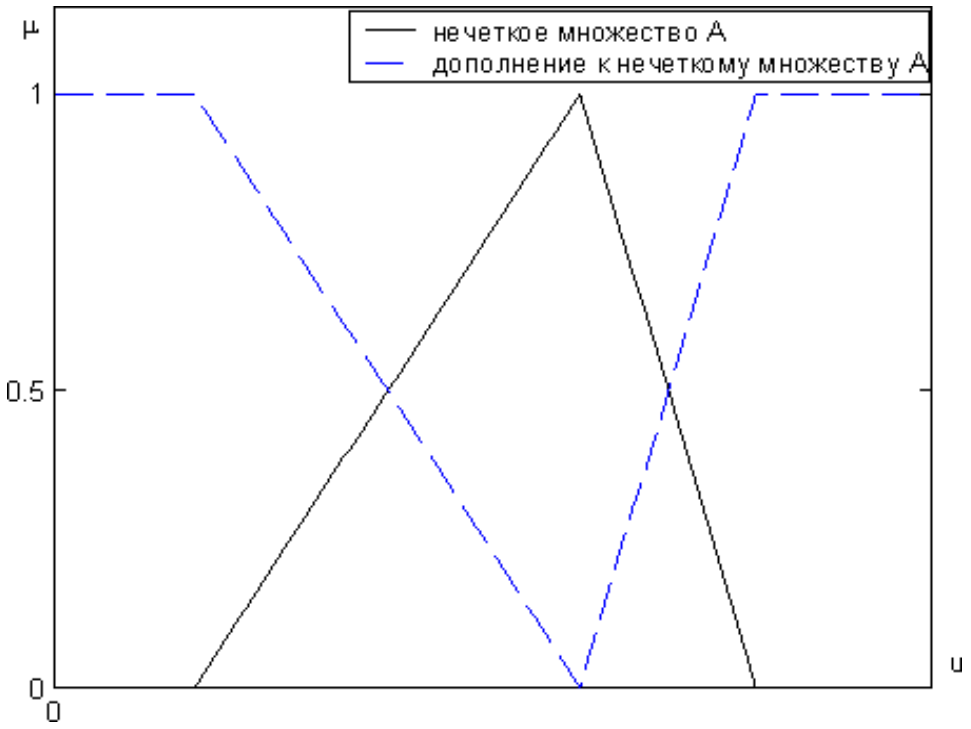
Малюнок 3 - До визначення опуклої нечіткої множини

Визначення 19. Нечіткі множини $\tilde{A}$ та $\tilde{B}$ *рівні* ($\tilde{A} = \tilde{B}$) якщо $\mu_A(u) = \mu_B(u), \forall u \in U$.

1.3. Операції над нечіткими множинами

Визначення нечітких теоретико-множинних операцій об'єднання, перетину та доповнення можуть бути узагальнені із звичайної теорії множин. На відміну від звичайних множин, теоретично нечітких множин ступінь належності не обмежена лише бінарними значеннями 0 і 1 -вона може набувати значень з інтервалу [0, 1]. Тому нечіткі теоретико-множинні операції можуть бути визначені по-різному. Зрозуміло, що виконання нечітких операцій об'єднання, перетину та доповнення над нечіткими множинами має дати такі ж результати, як і при використанні звичайних канторівських теоретико-множинних операцій. Нижче наведено визначення нечітких теоретико-множинних операцій, запропонованих Л. Заде.

Визначення 20. *Доповненням нечіткої множини* $\tilde{A}$ заданого на U називається нечітка множина $\overline{\tilde{A}}$ з функцією приналежності $\mu_{\overline{A}}(u) = 1 - \mu_A(u)$ всім $u \in U$. На рис. 4 наведено приклад виконання операції нечіткого доповнення.



Малюнок 4 - Доповнення нечіткої множини

Визначення 21. *Перетином нечітких множин* $\tilde{A}$ і $\tilde{B}$ заданих на U називається нечітка множина $\tilde{C} = \tilde{A} \cap \tilde{B}$ з функцією приналежності всім $u \in U$. Операція перебування мінімуму також позначається знаком $\wedge$, тобто. $\mu_C(u) = \mu_A(u) \wedge \mu_B(u)$.

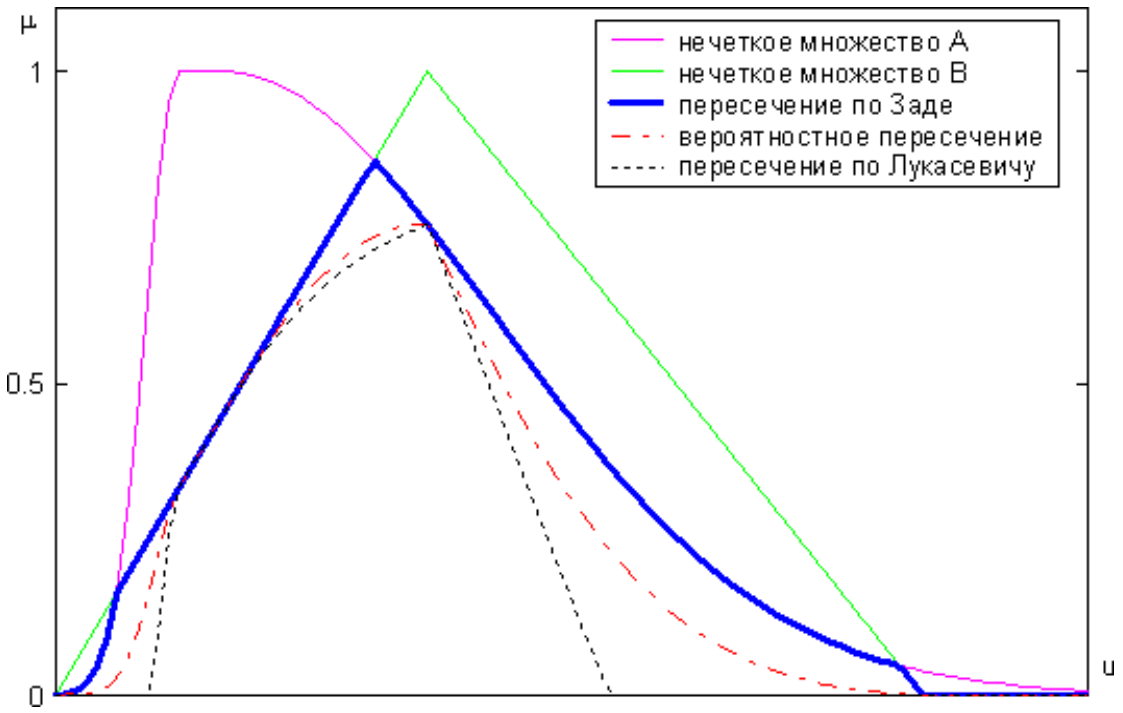
Визначення 22. *Об'єднанням нечітких множин* $\tilde{A}$ і $\tilde{B}$ заданих на U називається нечітка множина $D = \tilde{A} \cup \tilde{B}$ з функцією приналежності $\mu_D(u) = \max(\mu_A(u), \mu_B(u))$ всім $u \in U$. Операція перебування максимуму також позначається знаком $\vee$, тобто. $\mu_D(u) = \mu_A(u) \vee \mu_B(u)$.

Узагальнені визначення операцій нечіткого перетину та об'єднання - трикутної норми (t-норми) та трикутної конорми (t-конорми або s-норми) наведені нижче.

Визначення 23. *Трикутною нормою (t-нормою)* називається бінарна операція T на одиничному інтервалі $[0,1]\times[0,1]\to[0,1]$, що задовольняє наступним аксіомам для будь-яких $a,b,c\in[0,1]$:

1. $T(a,1)=a$ (гранична умова);
2. $T(a,b)\le T(a,c)$ якщо $b\le c$(монотонність);
3. $T(a,b)=T(b,a)$ (комутативність);
4. $T(a,T(b,c))=T(T(a,b),c)$ (Асоціативність).

Найчастіше використовуються такі t-норми: перетин по Заде- $T(a,b)=\min(a,b)$; ймовірнісний перетин- $T(a,b)=ab$; перетин по Лукасевичу- $T(a,b)=\max(a+b-1,0)$. Приклади виконання перетину нечітких множин з використанням цих t-норм показано на рис. 5.



Малюнок 5 - Перетин нечітких множин з використанням різних t-норм

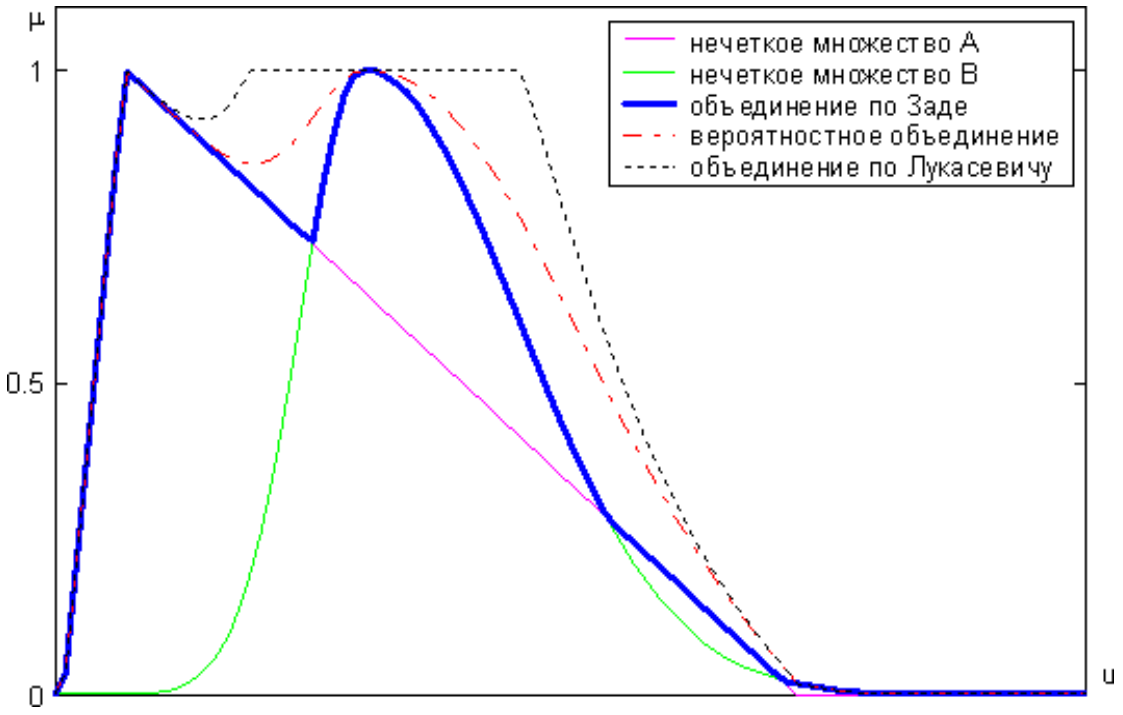
Визначення 25. *Трикутною конормою (s-нормою)* називається бінарна операція S на одиничному інтервалі $[0,1]\times[0,1]\to[0,1]$, що задовольняє наступним аксіомам для будь-яких $a,b,c\in[0,1]$:

1. $S(a,0)=a$ (гранична умова);
2. $S(a,b)\le S(a,c)$ якщо $b\le c$(монотонність);
3. $S(a,b)=S(b,a)$(комутативність);
4. $S(a,S(b,c))=S(S(a,b),c)$(Асоціативність).

Найчастіше використовуються такі s-норми: об'єднання по Заді- $S(a,b)=\max(a,b)$; імовірнісне об'єднання- $S(a,b)=a+b-ab$; об'єднання по Лукасевичу- $S(a,b)=\min(a+b,1)$.

Приклади виконання об'єднання нечітких множин із використанням цих s-норм показано на рис. 6.

Найбільш відомі трикутні норми наведено у табл. 1.



Малюнок 6 - Поєднання нечітких множин з використанням різних s-норм

Таблиця 1 – Приклади трикутних норм

T(a,b)	S(a,b)	Параметр
$\min(a,b)$	$\max(a,b)$	-
ab	$a+b-ab$	-
$\max(a+b-1,0)$		-
$\begin{cases} a, \text{если } b=1 \\ b, \text{если } a=1 \\ 0, \text{если } a,b \neq 1 \end{cases}$	$\begin{cases} a, \text{если } b=0 \\ b, \text{если } a=0 \\ 1, \text{если } a,b \neq 0 \end{cases}$	-
$\frac{ab}{\gamma+(1-\gamma)(a+b-ab)}$	$\frac{a+b-(2-\gamma)ab}{1-(1-\gamma)ab}$	$\gamma > 0$
$\frac{ab}{\max(a,b,\alpha)}$	$\frac{a+b-ab-\min(a,b,1-\alpha)}{\max(1-a,1-b,\alpha)}$	$\alpha \in [0,1]$
$\left(1+\sqrt[\lambda]{\left(\frac{1}{a}-1\right)^{\lambda}+\left(\frac{1}{b}-1\right)^{\lambda}}\right)^{-1}$	$\left(1+\sqrt[\lambda]{\left(\frac{1}{a}-1\right)^{-\lambda}+\left(\frac{1}{b}-1\right)^{-\lambda}}\right)^{-1}$	$\lambda > 0$

$1-\sqrt[\lambda]{(1-a)^{\lambda}+(1-b)^{\lambda}-(1-a)^{\lambda}(1-b)^{\lambda}}$ $\sqrt[\lambda]{a^{\lambda}+b^{\lambda}-a^{\lambda}b^{\lambda}}$ $\lambda>0$

$\max\left(1-\sqrt[\rho]{(1-a)^{\rho}+(1-b)^{\rho}},0\right)$ $\min\left(1-\sqrt[\rho]{a^{\rho}+b^{\rho}},1\right)$ $\rho\geq 1$

$\log_{\omega}\left(1+\frac{(\omega^{a}-1)(\omega^{b}-1)}{\omega-1}\right)$ $\omega>0,\omega\neq 1$

$\max\left(\frac{a+b-1+\lambda ab}{1+\lambda},0\right)$ $\min(a+b+\lambda ab,1)$ $\lambda\geq -1$

1.4. Нечітка арифметика

У цьому розділі розглядаються способи розрахунку значень чітких функцій алгебри від нечітких аргументів. Матеріал ґрунтується на поняттях нечіткого числа та принципу нечіткого узагальнення. Наприкінці розділу наводяться правила виконання арифметичних операцій над нечіткими числами.

Визначення 25. *Нечітким числом* називається опукла нормальна нечітка множина зі шматково-безперервною функцією приналежності, задана на безлічі дійсних чисел. Наприклад, нечітке число "близько 10" можна встановити наступною функцією приналежності: $\mu_{\text{"около10"}}(u)=\frac{1}{1+(u-10)^2}$.

Визначення 26. *Нечітке число* $\tilde{A}$ називається *позитивним (негативним)* якщо $\mu_A(u)=0$, $\forall u<0$($\forall u>0$).

Визначення 27. *Принцип узагальнення Заді.* Якщо $y=f(x_1,x_2,\ldots,x_n)$ -функція від n незалежних змінних і аргументи $x_1,x_2,\ldots,x_n$ задані нечіткими числами $\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_n$, відповідно, значенням функції $\tilde{y}=f(\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_n)$ називається нечітке число $\tilde{y}$ з функцією приналежності:

$$\mu_{\tilde{y}}(y^*)=\sup_{\substack{y^*=f(x_1^*,x_2^*,\ldots,x_n^*)\\ x_i^*\in \operatorname{supp}(\tilde{x}_i),\,i=\overline{1,n}}}\ \min_{i=1,n}\left(\mu_{\tilde{x}_i}\left(x_i^*\right)\right).$$

Принцип узагальнення дозволяє визначити функцію належності нечіткого числа, відповідного значення чіткої функції від нечітких аргументів.
Комп'ютерно-орієнтована реалізація принципу нечіткого узагальнення здійснюється за таким алгоритмом:

Крок 1. Зафіксувати значення $y=y^*$.

Крок 2. Знайти всі n-ки $\{x^*_{1,j}, x^*_{2,j}, \ldots, x^*_{n,j}\}$, $j = \overline{1,k}$, які відповідають умовам $y^* = f\left(x^*_{1,j}, x^*_{2,j}, \ldots, x^*_{n,j}\right)$; $x^*_{i,j} \in \operatorname{sup} p(\tilde{x}_i)$, $i = \overline{1,n}$.

Крок 3. Ступінь приналежності елемента y^* нечіткого числа $\tilde{y}$ обчислити за такою формулою: $\mu_{\tilde{y}}(y^*) = \max\limits_{j=\overline{1,k}} \min\limits_{i=\overline{1,n}} \left(\mu_{\tilde{x}_i}\left(x^*_{i,j}\right)\right)$.

Крок 4. Перевірити умову "Взято всі елементи у?". Якщо так, то перейти до кроку 5. Інакше зафіксувати нове значення y^* і перейти до кроку 2.

Крок 5. Кінець.

Наведений алгоритм грунтується на поданні нечіткого числа дискретному універсальному множині, тобто. $\tilde{x} = \sum\limits_{p=1}^{P} \mu_{\tilde{x}}(x_p)/x_p$. Зазвичай вихідні дані $\tilde{x}_i$, $i = \overline{1,n}$ задаються шматково-безперервними функціями власності: $\tilde{x}_i = \int\limits_{x_i \in R} \mu_{\tilde{x}_i}(x_i)/x_i$. Для обчислення значень функції $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ аргументи $\tilde{x}_i$ дискретизують, тобто . $i = \overline{1,n}$ представляють у вигляді $\tilde{x}_i = \sum\limits_{p=1}^{P} \mu_{\tilde{x}_i}(x_{i,p})/x_{i,p}$. Число точок P вибирають так, щоб забезпечити необхідну точність обчислень. На виході цього алгоритму виходить нечітка множина, також задана на універсальній дискретній множині. Результуючу шматково-безперервну функцію приналежності нечіткого числа $\tilde{y}$ отримують як верхню точку, що обгинає $(y^*, \mu_{\tilde{y}}(y^*))$.

Приклад 4. Нечіткі числа $\tilde{x}_1$ та $\tilde{x}_2$ задані наступними трапецієподібними функціями приналежності:

$$\mu_{\tilde{x}_1}(x) = \begin{cases} 0, & \text{если } x < 1 \text{ или } x > 4 \\ x - 1, & \text{если } x \in [1,2] \\ 1, & \text{если } x \in (2,3) \\ 4 - x, & \text{если } x \in [3,4] \end{cases} \quad \text{та} \quad \mu_{\tilde{x}_2}(x) = \begin{cases} 0, & \text{если } x < 2 \text{ или } x > 8 \\ x - 2, & \text{если } x \in [2,3] \\ 1, & \text{если } x \in (3,4) \\ 2 - 0.25x, & \text{если } x \in [4,8] \end{cases}.$$

Необхідно знайти нечітке число $\tilde{y} = \tilde{x}_1 \cdot \tilde{x}_2$ з використанням принципу узагальнення визначення 27.

Задамо нечіткі аргументи на чотирьох точках (дискретах): {1, 2, 3 4} для $\tilde{x}_1$ {2, 3, 4 8} для $\tilde{x}_2$. Тоді: $\tilde{x}_1 = \frac{0}{1} + \frac{1}{2} + \frac{1}{3} + \frac{0}{4}$ і $\tilde{x}_2 = \frac{0}{2} + \frac{1}{3} + \frac{1}{4} + \frac{0}{8}$. Процес виконання множення над нечіткими числами зведено у табл. 2. Кожен стовпець таблиці відповідає одній ітерації алгоритму нечіткого узагальнення. Результуюча нечітка множина задано першим і останнім рядками таблиці. У першому рядку записані елементи універсальної

множини, а в останньому рядку - ступеня їхньої приналежності до значення виразу $\tilde{x}_1 \cdot \tilde{x}_2$. Через війну отримуємо: $\tilde{y} = \frac{0}{2} + \frac{1}{6} + \frac{1}{9} + \frac{0}{32}$. Припустимо, що тип функція приналежності $\tilde{y}$ буде таким же, як і аргументів $\tilde{x}_1$ і $\tilde{x}_2$, тобто трапецієподібної. І тут функція власності задається выражением: $\mu_{\tilde{y}}(y) = \begin{cases} 0, & \text{если } y < 2 \text{ или } y > 32 \\ (y-2)/3, & \text{если } y \in [1,2] \\ 1, & \text{если } y \in (6,9) \\ (32-y)/23, & \text{если } y \in [3,4] \end{cases}$. На рис. 7 показані результати виконання операції $\tilde{y} = \tilde{x}_1 \cdot \tilde{x}_2$ з поданням нечітких множників на 4 дискретах. Червоними зірочками показані елементи нечіткої множини $\tilde{y}$ з табл. 2, а тонкою червоною лінією - трапецієподібна функція приналежності.

Досліджуємо, як змінитися результат нечіткого узагальнення зі збільшенням числа дискрет, у яких задаються аргументи. Нечітке число $\tilde{y}$ при заданні аргументів $\tilde{x}_1$ та $\tilde{x}_2$ на 30 дискретах наведено на рис. 7. Синіми точками показані елементи нечіткої множини $\tilde{y}$, знайдені за принципом узагальнення, а зеленою лінією - верхня -функція приналежності , що обгинає цих точок $\tilde{y}$. Функція приналежності результату має форму криволінійної трапеції, трохи вигнутої вліво.

Таблиця 2 - Наприклад 4

$y^* = x_1^* \cdot x_2^*$	**2**	**3**	**4**	**6**	**8**	**9**	**12**	**16**	**24**	**32**
x_1^*	1	1	1	2	2	3	1	2	4	3
x_2^*	2	3	4	2	3	2	8	4	2	3
$\mu_{\tilde{x}_1}(x_1^*)$	0	0	0	1	1	1	0	1	0	1
$\mu_{\tilde{x}_2}(x_2^*)$	0	1	1	0	1	0	0	1	0	1
$\min(\mu_{\tilde{x}_1}(x_1^*), \mu_{\tilde{x}_2}(x_2^*))$	0	0	1	0	1	0	0	1	0	1
$\mu_{\tilde{y}}(y^*)$	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**0**	**0**	**0**

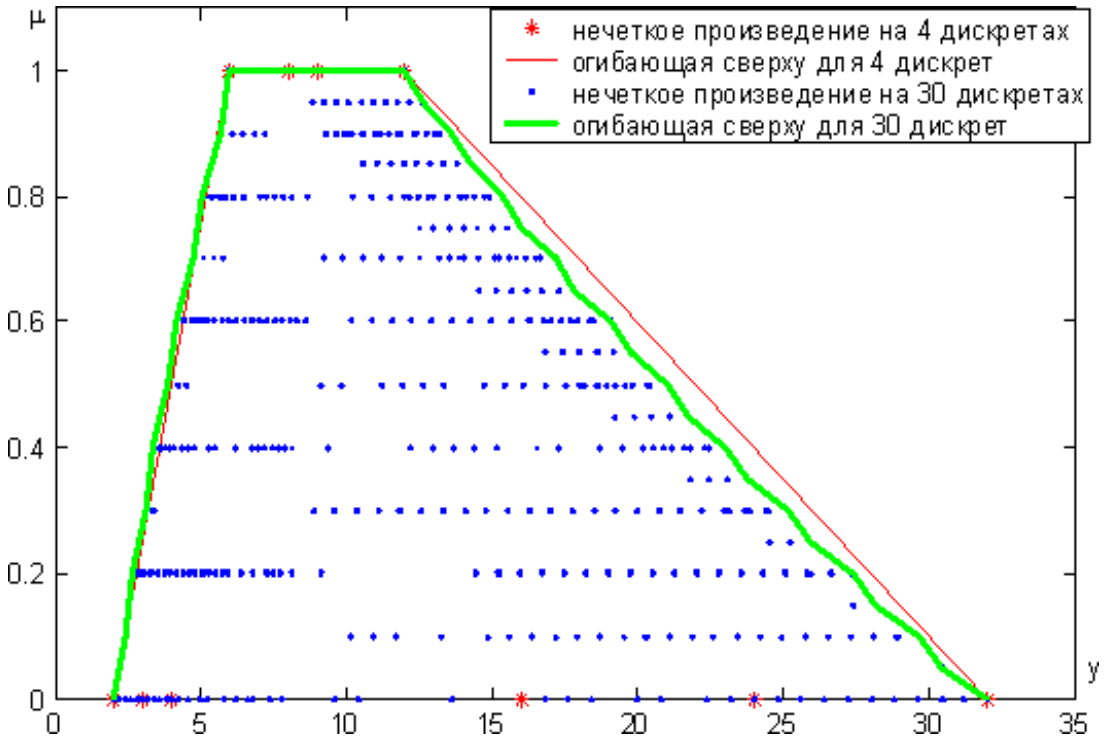


Рисунок 7 - Наприклад 4

Застосування принципу узагальнення Заде пов'язані з двома труднощами:

1. Великий обсяг обчислень - кількість елементів результуючої нечіткої множини, які необхідно обробити, так само $P_1 \cdot P_2 \cdot \ldots \cdot P_n$, де P_i -кількість точок, на яких заданий і-й нечіткий аргумент $i = \overline{1,n}$;
2. необхідність побудови верхньої огинаючої елементів результуючої нечіткої множини.

Більш практичним є застосування α-рівневого принципу узагальнення. У цьому випадку нечіткі числа подаються у вигляді розкладів по α-рівневих множин: $\tilde{x} = \bigcup_{\alpha \in [0,1]} (\underline{x}_\alpha, \overline{x}_\alpha)$, де $\underline{x}_\alpha$ ($\overline{x}_\alpha$) -мінімальне (максимальне) значення $\tilde{x}$ на αрівні.

Визначення 28. *α-рівневий принцип узагальнення* Якщо $y = f(x_1, x_2, \ldots, x_n)$ -функція від n незалежних змінних та аргументи x_i задані нечіткими числами $\tilde{x}_i = \bigcup_{\alpha \in [0,1]} (\underline{x}_{i,\alpha}, \overline{x}_{i,\alpha})$, $i = \overline{1,n}$, то значенням функції $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ називається нечітке число $\tilde{y} = \bigcup_{\alpha \in [0,1]} (\underline{y}_\alpha, \overline{y}_\alpha)$, де $\underline{y}_\alpha = \inf_{\substack{x_{i,\alpha} \in [\underline{x}_{i,\alpha}, \overline{x}_{i,\alpha}], \\ i=\overline{1,n}}} (f(x_{1,\alpha}, x_{2,\alpha}, \ldots, x_{n,\alpha}))$ $\overline{y}_\alpha = \sup_{\substack{x_{i,\alpha} \in [\underline{x}_{i,\alpha}, \overline{x}_{i,\alpha}], \\ i=\overline{1,n}}} (f(x_{1,\alpha}, x_{2,\alpha}, \ldots, x_{n,\alpha}))$ i .

Застосування α-рівневого принципу узагальнення зводиться до рішення для кожного αрівня наступного завдання оптимізації: знайти максимальне і мінімальне значення функції $y = f(x_1, x_2, \ldots, x_n)$ за умови, що аргументи можуть приймати значення з

відповідних α-рівневих множин. Кількість α-рівнів вибирають так, щоб забезпечити необхідну точність обчислень.

Приклад 5 . Розв'язати задачу з прикладу 4 застосовуючи α-рівневий принцип узагальнення.

Будемо використовувати 2 наступні αрівні: {0, 1}. Тоді нечіткі аргументи задаються так: $\tilde{x}_1 = (1,4)_0 \cup (2,3)_1$і $\tilde{x}_2 = (2,8)_0 \cup (3,4)_1$. По α-рівневому принципу узагальнення отримуємо: $\tilde{y} = (2,32)_0 \cup (6,12)_1$. На рис. 8 показаний результат множення двох нечітких чисел $\tilde{y} = \tilde{x}_1 \cdot \tilde{x}_2$: червоними горизонтальними лініями зображені αперетину, а тонкою червоною лінією - шматково-лінійна апроксимація функції належності нечіткого числа $\tilde{y}$.

Досліджуємо, як змінитися результат нечіткого узагальнення зі збільшенням числа α -рівнів. Нечітке число $\tilde{y}$при заданні аргументів $\tilde{x}_1$і $\tilde{x}_2$на 41 αрівні показано на рис. 8. Синіми горизонтальними лініями зображені αперерізи нечіткої множини, а жирною синьою лінією - шматково-лінійна апроксимація функції належності нечіткого числа $\tilde{y}$ для 41 αрівня. Порівнюючи рис. 7 та 8, бачимо, що результати узагальнення за визначеннями 27 та 28 близькі.

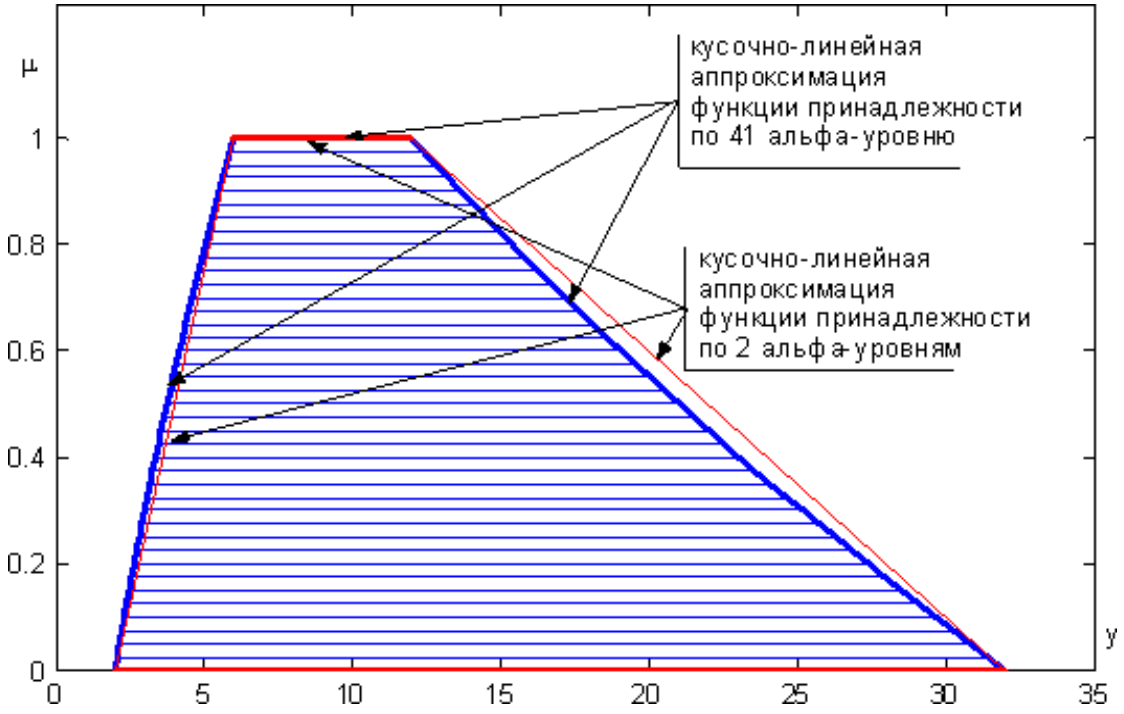


Рисунок 8 - Наприклад 5

Застосування α-рівневого принципу узагальнення дозволяє отримати правила виконання арифметичних операцій над нечіткими числами. Правила

виконання арифметичних операцій для позитивних нечітких чисел наведено у табл. 3. Ці правила необхідно застосовувати для кожного α-рівня.

Таблиця 3 - Правила виконання арифметичних операцій для позитивних нечітких чисел (для кожного α-рівня)

Арифметична операція	$\underline{y}$	$\overline{y}$
$\tilde{y} = \tilde{x}_1 + \tilde{x}_2$	$\underline{x}_1 + \underline{x}_2$	$\overline{x}_1 + \overline{x}_2$
	$\underline{x}_1 - \overline{x}_2$	
$\tilde{y} = \tilde{x}_1 \cdot \tilde{x}_2$	$\underline{x}_1 \cdot \underline{x}_2$	
$\tilde{y} = \frac{\tilde{x}_1}{\tilde{x}_2}$	$\frac{\underline{x}_1}{\overline{x}_2}$	$\frac{\overline{x}_1}{\underline{x}_2}$

Лекція №11. Синтез нечітких регуляторів систем управління нестаціонарними об'єктами

1. Синтез цифрових регуляторів систем управління на базі нечіткої логіки
2. Аналітичні вирази для керуючих впливів на виході нечіткого регулятора
2. Синтез цифрових нечітких регуляторів параметрами парового казана.

Нечітка логіка

Нечітка логіка це узагальнення традиційної аристотелевої логіки на випадок, коли істинність розглядається як лінгвістична змінна, що приймає значення типу: "дуже істинно", "менш істинно", "не дуже хибно" і т.п. Зазначені лінгвістичні значення видаються нечіткими множинами.

11.1. Лінгвістичні змінні

Нагадаємо, що лінгвістичною називається змінна, що приймає значення з безлічі слів або словосполучень деякої природної чи штучної мови. Безліч допустимих значень лінгвістичної змінної називається терм-множиною. Завдання значення змінної словами, без використання чисел, в людини природніше. Щодня ми приймаємо рішення на основі лінгвістичної інформації на кшталт: "дуже висока температура"; "тривала подорож"; "швидка відповідь"; "гарний букет"; "гармонійний смак" тощо. Психологи встановили, що у людському мозку майже вся числова інформація вербально перекодується і зберігається у вигляді лінгвістичних термів. Поняття лінгвістичної змінної відіграє важливу роль у нечіткому логічному висновку та у прийнятті рішень на основі наближених міркувань. Формально, лінгвістична змінна визначається в такий спосіб.

Визначення 44. *Лінгвістична змінна* визначається п'ятіркою $\langle x, T, U, G, M \rangle$, де x- ; ім'я змінної; T- ; терм-множина, кожен елемент якого (терм) представляється як нечітка множина на універсальній множині U; G- ; синтаксичні правила, часто у вигляді граматики, що породжують назву термів; M- ; семантичні правила, що задають функції належності нечітких термів, породжених синтаксичними правилами G.

приклад 9. Розглянемо лінгвістичну змінну під назвою $x =$"температура в кімнаті". Тоді четвірку, що залишилася, $\langle T, U, G, M \rangle$можна визначити так:

- універсальне безліч -; $U = [5, 35]$;
- терм-множина -; $T =$ {"холодно", "комфортно", "спекотно"} з такими функціями приладдям ($u \in U$):
  - $\mu_{\text{"холодно"}}(u) = \dfrac{1}{1 + \left(\dfrac{u - 10}{7}\right)^{12}}$
  - $\mu_{\text{"комфортно"}}(u) = \dfrac{1}{1 + \left(\dfrac{u - 20}{3}\right)^{6}}$
  - $\mu_{\text{"жарко"}}(u) = \dfrac{1}{1 + \left(\dfrac{u - 30}{6}\right)^{10}}$
- синтаксичні правила G, що породжує нові терми з використанням квантифікаторів "не", "дуже" і "менш-менш";
- семантичні правила Mу вигляді таблиці 4.

Таблиця 4 - Правила розрахунку функцій власності

Квантифікатор	Функція приладдя ($u \in U$)

не t $1 - \mu_t(u)$

дуже t $(\mu_t(u))^2$

більш-менш t $\sqrt{\mu_t(u)}$

Графіки функцій приналежності термів "холодно", "не дуже холодно", "комфортно", "менш комфортно", "жарко" і "дуже жарко" лінгвістичної змінної "температура в кімнаті" показані на рис. 13.

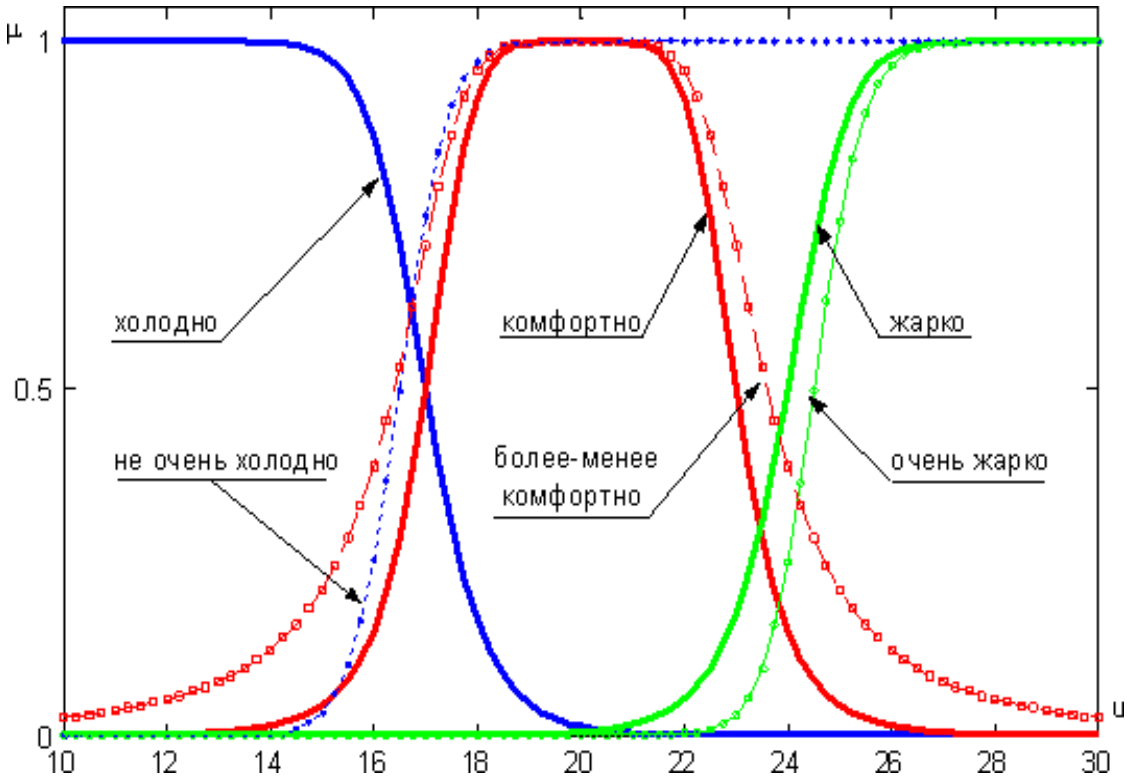


Рисунок 13 - Лінгвістична змінна "температура в кімнаті"

11.2. Нечітка істинність

Особливе місце у нечіткій логіці займає лінгвістична змінна "істинність". У класичній логіці істинність може приймати лише два значення: істинно та хибно. У нечіткій логіці істинність "розмита". Нечітка істинність визначається аксіоматично, причому різні автори роблять це по-різному. Інтервал [0, 1] використовується як універсальна множина для завдання лінгвістичної змінної "істинність". Звичайна, чітка істинність може бути представлена нечіткими множинами-синглтонами. І тут чіткому поняттю істинно відповідатиме функція власності

$\mu_{\text{"истинно"}}(u) = \begin{cases} 0, & u \neq 1 \\ 1, & u = 1 \end{cases}$, а чіткому поняттю хибно - ; $\mu_{\text{"ложно"}}(u) = \begin{cases} 0, & u \neq 0 \\ 1, & u = 0, \ u \in [0,1] \end{cases}$.

Для завдання нечіткої істинності Заде запропонував такі функції приналежності термів "істинно" та "хибно":

$$\mu_{\text{"истинно"}}(u) = \begin{cases} 0, & 0 \le u \le a \\ 2 \cdot \left(\frac{u-a}{1-a}\right)^2, & a < u \le \frac{a+1}{2} \\ 1 - 2 \cdot \left(\frac{u-1}{1-a}\right)^2, & \frac{a+1}{2} < u \le 1 \end{cases};$$

$$\mu_{\text{"ложно"}}(u) = \mu_{\text{"истинно"}}(1-u), \; u \in [0,1],$$

де $a \in [0,1]$-; параметр, що визначає носії нечітких множин "істинно" та "хибно". Для нечіткої множини "істинно" носієм буде інтервал $(a,1]$, а для нечіткої множини хибно" - ; $[0,a)$.

Функції приналежності нечітких термів "істинно" та "хибно" зображені на рис. 14. Вони побудовані за значенням параметра $a = 0.4$. Як видно, графіки функцій приналежності термів "істинно" і "хибно" є дзеркальними відображеннями.

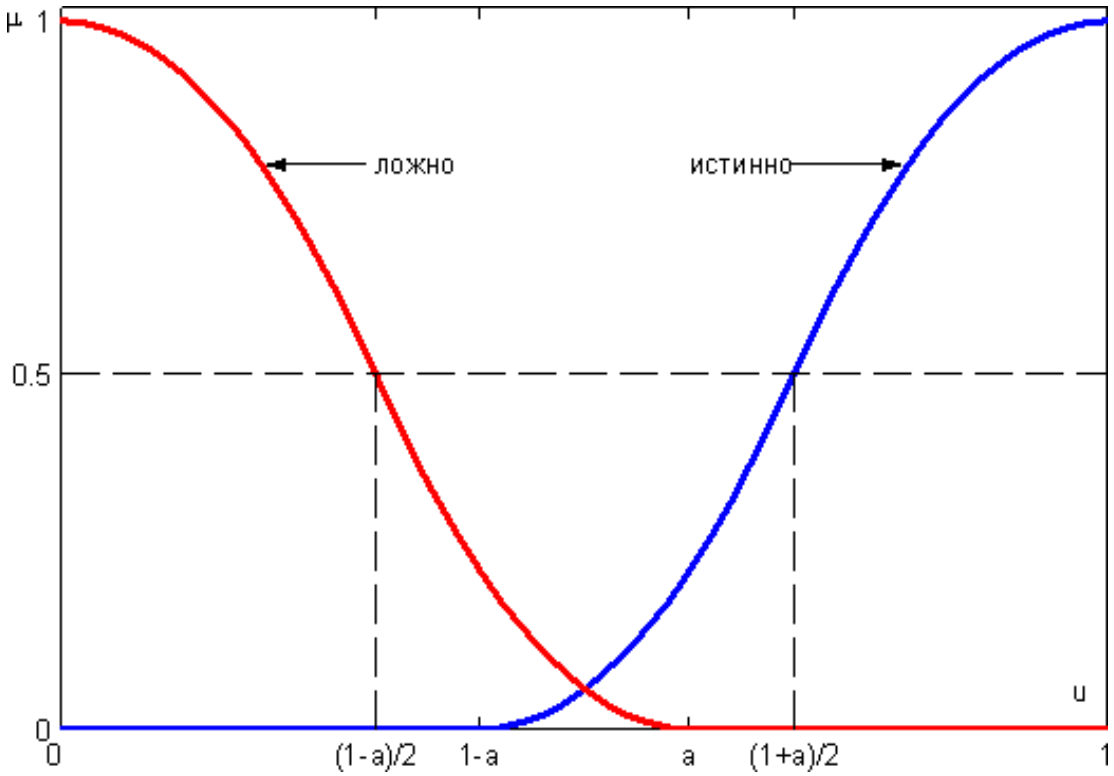


Рисунок 14 - Лінгвістична змінна "істинність" по Заді

Для завдання нечіткої істинності Балдвін запропонував такі функції приналежності нечітких "істинно" та "хибно":

$$\mu_{\text{"истинно"}}(u) = u;$$

$$\mu_{\text{"ложно"}}(u) = 1 - u;$$

де $u \in [0,1]$.

Квантифікатори "менш-менш" і "дуже" часто застосовують до нечіткими множинами "істинно" і "хибно", отримуючи таким чином терми "дуже хибно", "менш хибно", "більш істинно", "дуже істинно", "дуже, дуже істинно". Функції приналежності нових термів отримують, виконуючи операції концентрації та розтягування нечітких множин "істинно" та "хибно". Операція концентрації відповідає зведенню функції належності квадрат, а операція розтягування - зведенню в ступінь ½. Отже, функції приналежності термів "дуже, дуже хибно", "дуже хибно", "менш хибно", "більш-менш істинно", "істинно", "дуже істинно" і "дуже, дуже істинно" задаються так:

$$\mu_{\text{"очень ложно"}}(u) = \left(\mu_{\text{"ложно"}}(u)\right)^2;$$

$$\mu_{\text{"очень, очень ложно"}}(u) = \left(\mu_{\text{"очень ложно"}}(u)\right)^2$$

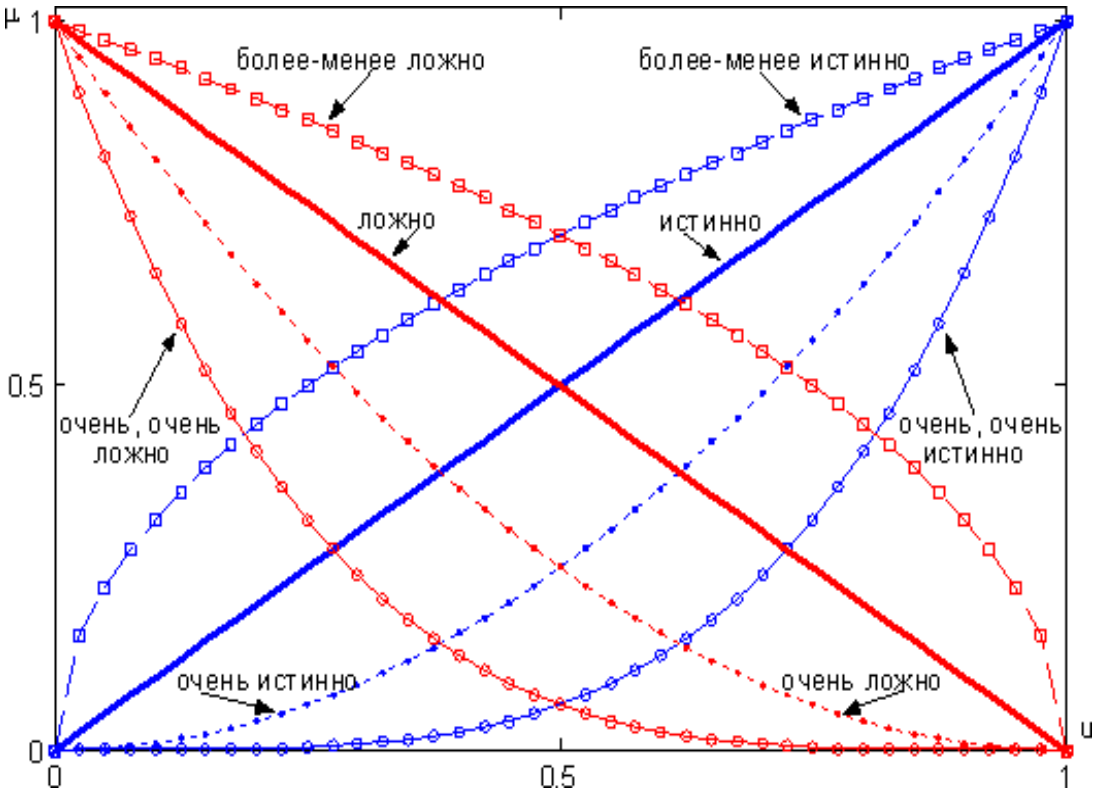
$$\mu_{\text{"более-менее ложно"}}(u) = \left(\mu_{\text{"ложно"}}(u)\right)^{1/2};$$

$$\mu_{\text{"более-менее истинно"}}(u) = \left(\mu_{\text{"истинно"}}(u)\right)^{1/2};$$

$$\mu_{\text{"очень истинно"}}(u) = \left(\mu_{\text{"истинно"}}(u)\right)^2;$$

$$\mu_{\text{"очень, очень истинно"}}(u) = \left(\mu_{\text{"очень истинно"}}(u)\right)^2.$$

Графіки функцій приналежності цих термів показано на рис. 15.



Малюнок 15 - Лінгвістична змінна "істинність" Балсусу

11.3. Нечіткі логічні операції

Спочатку коротко нагадати основні тези звичайної (бульової) логіки. Розглянемо два твердження A і B, кожне з яких може бути істинним чи хибним, тобто. приймати значення "1" чи "0". Для цих двох тверджень всього існує $2^{2^2} = 16$ різних логічних операцій, з яких змістовно інтерпретуються лише п'ять: І ($\wedge$), АБО ($\vee$), що виключає АБО ($\oplus$), імплікація ($\Rightarrow$) та еквівалентність ($\Leftrightarrow$). Таблиці істинності цих операцій наведено у табл. 5.

Таблиця 5 - Таблиці істинності булевої логіки

A	B	$A \wedge B$	$A \vee B$	$A \oplus B$	$A \Rightarrow B$	$A \Leftrightarrow B$
0	0	0	0	0	1	1
0	1	0	1	1	1	0

1	0	0	1	1	0	0
1	1	1	1	0	1	1

Припустимо, що логічне твердження може приймати не два значення істинності, а три, наприклад: "істинно", "хибно" та "невизначено". У цьому випадку ми матимемо справу не із двозначною, а тризначною логікою. Загальна кількість бінарних операцій, отже, і таблиць істинності, в тризначної логіці дорівнює $3^{3^2} = 729$. Нечітка логіка є різновидом багатозначної логіки, у якій значення істинності задаються лінгвістичними змінними чи термами лінгвістичної змінної "істинність". Правила виконання нечітких логічних операцій отримують із булевих логічних операцій за допомогою принципу узагальнення.

Визначення 45. Позначимо нечіткі логічні змінні через $\tilde{A}$ і $\tilde{B}$, а функції приналежності, що задають істинні значення цих змінних через $\mu_{\tilde{A}}(u)$ і $\mu_{\tilde{B}}(u)$, $u \in [0,1]$. *Нечіткі логічні операції* І ($\wedge$), АБО ($\vee$),

НЕ ($\blacksquare$) та імплікація ($\Rightarrow$) виконуються за такими правилами:

$$\mu_{\tilde{A}\wedge\tilde{B}}(u) = \min\left[\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u)\right];$$

$$\mu_{\tilde{A}\vee\tilde{B}}(u) = \max\left[\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u)\right];$$

$$\mu_{\overline{\tilde{A}}}(u) = 1 - \mu_{\tilde{A}}(u);$$

$$\mu_{\tilde{A}\Rightarrow\tilde{B}}(u) = \max\left[1 - \mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u)\right].$$

У багатозначній логіці логічні операції може бути задані таблицями істинності. У нечіткій логіці кількість можливих значень істинності може бути нескінченною, отже, у загальному вигляді табличне уявлення логічних операцій неможливе. Однак, у табличній формі можна представити нечіткі логічні операції для обмеженої кількості істиннісних значень, наприклад, для терм-множини {"істинно", "дуже істинно", "не істинно", "менш хибно", "хибно"}. Для тризначної логіки з нечіткими значеннями істинності T -; "істинно", F -; "хибно" та T+F - "невідомо" Л Заде запропонував такі лінгвістичні таблиці істинності:

$\tilde{A}$	$\tilde{B}$	$\overline{\tilde{A}}$	$\tilde{A}\wedge\tilde{B}$	$\tilde{A}\vee\tilde{B}$
T	T	F	T	T
T	F	F	F	T
T	T+F	F	T+F	T
F	T	T	F	T
F	F	T	F	F
F	T+F	T	F	T+F
T+F	T	T+F	T+F	T
T+F	F	T+F	F	T+F
T+F	T+F	T+F	T+F	T+F

Застосовуючи правила виконання нечітких логічних операцій з визначення 45, можна розширити таблиці істинності для більшої кількості термів. Як це зробити, розглянемо на наступному прикладі.

Приклад 10. Задано такі нечіткі істинні значення:

$$\text{истинно} = 0/0 + 0/0.2 + 0.25/0.4 + 0.5/0.6 + 0.9/0.8 + 1/1;$$

;

$$\text{почти истинно} = 0/0 + 0/0.05 + 0.4/0.4 + 0.7/0.6 + 1/0.8 + 0.8/1.$$

Застосовуючи правило з визначення 45, знайдемо нечітку істинність виразу "майже істинно АБО істинно":

$$\text{почти истинно} \vee \text{истинно} = 0/0 + 0.05/0.2 + 0.4/0.4 + 0.7/0.6 + 1/0.8 + 1/1.$$

Порівняємо отриману нечітку множину з нечіткою множиною "менш істинно". Вони майже рівні, отже:

$$\text{почти истинно} \vee \text{истинно} \approx \text{более - менее истинно}.$$

В результаті виконання логічних операцій часто виходить нечітка множина, яка не еквівалентна жодному з раніше введених нечітких значень істинності. У цьому випадку необхідно серед нечітких значень істинності знайти таке, що відповідає результату виконання нечіткої логічної операції максимально. Іншими словами, необхідно провести так звану *лінгвістичну апроксимацію* , яка може розглядатися як аналог апроксимації емпіричного статистичними розподілами стандартними функціями розподілу випадкових величин. Як приклад наведемо запропоновані Балдвіном лінгвістичні таблиці істинності для показаних на рис. 15 нечітких значень істинності:

$\tilde{A}$	$\tilde{B}$	$\tilde{A} \wedge \tilde{B}$	$\tilde{A} \vee \tilde{B}$
хибно	хибно	хибно	хибно
істинно	хибно	хибно	істинно
істинно	істинно	істинно	істинно
невизначено	хибно	хибно	невизначено
невизначено	істинно	невизначено	істинно
невизначено	невизначено	невизначено	невизначено
істинно	дуже істинно	істинно	дуже істинно
істинно	більш-менш істинно	більш-менш істинно	істинно

11.3. Нечітка база знань

Визначення 46. *Нечіткою базою знань* називається сукупність нечітких правил "Якщо - то", що визначають взаємозв'язок між входами та виходами досліджуваного об'єкта. Узагальнений формат нечітких правил такий:

Якщо *посилка правила,* **висновок** *правила.*

Посилання правила або антецедент є твердженням типу "x є низький", де "низький" - це терм (лінгвістичне значення), заданий нечіткою множиною на універсальній множині лінгвістичної змінної x. Квантифікатори "дуже", "менш-менш", "ні", "майже" тощо. можуть використовуватись для модифікації термів антецедента.

Висновок або наслідок правила є твердженням типу "y є d", в якому значення вихідної змінної (d) може задаватися:

1. нечітким термом: "y є високий";
2. класом рішень: "y є бронхіт"
3. чіткою константою: "y=5";
4. чіткою функцією від вхідних змінних: "y=5+4*x".

Якщо значення вихідний змінної у правилі задано нечітким безліччю, тоді правило може бути нечітким ставленням. Для нечіткого правила "Якщо x є $\tilde{A}$, тобто y $\tilde{B}$", нечітке ставлення $\tilde{R}$ задається на декартовому творі $U_X \times U_Y$, де U_X (U_Y)-; універсальна множина вхідної (вихідної) змінної. Для розрахунку нечіткого відношення можна застосовувати нечітку імплікацію та t-норму. При використанні як t-норми операції знаходження мінімуму розрахунок нечіткого відношення $\tilde{R}$ здійснюється так:

$$\mu_{\tilde{R}}(x,y) = \min\left[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)\right], \quad (x,y) \in U_X \times U_Y.$$

Приклад 11. Наступна нечітка база знань описує залежність між віком водія (x) та можливістю дорожньо-транспортної пригоди (y):

Якщо *x = молодий,* **то** *y = висока;*

Якщо *x = середній,* **то** *y = низька;*

Якщо *x = дуже старий,* **то** *y = висока.*

Нехай функції приладдя термів мають вигляд, показаний на рис. 16. Тоді нечіткі відносини, що відповідають правилам бази знань, будуть такими, як на рис. 17.

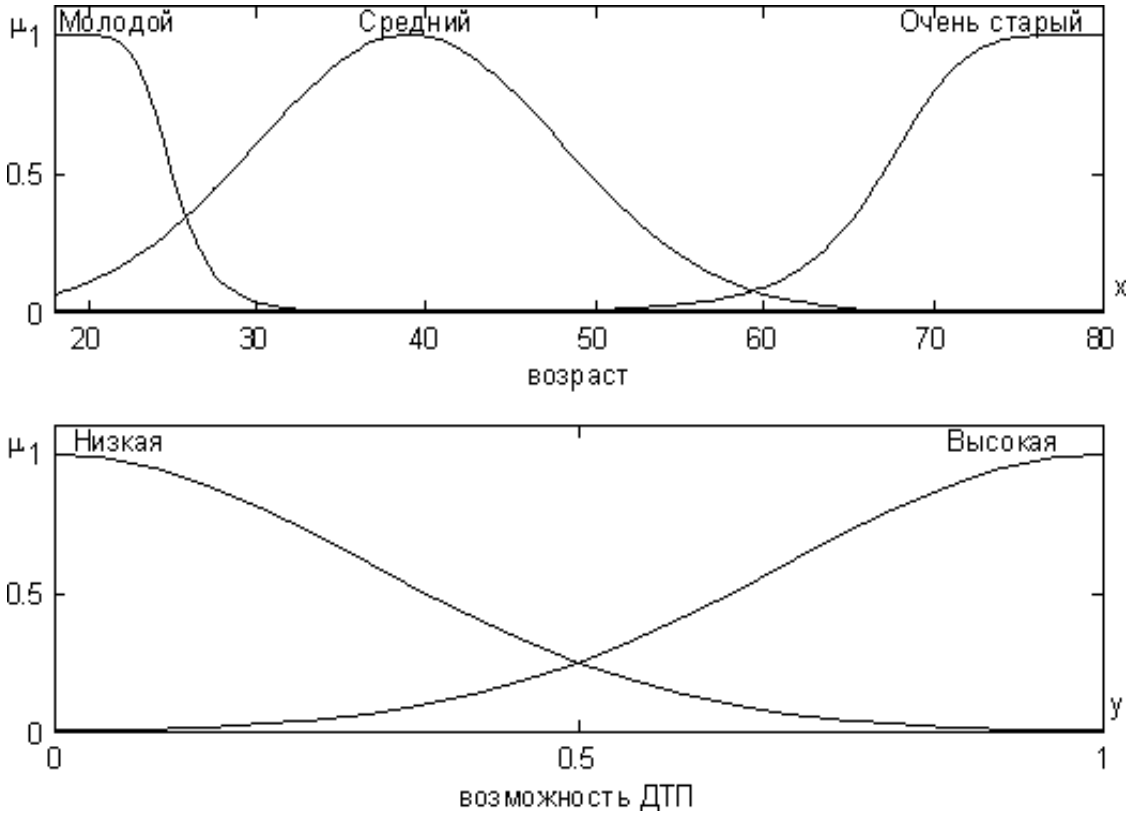


Рисунок 16 - Функції приладдя термів

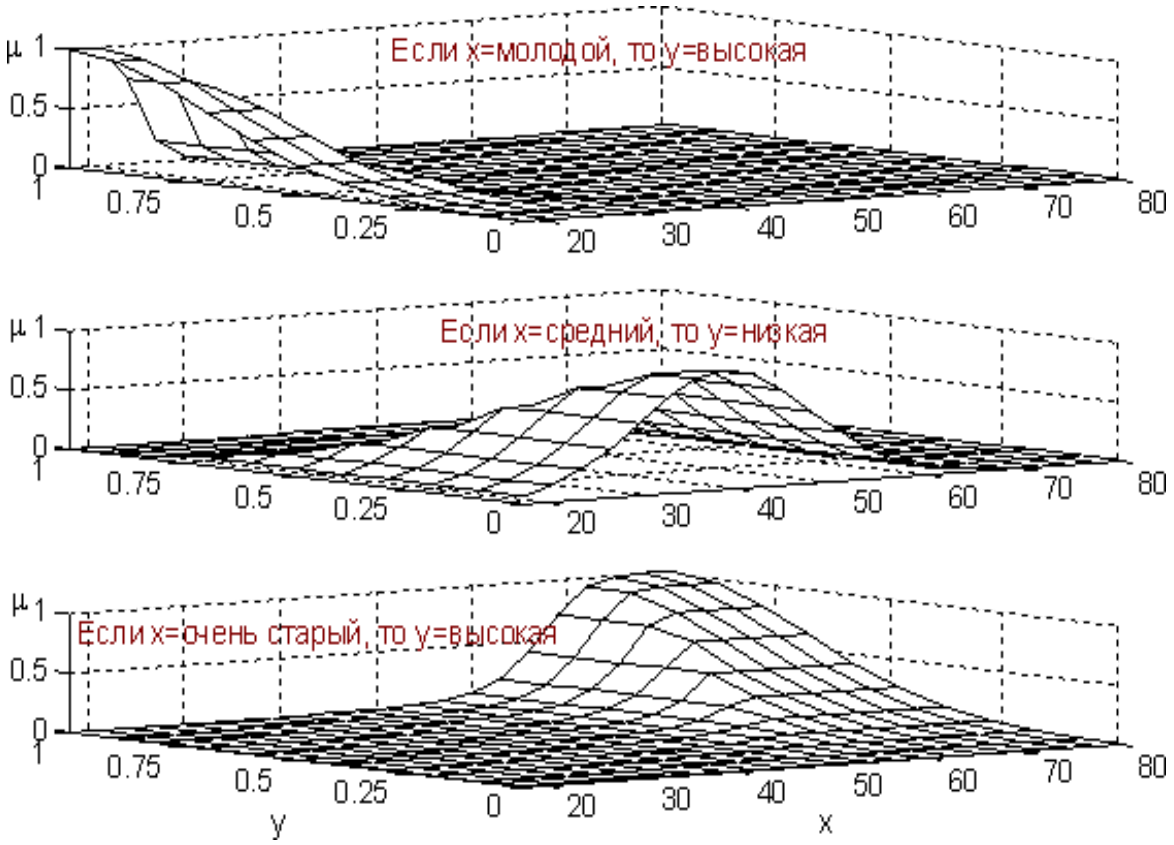


Рисунок 17 - Нечіткі відносини, що відповідають правилам бази знань на прикладі 11

Для завдання багатовимірних залежностей "входи-виходи" використовують нечіткі логічні операції І та АБО. Зручно правила формулювати те щоб усередині кожного правил змінні об'єднувалися логічної операцією І, а правила з урахуванням знань пов'язувалися операцією АБО. В цьому випадку нечітку базу знань, що зв'язує входи $X = (x_1, x_2, \ldots, x_n)$ з виходом y, можна уявити в наступному вигляді:

ЯКЩО $(x_1 = a_{1,j1})$ И $(x_2 = a_{2,j1})$ И ... И $(x_n = a_{n,j1})$

АБО $(x_1 = a_{1,j2})$ И $(x_2 = a_{2,j2})$ И ... И $(x_n = a_{n,j2})$

…

АБО $(x_1 = a_{1,jk_j})$ И $(x_2 = a_{2,jk_j})$ И ... И $(x_n = a_{n,jk_j})$,

ТО $y = d_j$, $j = \overline{1,m}$,

де $a_{i,jp}$ - нечіткий терм, яким оцінюється змінна x_i у рядку з номером jp ($p = \overline{1,k_j}$);

k_j - кількість рядків-кон'юнкцій, у яких вихід y оцінюється значень d_j;

m - кількість різних значень, що використовуються для оцінки вихідної змінної y.

Наведену вище базу знань зручно представляти таблицею, яку іноді називають матрицею знань (табл. 6).

Таблиця 6 - Нечітка база знань

x_1	x_2	…	x_n	y

$a_{1,1,1}$	$a_{2,1,1}$	…	$a_{n,1,1}$	
$a_{1,1,2}$	$a_{2,1,2}$	…	$a_{n,1,2}$	
				d_1
…	…	…	…	
$a_{1,1,k_1}$	$a_{2,1,k_1}$	…	$a_{n,1,k_1}$	
$a_{1,2,1}$		…	$a_{n,2,1}$	
$a_{1,2,2}$	$a_{2,2,2}$	…	$a_{n,2,2}$	
				d_2
…	…	…	…	
$a_{1,2,k_2}$	$a_{2,2,k_2}$	…	$a_{n,2,k_2}$	
…				
$a_{1,m,1}$	$a_{2,m,1}$	…	$a_{n,m,1}$	
$a_{1,m,2}$	$a_{2,m,2}$	…	$a_{n,m,2}$	
				d_m
…	…	…	…	
a_{1,m,k_m}	a_{2,m,k_m}	…	a_{n,m,k_m}	

Для обліку різного ступеня впевненості експерта адекватності правил використовують вагові коефіцієнти. Нечітку базу знань з таблиці 6 з ваговими коефіцієнтами правил можна записати так:

$$\bigcup_{p=1}^{k_j}\left(\bigcap_{i=1}^{n} x_i = a_{i,jp} \text{ с весом } w_{jp}\right) \to y = d_j, \quad j = \overline{1,m},$$

де $\bigcup$-; нечітка логічна операція АБО;

$\bigcap$ - ;нечітка логічна операція І;

$w_{jp} \in [0,1]$ - ; ваговий коефіцієнт правила з номером jp.

.7. Нечітка логіка

11.5. Нечіткий логічний висновок

11.5.1. Композиційне правило нечіткого висновку

Звичайний, булевий логічний висновок базується на наступних тавтологіях:

- модус поненс: $(A \wedge (A \Rightarrow B)) \Rightarrow B$;
- модус толленс: $((A \Rightarrow B) \wedge \overline{B}) \Rightarrow \overline{A}$;
- силогізм: $((A \Rightarrow B) \wedge (B \Rightarrow C)) \Rightarrow (A \Rightarrow C)$;
- Контрапозиція: $(A \Rightarrow B) \Rightarrow (\overline{B} \Rightarrow \overline{A})$.

У чіткому логічному висновку найчастіше застосовується правило модус поненс, яке можна записати так:

Посилання	A є істинно
Імплікація	Якщо A, то B
Логічний висновок	B є істинно

Модус поненс виводить висновок "B є істинним", якщо відомо, що "A є істинним" і існує правило "Якщо A, то B" (A і B - чіткі логічні твердження). Однак, якщо прецедент відсутній, то модус поненс не зможе вивести жодного, навіть наближеного висновку. Навіть якщо відомо, що близьке до A твердження A' є істинним, модус поненс не може бути застосований. Одним із можливих способів прийняття рішень за невизначеної інформації є застосування нечіткого логічного висновку.

Визначення 47. *Нечітким логічним висновком* називається отримання висновку у вигляді нечіткої множини, що відповідає поточним значенням входів, з використанням нечіткої бази знань та нечітких операцій.

Основу нечіткого логічного висновку становить композиційне правило Заде.

Визначення 48. *Композиційне правило виведення Заде* формулюється наступним чином: якщо відомо нечітке відношення $\tilde{R}$ між вхідною (x) і вихідною (y) змінними, то при нечіткому значенні вхідної змінної $x = \tilde{A}$, нечітке значення вихідної змінної визначається так:

$$y = \tilde{A} \circ \tilde{R},$$

де $\circ$ – максмінна композиція.

Приклад 12. Дано нечітке правило "Якщо $x = \tilde{A}$, то $y = \tilde{B}$" з нечіткими множинами: $\tilde{A} = 0/1 + 0.1/2 + 0.5/3 + 0.8/4 + 1/5$ і $\tilde{B} = 1/5 + 0.8/10 + 0.4/15 + 0.2/20$. Визначити значення вихідної змінної y, якщо $x = \tilde{C} = 0.3/1 + 0.5/2 + 1/3 + 0.7/4 + 0.4/5$.

На початку розрахуємо нечітке відношення, що відповідає правилу "Якщо $x = \tilde{A}$, то $y = \tilde{B}$", застосовуючи як t-норму операцію знаходження мінімуму:

$$\tilde{R} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 \\ 0.5 & 0.5 & 0.4 & 0.2 \\ 0.8 & 0.8 & 0.4 & 0.2 \\ 1 & 0.8 & 0.4 & 0.2 \end{bmatrix}.$$

Тепер, за формулою, $y = \tilde{C} \circ \tilde{R}$ розрахуємо нечітке значення вихідної змінної:

$y = 0.7/5 + 0.7/10 + 0.4/15 + 0.2/20$.

11.5.2. Нечіткий логічний висновок Мамдані

Нечіткий логічний висновок за алгоритмом Мамдані виконується за нечіткою базою знань:

$$\bigcup_{p=1}^{k_j} \left(\bigcap_{i=1}^{n} x_i = a_{i,jp} \;\; \text{с весом } w_{jp} \right) \rightarrow y = d_j, \quad j = \overline{1,m},$$

в якій значення вхідних та вихідний змінної задані нечіткими множинами. Введемо такі позначення, необхідні подальшого викладу матеріалу:

$\mu_{jp}(x_i)$ - функція власності входу x_i нечіткому терму $a_{i,jp}$, тобто. $a_{i,jp} = \int_{\underline{x}_i}^{\bar{x}_i} \mu_{jp}(x_i)/x_i$, $x_i \in \left[\underline{x}_i, \bar{x}_i\right]$.

$\mu_{d_j}(y)$ - функція власності виходу y нечіткому терму d_j, тобто. $d_j = \int_{\underline{y}}^{\bar{y}} \mu_{d_j}(y)/y$, $y \in \left[\underline{y}, \bar{y}\right]$.

Ступені приналежності вхідного вектора нечітким термам d_j з бази знань розраховується так:

$$\mu_{d_j}\left(X^*\right) = \bigvee_{p=1,k_j} w_{jp} \cdot \bigwedge_{i=1,n} \left[\mu_{jp}\left(x_i^*\right) \right], \; j = \overline{1,m},$$

де $\vee(\wedge)$ - операція із s-норми (t-норми), тобто. з безлічі реалізацій логічної операцій АБО (І). Найчастіше використовуються такі реалізації: для операції АБО – знаходження максимуму та для операції І – знаходження мінімуму.

В результаті отримуємо таку нечітку множину $\tilde{y}$, що відповідає вхідному вектору X^*:

$$\tilde{y} = \frac{\mu_{d_1}(X^*)}{d_1} + \frac{\mu_{d_2}(X^*)}{d_2} + \ldots + \frac{\mu_{d_m}(X^*)}{d_m}.$$

Особливістю цієї нечіткої множини є те, що універсальною множиною для нього є терм-множина вихідний змінної y. Такі нечіткі множини називаються нечіткими множинами другого порядку.

Для переходу від нечіткої множини, заданої на універсальній множині нечітких термів $\{d_1, d_2, \ldots, d_m\}$ до нечіткої множини на інтервалі $\left[\underline{y}, \bar{y}\right]$ необхідно: 1) "зрізати" функції приналежності $\mu_{d_j}(y)$ на рівні $\mu_{d_j}(X^*)$; 2) об'єднати (агрегувати) отримані нечіткі множини. Математично це записується так:

$$\tilde{y} = \underset{j=1,m}{agg} \left(\int_{\underline{y}}^{\bar{y}} \min\left(\mu_{d_j}(X^*), \; \mu_{d_j}(y) \right) / y \right),$$

де agg - агрегування нечітких множин, що найчастіше реалізується операцією знаходження максимуму.

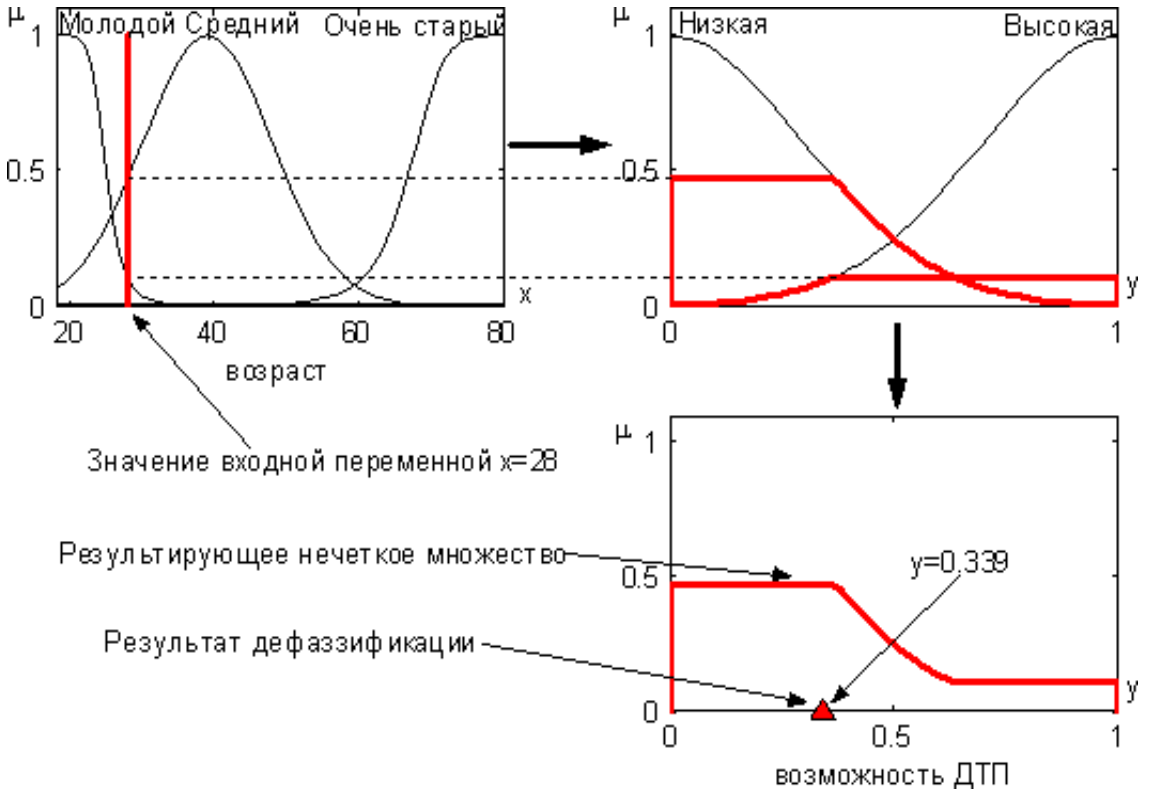
Чітке значення виходу y, що відповідає вхідному вектору X^* визначається в результаті дефазифікації нечіткої множини $\tilde{y}$. Найчастіше застосовується дефаззифікація методом центру тяжкості:

$$y = \frac{\int_{\underline{y}}^{\overline{y}} y \cdot \mu_{\tilde{y}}(y)dy}{\int_{\underline{y}}^{\overline{y}} \mu_{\tilde{y}}(y)dy},$$

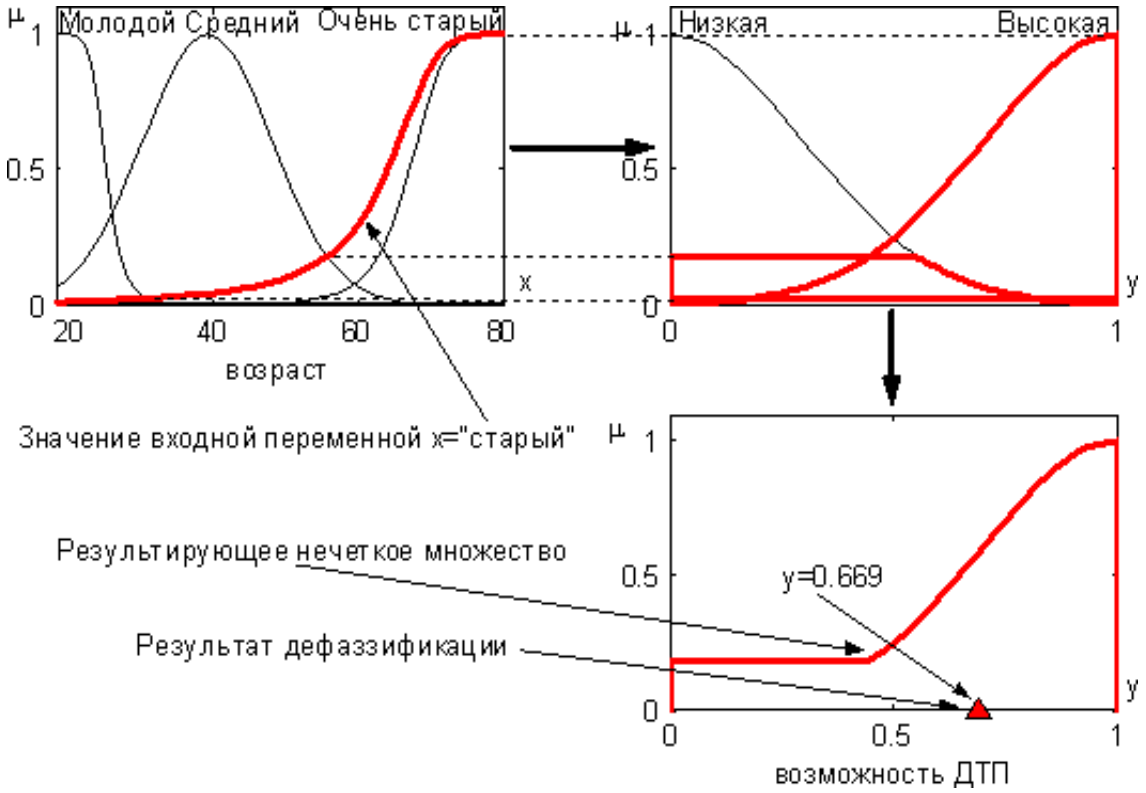
де $\int$ – тут символ інтеграла.

Приклад 13. За нечіткою базою знань з прикладу 11 виконати нечіткий логічний висновок при значеннях вхідної змінної $x = 28$ та $x =$ "старый".

Виконання нечіткого логічного висновку при значеннях вхідної змінної $x = 28$ та $x =$ "старый" показано на рис. 18 та 19. Операція агрегування здійснювалася знаходженням максимуму. Дефаззифікація проводилася методом центру тяжкості. На рис. 20 показана залежність "вхід-вихід", що відповідає нечіткій базі знань з прикладу 11. Ділянки графіка, відповідні першому, другому та третьому правилу бази знань позначені на малюнку #1, #2 і #3.



Малюнок 18 - Нечіткий логічний висновок Мамдані при чіткому значенні вхідної змінної



Малюнок 19 - Нечіткий логічний висновок Мамдані при нечіткому значенні вхідної змінної

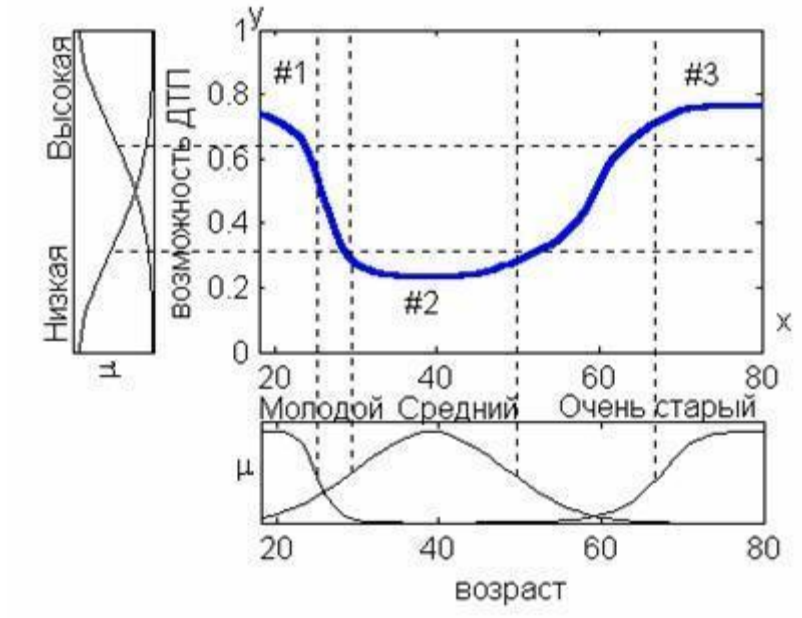


Рисунок 20 -Залежність "вхід-вихід" для нечіткої бази знань з прикладу 11

11.5.3. Нечіткий логічний висновок Сугено

Нечіткий логічний висновок за алгоритмом Сугено (іноді говорять алгоритм Такагі-Сугено) виконується за нечіткою базою знань:

$$\bigcup_{p=1}^{k_j}\left(\bigcap_{i=1}^{n} x_i = a_{i,jp} \;\text{ с весом } w_{jp}\right) \rightarrow y = b_{j,0} + b_{j,1}\cdot x_1 + b_{j,2}\cdot x_2 + \ldots + b_{j,n}\cdot x_n, \quad j = \overline{1,m},$$

де b_{ji} – деякі числа.

База знань Сугено аналогічна базі знань Мамдані крім висновків правил d_j, які задаються не нечіткими термами, а лінійною функцією від входів : $d_j = b_{j,0} + \sum_{i=\overline{1,n}} b_{j,i} \cdot x_i$. Правила в основі знань Сугено є свого роду перемикачами з одного лінійного закону "входи - вихід" на інший, також лінійний. Кордони підобластей розмиті, отже, одночасно можуть виконуватися кілька лінійних законів, але з різними ступенями. Ступіні приналежності вхідного вектора до значень $d_j = b_{j,0} + \sum_{i=\overline{1,n}} b_{j,i} \cdot x_i$ розраховується так:

$$\mu_{d_j}(X^*) = \bigvee_{p=1,k_j} w_{jp} \cdot \bigwedge_{i=1,n} \left[\mu_{jp}\left(x_i^*\right) \right], \; j = \overline{1,m},$$

де $\vee(\wedge)$- операція із s-норми (t-норми), тобто. з безлічі реалізацій логічної операцій АБО (І). У нечіткому логічному висновку Сугено найчастіше використовуються такі реалізації трикутних норм: імовірнісне АБО як s-норма і твір як t-норма.

В результаті отримуємо таку нечітку множину $\tilde{y}$, що відповідає вхідному вектору X^*:

$$\tilde{y} = \frac{\mu_{d_1}(X^*)}{d_1} + \frac{\mu_{d_2}(X^*)}{d_2} + \ldots + \frac{\mu_{d_m}(X^*)}{d_m}.$$

Звернемо увагу, що на відміну від результату висновку Мамдані, наведена вище нечітка множина є звичайною нечіткою множиною першого порядку. Воно задано на безлічі чітких чисел. Результуюче значення виходу y визначається як суперпозиція лінійних залежностей, що виконуються у цій точці X^* $n-$мірного факторного простору. Для цього дефаззифікують нечітку множину $\tilde{y}$, знаходячи зважену середню $y = \frac{\sum_{j=1,m} \mu_{d_j}(X^*) \cdot d_j}{\sum_{j=1,m} \mu_{d_j}(X^*)}$ або зважену суму $y = \sum_{j=1,m} \mu_{d_j}(X^*) \cdot d_j$.

Приклад 14. Відома нечітка база знань:

Якщо x = низький, **то** $y = 3x$;

Якщо x = високий, **то** $y = 3 - x$.

Функції приладдя термів задані такими виразами:

$$\mu_{\text{"низкий"}}(x) = \exp\left(-\frac{x^2}{0.18}\right) \text{ та } \mu_{\text{"высокий"}}(x) = \exp\left(-\frac{(x-1)^2}{0.18}\right), \; x \in [0,1].$$

Необхідно виконати нечіткий логічний висновок при значенні вхідної змінної $x = 0.4$.

Виконання нечіткого логічного висновку показано на рис. 21. Дефаззифікація проводилася методом центру тяжкості (зваженого середнього). На рис. 22 показано залежність "вхід-вихід" для наведеної вище нечіткої бази знань. Ділянки графіка, відповідні першому та другому правилу бази знань позначені малюнку #1 і #2.

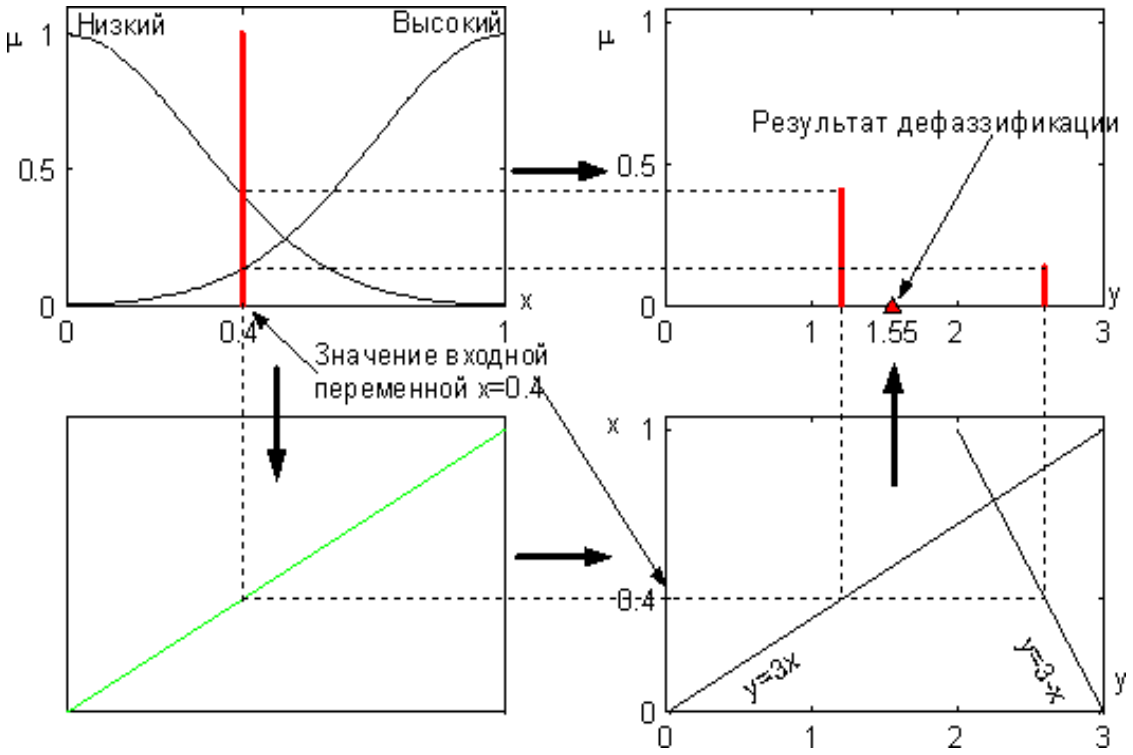


Рисунок 21 - Виконання нечіткого логічного висновку Сугено для прикладу 14

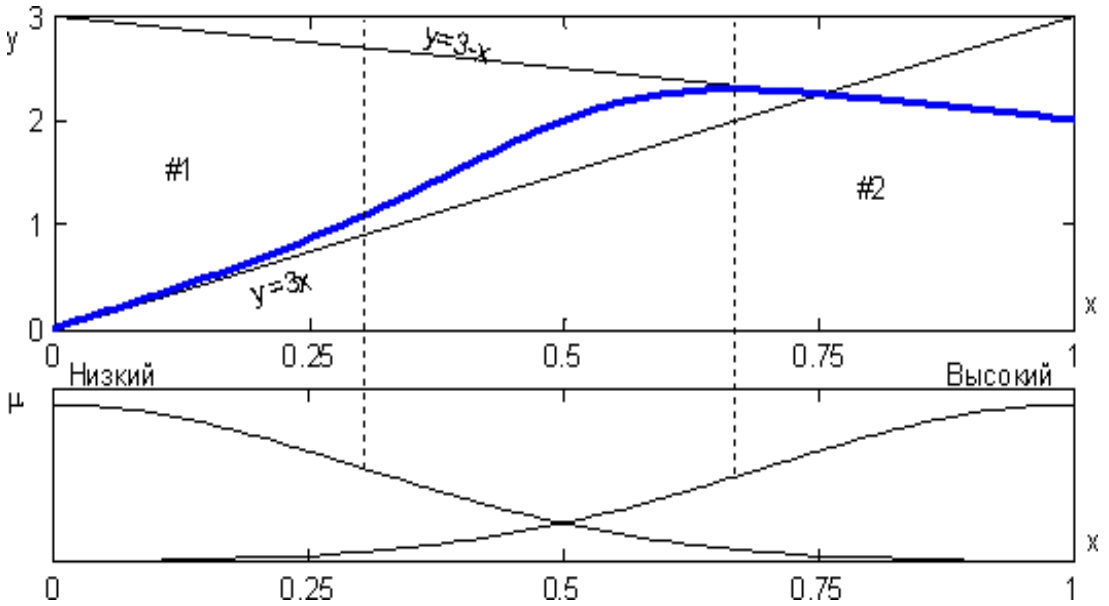


Рисунок 22 - Залежність "вхід-вихід" для нечіткої бази знань з прикладу 14

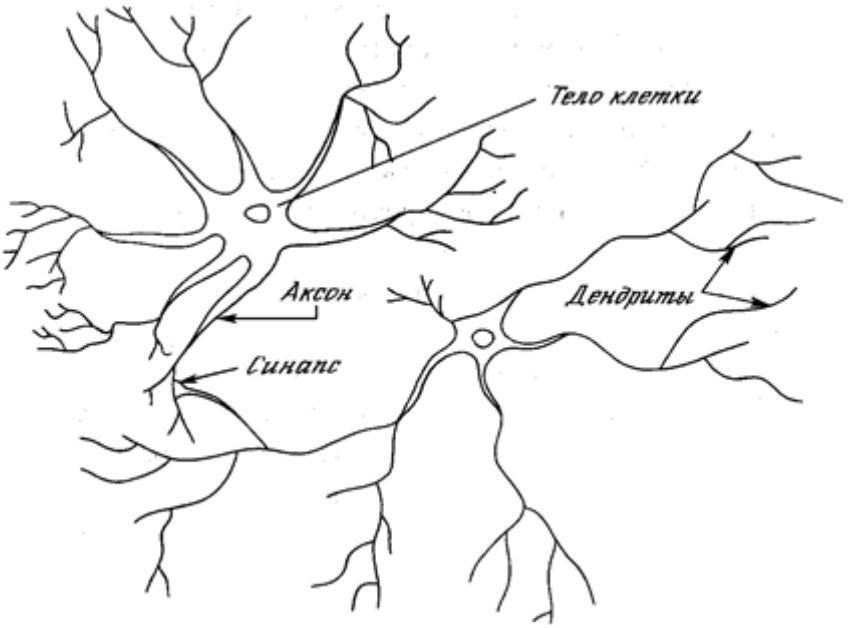
Тема 6. Нейро-Мережеві алгоритми

Лекція №12 Нейронні мережі. Базові поняття

1. Поняття нейрокомп'ютера
2. Класифікація нейронних мереж
3. Штучний нейрон
4. Правило навчання Хебба
5. Парадигми навчання

Модель штучного нейрона.

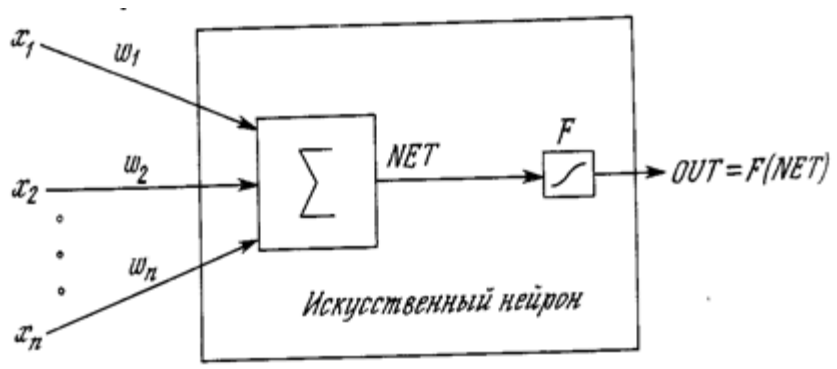
Штучна нейронна мережа (ІНС) – це спрощена модель біологічного мозку, точніше нервової тканини. Нервова система людини, побудована з елементів, званих нейронами, має разючу складність. Близько 10^{11} нейронів беруть участь у приблизно 10^{15} передавальних зв'язках, що мають довжину метр і більше. Кожен нейрон має багато якостей, загальними з іншими елементами тіла, але його унікальною здатністю є прийом, обробка та передача електрохімічних сигналів по нервових шляхах, які утворюють комунікаційну систему мозку. Структура пари типових біологічних нейронів показано на рис. 39. Нейрон складається з тіла (соми), що містить ядро, і відростків - дендритів , за якими в нейрон надходять вхідні сигнали. Один із відростків, що гілкується на кінці, служить для передачі вихідних сигналів даного нейрона іншим нервовим клітинам. Він називається аксоном. Сполука аксона з дендритом іншого нейрона називається синапсом. Нейрон збуджується і передає сигнал через аксон, якщо кількість збуджуючих сигналів, що прийшли по дендритах, більше, ніж число гальмують. Незважаючи на те, що ця функціональна схема нейрона має багато ускладнень і винятків, більшість штучних нейронних мереж моделюють лише ці прості властивості.



Мал. 39. Біологічний нейрон.

Штучний нейрон (мал. 40), далі просто нейрон, задається сукупністю своїх входів, позначених $x_1, x_2, \ldots, x_n$ вагами входів , функцією стану NET і функцією активації F. Функція стану визначає стан нейрона в залежності від значень його входів, ваг входів і, можливо, попередніх станів. Найчастіше використовуються функції стану, що обчислюються як сума творів значень входів на ваги відповідних входів по всіх входах. Підсумовуючий блок, що відповідає тілу біологічного нейрона, складає зважені входи алгебраїчно, створюючи вихід, який називатимемо NET.

Сигнал NET перетворюється функцією активації F і дає вихідний сигнал нейрона $OUT = F(NET)$. NET = x1 * w1 + x2 * w2 + ... + xn * wn

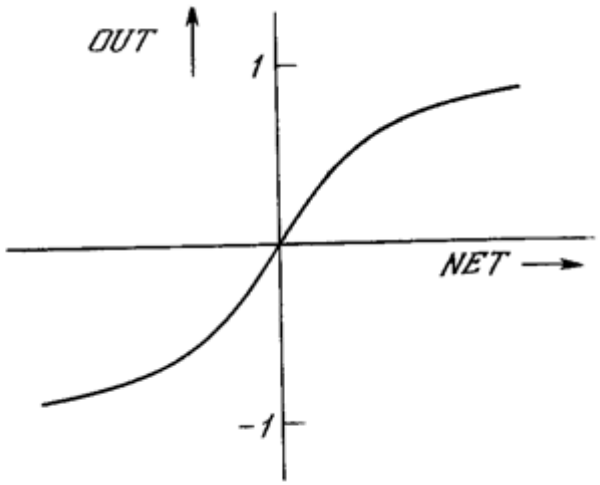


Мал. 41. Штучний нейрон із активаційною функцією F.

На рис. 41 блок, позначений F приймає сигнал NET і видає сигнал OUT. Якщо блок F звужує діапазон зміни величини NET так, що при будь-яких значеннях NET значення OUT належать деякому кінцевому інтервалу, то F називається функцією, що «стискає» . Як «стискаючу» функцію часто використовується сигмоїдальна (S-подібна) функція, показана на рис. 42.

Мал. 42. Сігмоїдальна функція.

Ця функція має вигляд: $OUT = 1/(1 + e^{-NET})$. За аналогією з електронними системами активаційну функцію вважатимуться нелінійною підсилювальною характеристикою штучного нейрона. Коефіцієнт посилення обчислюється як відношення збільшення величини OUT до невеликого збільшення величини NET, що викликало його. Центральна область сигмоїдальної функції, що має великий коефіцієнт посилення, вирішує проблему обробки слабких сигналів, у той час як області з падаючим посиленням на позитивному та негативному кінцях підходять для великих збуджень. Таким чином, нейрон функціонує з великим посиленням у широкому діапазоні рівня вхідного сигналу. Інший широко використовується функцією активації є гіперболічний тангенс OUT = th(x). Подібно до сигмоїдальної функції гіперболічний тангенс є S-подібною функцією, але він симетричний щодо початку координат, і в точці NET = 0 значення вихідного сигналу OUT дорівнює нулю (рис. 43). На відміну від сигмоїдальної функції гіперболічний тангенс набуває значень різних знаків, що виявляється вигідним при пошуку екстремуму в просторі параметрів ІНС.



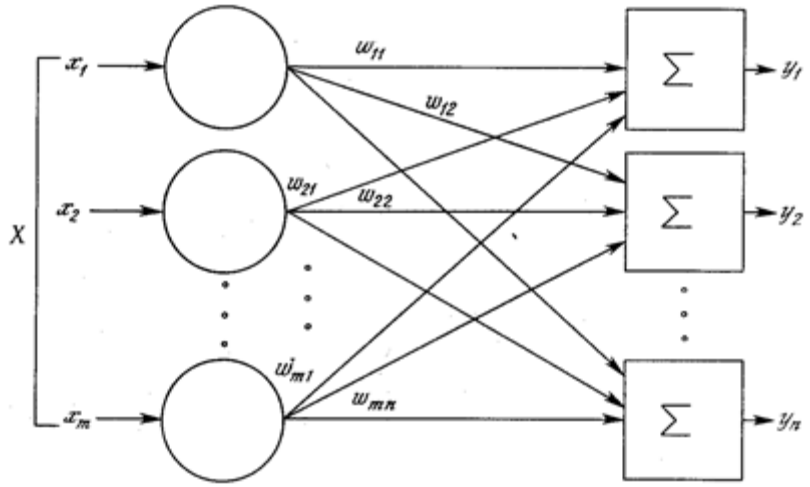
Мал. 43. Функція гіперболічного тангенсу.

Найпростішими формами функції активації є лінійна та ступінчаста (порогова) функції. Лінійна функція активації диференційована і легко обчислюється, що у ряді випадків дозволяє зменшити помилки вихідних сигналів в ІНС. Проте вона універсальна і забезпечує рішення багатьох погано формалізованих завдань. Ступінчаста функція зазвичай задається у вигляді одиничної функції з жорсткими обмеженнями: вона дорівнює нулю, якщо NET < 0, і одиниці, якщо NET 0. Така функція не забезпечує достатньої гнучкості ІНС при навчанні. Розглянута проста модель штучного нейрона ігнорує багато властивостей свого біологічного двійника. Наприклад, вона не бере до уваги затримки в часі, що впливають на динаміку системи. Незважаючи на ці обмеження, мережі, побудовані з цих нейронів, виявляють властивості, що сильно нагадують біологічну систему.

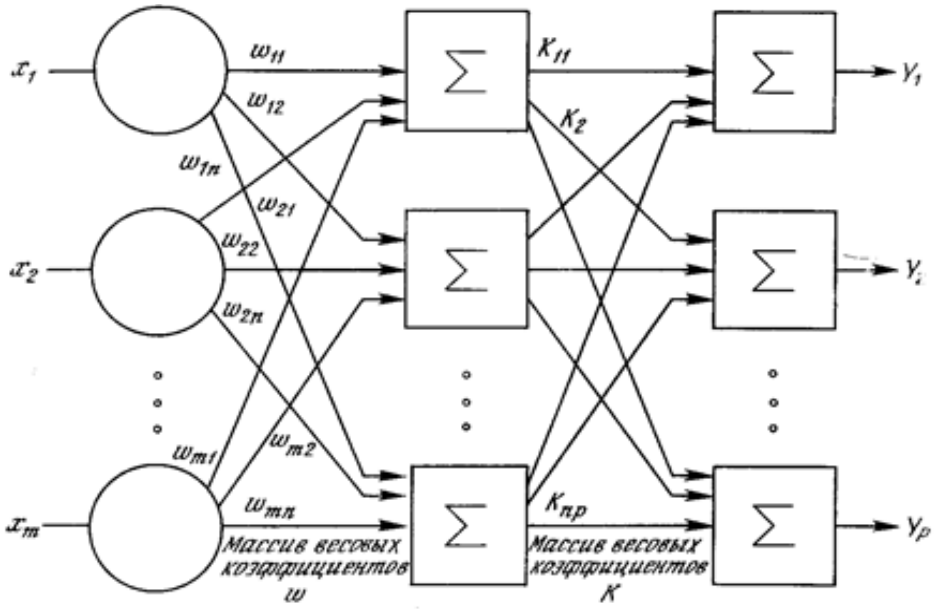
Моделі нейронних мереж.

Нейронна мережа є сукупністю штучних нейронів, організованих шарами. Реальна нейронна мережа може містити один або більше шарів і відповідно характеризуватись як одношарова або як багатошарова, зі зворотними зв'язками і без них. Багатошарові мережі відрізняються від одношарових тим, що між вхідними та вихідними даними розташовуються кілька так званих прихованих шарів нейронів, що додають більше нелінійних зв'язків у модель. Найпростіша мережа складається з групи нейронів, що утворюють один шар (рис. 44). На малюнку вершини - кола зліва служать лише розподілу вхідних сигналів. Вони не виконують якихось обчислень, і тому не будуть вважатися шаром. Тому вони позначені колами, щоб відрізняти їх від обчислювальних нейронів, позначених квадратами. Кожен елемент з множини входів X окремою вагою з'єднаний з кожним штучним нейроном, і кожен нейрон видає зважену суму входів у мережу. У штучних та біологічних мережах багато сполук можуть бути відсутні, тому на малюнку всі сполуки показані з метою спільності.

Багатошарові мережі мають більші обчислювальні можливості, ніж одношарові. Пошарова організація нейронів копіює шаруваті структури певних відділів мозку. Багатошарові мережі можуть утворюватися каскадами шарів: вихід одного шару є входом наступного шару. Двошарова мережа з усіма сполуками показано на рис. 45. Якщо функція активації є лінійною, то нейронна мережа також буде лінійною. Двошарова лінійна мережа еквівалентна одному шару з ваговою матрицею, що дорівнює добутку двох вагових матриць, тобто $\mathrm{NET} = (\mathrm{XW}_1)\mathrm{W}_2 = \mathrm{X}(\mathrm{W}_1\mathrm{W}_2)$. Отже, будь-яка багатошарова лінійна мережа може бути замінена еквівалентною одношаровою мережею. Для розширення обчислювальних можливостей ІНС застосовується нелінійна функція активації.



Мал. 44. Одношарова нейронна мережа



Мал. 45. Двошарова нейронна мережа.

По архітектурі зв'язків нейронні мережі можуть бути згруповані в два класи: мережі прямого поширення, в яких зворотні зв'язки відсутні (немає з'єднань, що йдуть від виходів деякого шару до входів цього ж шару або попередніх шарів) та мережі рекурентного типу, в яких можливі зворотні зв'язки. У мережах прямого поширення немає пам'яті, їх вихід повністю визначається поточними входами та значеннями ваги. Мережі прямого поширення поділяються на одношарові перцепротрони (мережі) та багатошарові перцептрони (мережі). Назва перцептрон для нейромереж введено Ф. Розенблаттом, розробником першої нейромережі (1957). Він же довів збіжність області рішень для перцептрону під час навчання. Реккурентні мережі можуть мати властивості, подібні до короткочасної людської пам'яті.

Клас рекурентних нейронних мереж набагато ширший і складніший за своїм пристроєм. Поведінка рекурентних мереж описується диференціальними чи різницевими рівняннями, зазвичай першого порядку. Це значно розширює сфери застосування нейромереж та способи їх навчання. Мережа організована так, що кожен нейрон отримує вхідну інформацію від інших нейронів, можливо, і від самого себе, і навколишнього середовища. Цей тип мереж має значення для моделювання нелінійних динамічних систем. Серед рекурентних мереж можна виділити мережі Хопфілда та мережі Кохонена. При всьому різноманітті можливих змін ІНС на практиці набули поширення лише деякі з них. Класичними моделями нейронних мереж є:

1. Багатошарові перцептрони містять крім вхідного та вихідного шарів так звані приховані шари. Вони є нейрони, які мають безпосередніх входів вихідних даних, а пов'язані лише з виходами вхідного шару і з входом вихідного шару. Таким чином, приховані шари додатково перетворюють

інформацію та додають нелінійності моделі. Багатошаровий перцептрон із сигмоїдальними функціями активації здатний апроксимувати будь-яку функціональну залежність (доведено у вигляді теореми), проте при цьому не відомо ні потрібне число шарів, ні потрібна кількість прихованих нейронів, ні необхідний для навчання мережі час. Ці проблеми досі не вирішені;

2. Мережі Хопфілда будуються з N нейронів, пов'язаних кожен з кожним, крім самого себе, причому всі нейрони є вихідними. Нейронну мережу можна використовувати як асоціативну пам'ять, а також для обробки неупорядкованих, упорядкованих у часі або просторі зразків (рукописні літери, часові ряди, графіки);

3. Мережі Кохонена ще називають «картами ознак, що самоорганізуються». Мережа розрахована на самостійне навчання. У процесі навчання на вхід мережі подаються різні зразки. Мережа вловлює особливості їх структури і поділяє зразки на кластери, а потім вже навчена мережа відносить кожен знову приклад до одного з кластерів, керуючись деяким критерієм «близькості». Мережа складається з одного вхідного та одного вихідного шару. Кількість елементів у вихідному шарі безпосередньо визначає, скільки різних кластерів мережа зможе розпізнати. Кожен із вихідних елементів отримує на вхід весь вхідний вектор. Як і у будь-якій нейронній мережі, кожному зв'язку приписана деяка синаптична вага. Найчастіше кожен вихідний елемент з'єднаний зі своїми сусідами. Ці внутрішньошарові зв'язку відіграють важливу роль у процесі навчання, так як коригування ваги відбувається тільки в околиці того елемента, який найкраще відгукується на черговий вхід. Вихідні елементи змагаються між собою право вступити у дію «отримати урок». Виграє той із них, чий вектор ваг виявиться найближчим до вхідного вектора.

Структура ІНС, включаючи визначення числа шарів та числа нейронів у кожному шарі, формується до початку навчання, тому успішне вирішення цієї проблеми багато в чому залежить від досвіду та мистецтва аналітика, який проводить дослідження.

Навчання нейронних мереж.

Головна відмінність та перевага нейромереж перед класичними засобами прогнозування та класифікації полягає в їх здатності до навчання. На етапі навчання відбувається обчислення синаптичних коефіцієнтів у процесі вирішення нейронної мережею завдань, у яких необхідна відповідь визначається не за правилами, а за допомогою прикладів, згрупованих у навчальні множини. Нейросеть на етапі навчання сама виконує роль експерта у процесі підготовки даних для побудови експертної системи. Передбачається, що правила перебувають у структурі навчальних даних. Такі дані є рядами прикладів із зазначенням для кожного з них значення вихідного параметра, яке було б бажано отримати. Дії, що при цьому відбуваються, можна назвати навчання з учителем. На вхід мережі подається вектор вихідних даних, а вихідний вузол повідомляється бажане значення результату обчислень. Контрольоване навчання нейромережі можна як рішення оптимізаційної завдання. Її метою є мінімізація функції помилок на даному множині прикладів шляхом вибору значень ваг W. Досягнення мінімуму називається збіжністю процесу навчання. Оскільки помилка залежить від ваги нелінійно, отримати рішення в аналітичній формі неможливо, і пошук глобального мінімуму здійснюється за допомогою ітераційного процесу, так званого навчального алгоритму. В даний час використовуються більше сотні різних навчальних алгоритмів, що відрізняються один від одного стратегією оптимізації та критерієм помилок. Зазвичай як міру похибки береться середня квадратична помилка

$$E = \sqrt{\frac{\sum_{i=1}^{M} (d_i - y_i)^2}{M}},$$

де M – число прикладів у множині, що навчається.

Серед навчальних алгоритмів найпоширенішим є алгоритм зворотного розповсюдження помилок. Відповідно до методу перед початком навчання мережі всім міжнейронних зв'язків надаються

невеликі випадкові значення ваг. Кожен крок навчальної процедури і двох фаз. Під час першої фази вхідні елементи мережі встановлюються заданий стан. Вхідні сигнали поширюються мережею, породжуючи деякий вихідний вектор. При цьому використовуються функції сигмоїдної активації. Отриманий вихідний вектор порівнюється з потрібним (правильним) вектором. Якщо вони збігаються, то вагові коефіцієнти зв'язків не змінюються. В іншому випадку обчислюється різниця між фактичними та необхідними вихідними значеннями, яка передається послідовно від вихідного шару до вхідного шару. На основі цієї інформації проводиться модифікація ваги зв'язків відповідно до формули:

$$\Delta W_{ij}(t+1) = \mu \Delta W_{ij}(t) - (1-\mu)\varepsilon \frac{\partial E}{\partial W_{ij}},$$

де $\Delta W_{ij}(t)$ - зміна ваги зв'язку на етапі ітерації t навчальної пари вхід-вихід і-го нейрона, що отримує вхідний сигнал, і j-го нейрона, що посилає вихідний сигнал; m - коефіцієнт, що задається користувачем в інтервалі (0,1); e – величина кроку мережі або міра точності навчання мережі. У сучасних нейромережевих пакетах користувач може сам визначати, як буде змінюватися величина кроку мережі, дуже часто за замовчуванням береться лінійна або експонентна залежність величини кроку від кількості ітерацій мережі. Чим менший крок мережі, тим більше часу знадобиться на навчання мережі і тим можливішим стане її потрапляння в околицю локального мінімуму помилки.

Загальна помилка функціонування мережі визначається за такою формулою:

$$D = \frac{1}{2}\sum_p \sum_j (T_{jp} - R_{jp})^2,$$

де T_{jp} - R_{jp} відповідно бажане і дійсне значення вихідного сигналу j - го блоку для p - ї навчальної пари нейронів вхід-вихід.

Коли величина помилки досягає прийнятно малого рівня, навчання зупиняють, і мережа готова до виконання покладених на неї завдань. Важливо, що вся інформація, яку мережа набуває завдання, міститься у наборі прикладів. Тому якість навчання мережі залежить кількості прикладів у навчальної вибірці, і навіть від цього, наскільки повно ці приклади описують завдання. Вважається, що для повноцінного тренування потрібно хоча б кілька десятків (а краще за сотні) прикладів. Якщо не всім прикладів навчальної вибірки відомі правильні відповіді, то навчання мережі проводиться без вчителя. У цьому випадку застосування самонастроюваних мереж Кохонена дає можливість визначити внутрішню структуру даних, що надходять у мережу, і розподілити зразки за категоріями.

Незважаючи на численні успішні застосування алгоритму зворотного поширення під час навчання ІНС, він має недоліки. Найбільше неприємностей завдає невизначено довгий процес навчання. У складних завданнях для навчання мережі можуть знадобитися дні або навіть тижні, вона може взагалі не навчитися. Тривалий час навчання може бути результатом неоптимального вибору етапу мережі. Невдачі у навчанні мережі зазвичай виникають із двох причин.

1. Параліч мережі. У процесі навчання мережі значення ваги можуть в результаті корекції стати дуже великими величинами. Це може призвести до того, що всі або більшість нейронів будуть функціонувати при дуже великих значеннях OUT, в області, де похідна функції стискання дуже мала. Оскільки посилана у процесі навчання помилка пропорційна цієї похідної, процес навчання може практично завмерти. У теоретичному відношенні цю проблему погано вивчено. Зазвичай цього уникають зменшенням розміру кроку, але це збільшує час навчання. Для запобігання паралічу застосовуються різні евристики, але поки що вони можуть розглядатися лише як експериментальні.

2. Попадання до локального мінімуму. Зворотне поширення використовує різновид градієнтного спуску, тобто здійснює спуск по поверхні помилки, безперервно підлаштовуючи ваги в напрямку мінімуму. Поверхня помилки складної мережі сильно порізана і складається з пагорбів, долин, складок та ярів у просторі високої розмірності. Мережа може потрапити в локальний мінімум

(неглибоку долину), коли поряд є набагато глибший мінімум. У точці локального мінімуму всі напрямки ведуть вгору, і мережа не може вибратися з нього.

Способи реалізації ІНС.

Нейронні мережі можуть бути реалізовані програмним або апаратним способом. Варіантом апаратної реалізації є нейрокомп'ютери. Більшість сучасних нейрокомп'ютерів є персональним комп'ютером, з додатковою нейроплатою. Підвищений інтерес викликають спеціалізовані нейрокомп'ютери, в яких реалізовані принципи архітектури нейронних мереж. У випадках, коли розробка чи використання апаратних реалізацій нейронних мереж обходиться занадто дорого, застосовують дешевші програмні реалізації. Програму моделювання нейронної мережі зазвичай називають програмою-імітатором або нейропакетом, розуміючи під цим програмну оболонку, що емулює серед користувача нейрокомп'ютера на звичайному комп'ютері. В даний час на ринку програмного забезпечення є безліч найрізноманітніших програм для моделювання нейронних мереж. У них втілено практично всі відомі алгоритми навчання та топології нейромереж. Загальна кількість фірм, які розробляють ці кошти, перевищує 150. Незважаючи на складність закладених у нейропакетах методів, використовувати їх досить просто. Вони дозволяють сконструювати, навчити, протестувати і використовувати у роботі нейронну мережу з урахуванням розуміння кількох базових теоретичних положень, викладених вище. Перші подібні програмні продукти з'явилися на Заході в середині 80-х років XX ст. Одним із лідерів цього ринку став нейромережевий пакет BrainMaker американської фірми California Scientific Software. Нині нейрокомп'ютерний сегмент ринку програмного забезпечення бурхливо розвивається, і нових пакетів – природний процес. Сучасні продукти багато в чому краще за своїх попередників - покращено інтерфейс користувача, з'явилися додаткові нейропарадигми, в пакетах реалізовані можливості взаємодії з іншими додатками за допомогою механізмів OLE, ActiveX та ін. Плата за ці можливості використовувати більш потужний, ніж раніше комп'ютер, зростання вимог до обсягу оперативної та дискової пам'яті. Крім того, новий пакет не гарантує якіснішого рішення прикладного завдання користувача. Більшою мірою це якість залежить від правильно поставленого завдання, вибору даних. Ефективним реальним інструментом для розрахунку та проектування нейронних мереж при вирішенні численних прикладних завдань у багатьох галузях техніки, в тому числі електротехніки та електроніки, є електроустаткування пакет прикладних програм Neural Network Toolbox (ППП NNT), що функціонує під управлінням системи Matlab версії 5.3 і 6.0. Слід також звернути увагу на інтерфейс ППП NNT із системою Simulink, що дозволяє наочно відобразити архітектуру мережі та виконати моделювання як статичних, так і динамічних нейронних мереж. Остання обставина особливо важлива при побудові систем керування різними електротехнічними об'єктами, діагностики їх стану та поведінки.

Лекція №13. Нейроуправління

1. Ідентифікація динамічних ланок
2. Нейроемулятори та нейропредиктори
3. Концепція нейроуправління
4. Інверсне нейроуправління

Штучні нейронні мережі як засіб управління

Принципи застосування ІНС у задачах управління

При побудові систем управління з допомогою штучних нейронних мереж використовують їх властивість до навчання, тобто. відтворення заданих реакцій, а також до адаптації - підстроювання реакцій в процесі функціонування умов, що змінюються, завдання управління, структурних і параметричних змінах об'єкта. Крім того, ІНС, будучи спочатку нелінійним об'єктом, може виступати як блок лінеаризації об'єкта або як нелінійний регулятор.

Доцільність застосування ІНС виникає при роботі з об'єктом, математичний опис якого неповний, дуже складний або взагалі частковий. Про об'єкт можуть бути відомі тільки його реакції на деяку обмежену кількість керуючих впливів, знову ж таки, обмеженому діапазоні. Здатність нейронних мереж до інтерполяції та екстраполяції (функціонування за аналогією), виявлення спільності ситуацій є важливою перевагою при використанні їх для роботи зі слабко визначеними об'єктами.

Окремою особливістю ІНС є відсутність необхідності складних аналітичних побудов при їх проектуванні та навчанні.

Наслідувальне нейроуправління

При наслідуючому нейроуправлінні штучна нейронна мережа є нейроемулятором звичайного регулятора. На етапі навчання метою є вихід регулятора, вхідними сигналами - сигнали завдання та зворотного зв'язку, затримані в часі на необхідну кількість тактів.

Навчання проводять для кількох робочих точок у всьому діапазоні регулювання. Зауважимо, що з нелінійному об'єкті параметри еталонного регулятора залежить від робочої точки.

Нейронна мережа неспроможна у разі сформувати якісніше управління, ніж традиційний регулятор. Але при керуванні об'єктом зі змінними параметрами, коли звичайний регулятор потребує переналаштування, ІНС може запам'ятати кілька налаштувань регулятора і за зміни режиму роботи об'єкта продовжити керування ним без перенавчання.

Нейроуправління з еталонною моделлю

Цей варіант є розвитком попереднього, коли структура та параметри регулятора не визначені. Тоді навчання ІНС може бути проведене на основі еталонної моделі контуру регулювання, що відповідає бажаному показнику якості.

ІНС навчається на основі сигналу неузгодженості еталонної моделі та виходу моделі об'єкта управління. Реальний об'єкт у такому навчанні використовувати не можна, оскільки можливі неприпустимі стану системи. З цієї причини математична модель об'єкта повинна мати прийнятну точність принаймні в робочому діапазоні регулювання. Якщо таку модель отримати не вдається, можна скористатися ІНС як нейроемулятор, тобто. пристроєм, що відтворює реакції об'єкта при вибраних впливах. Навчання проводиться традиційним способом: на основі експерименту та/або статистичних даних формується набір навчальних шаблонів, на

основі яких навчають нейроемулятор. Далі на його основі може бути натренований нейроконтролер.

Прогнозуюче нейроуправління

При прогнозуючому управлінні контролер повинен вибрати оптимальну стратегію управління кілька тактів розрахунку вперед, оцінюючи можливу реакцію об'єкта. Для цього потрібна модель об'єкта керування, яка може бути розрахована досить швидко – швидше, ніж протікають у реальному часі. За такт управління слід оцінити кілька стратегій та вибрати оптимальну у певному сенсі.

Як модель об'єкта доцільний для застосування нейроемулятор. Високий паралелізм нейронної мережі дозволяє досить швидко виконати обчислення та оцінити реакцію об'єкта на керування.

Реалізація нейроконтролерів

Високий ступінь паралелізму, властивий нейронним мережам, може бути реалізований тільки при використанні спеціалізованих засобів управління. Це або так звані нейропроцесори, або програмовані логічні матриці, які також використовують паралельні алгоритми обробки даних.

Якщо реалізації нейроконтролера використовується традиційний (або сигнальний) мікропроцесор, властивість паралелізму втрачається навіть у багатоядерних системах. Також слід сказати і про програмну реалізацію нейронної мережі на обчислювальній машині, яка, фактично, емулює нейронну мережу.

Необхідність спеціального апаратного забезпечення є певною перешкодою для поширення нейроконтролерів, які працюють у реальному часі. Справа в тому, що при емуляції нейронної мережі виникає необхідність у великій кількості множень та функціональних перетворень (згадаймо модель нейрона). Тому як регулятор (контролер) нейронні мережі використовують сьогодні для об'єктів, процеси в яких протікають істотно повільніше по відношенню до процесів розрахунку керуючих впливів.

Запитання для самоконтролю

1. Перерахуйте основні структури систем керування, які використовують штучні нейронні мережі.
2. Поясніть поняття "Нейроконтролер".
3. Поясніть поняття «Нейроемулятор».
4. У чому полягає наслідуюче нейроуправління?
5. У чому полягає прогнозне нейроуправління?

Тема 6. Проблеми оптимальної експлуатації технологічного обладнання

Лекція № 14. Необхідні умови оптимальності для Основне завдання синтезу закону управління. Метод динамічного програмування

1. Завдання синтезу раціонального закону управління
2. Принцип оптимальності динамічного програмування
3. Зведення загальних процедур методу динамічного програмування для обчислення оптимального закону управління u * = v * (t, x)

Лекція №15. Прикладні аспекти моделювання складних технічних та технологічних процесів.

1. Розв'язання прикладних завдань.
2. Застосування пактів прикладних програм.

Нейроуправління (англ. Neurocontrol) - окремий випадок інтелектуального управління, що використовує штучні нейронні мережі для вирішення завдань управління динамічними об'єктами. Нейроуправління знаходиться на стику таких дисциплін як штучний інтелект, нейрофізіологія, теорія автоматичного управління, робототехніка. Нейронні мережі мають ряд унікальних властивостей, які роблять їх потужним інструментом для створення систем управління: здатністю до навчання на прикладах і узагальнення даних, здатністю адаптуватися до зміни властивостей об'єкта управління і зовнішнього середовища, придатністю для синтезу нелінійних регуляторів, високою стійкістю до пошкоджень своїх елементів в силу закладеного в нейросеті. Термін "нейроуправління", вперше був використаний одним з авторів методу зворотного поширення помилки Полом Дж. Вербосом в 1976 [1] [2]. Відомі численні приклади практичного застосування нейронних мереж для вирішення завдань управління літаком [3] [4], вертольотом [5], автомобілем-роботом [6], швидкістю обертання вала двигуна [7], гібридним двигуном автомобіля [8], електропіччю [9], турбогенератором [10] 3], моделлю перевернутого маятника [14].

Зміст

Методи нейроуправління Схема прямого
нейроуправління

зі зворотним зв'язком. На такті k нейроконтролер отримує на вхід статутне значення r(k+1) та оцінку поточного стану об'єкта S(k) і генерує керуючий вплив u(k), переводячи об'єкт управління у нове положення y(k+1). За способом використання нейронних мереж методи нейроуправління діляться на прямі методи та непрямі методи. У прямих методах нейронна мережа навчається безпосередньо генерувати керуючі на об'єкт, у непрямих методах нейронна мережа навчається виконувати допоміжні функції: ідентифікація об'єкта управління, придушення шуму, оперативне налаштування коефіцієнтів ПІД-контролера. Залежно від числа нейромереж, що становлять нейроконтролер, системи нейроуправління діляться на одномодульні та багатомодульні. Системи нейроуправління, що застосовуються разом із традиційними регуляторами, називаються гібридними.

У завданнях нейроуправління представлення об'єкта управління використовують модель чорного ящика, у якому спостерігаються поточні значення входу і виходу. Стан об'єкта вважається недоступним для зовнішнього спостереження, хоча розмір вектора станів зазвичай вважається

фіксованою. Динаміку поведінки об'єкта управління можна у дискретному вигляді:

де: – стан об'єкта управління порядку N на такті k; - значення P-вимірного вектора управління на такті k, - значення V-вимірного виходу об'єкта управління на такті k + 1.

Для оцінки поточного стану об'єкта управління S(k) може бути використана модель NARX, що складається з минулих положень об'єкта y і затриманих сигналів управління u:

Вектор оцінки стану S може бути також представлений без використання затриманих сигналів:

хема наслідуючого
нейроуправління
: зліва – режим навчання нейронної мережі; справа – режим управління
Наслідувальне нейроуправление [15] [16] [17] (Neurocontrol learning based on mimic, Controller Modeling, Supervised Learning Using an Existing Controller) охоплює системи нейроуправления, у яких нейроконтроллер навчається з прикладів динаміки звичайного контролера по зворотному зв'язку, побудованого, наприклад, на основі. Після навчання нейронна мережа точно відтворює функції вихідного контролера. Як приклад динаміки контролера може бути використана запис поведінки людини-оператора. Звичайний контролер зворотного зв'язку (чи людина-оператор) управляє об'єктом управління штатному режимі. Значення величин на вході та виході контролера протоколюються, і на основі протоколу формується навчальна вибірка для нейронної мережі, що містить M пар значень входу Pi та очікуваних реакцій Ti нейромережі:

Після навчання за допомогою, наприклад, методу зворотного поширення помилки, нейронна мережа підключається замість вихідного контролера. Отриманий нейроконтролер може замінити людину в управлінні пристроєм, а також бути вигіднішим економічно, ніж вихідний контролер.
Схема узагальненого інверсного
нейроуправління
: ліворуч – режим навчання інверсного нейроемулятора; справа – режим управління об'єктом У схемі **узагальненого інверсного нейроуправления** (Generalized Inverse Neurocontrol, Direct Inverse Neurocontrol, Adaptive Inverse Control) [18] [19] як контролера використовується нейронна модель інверсної динаміки об'єкта управління, звана інверсний нейроэмулятор. Інверсний нейроемулятор є нейронною мережею, навченою в режимі офф-лайн імітувати зворотну динаміку об'єкта управління на основі записаних траєкторій поведінки динамічного об'єкта. Для отримання таких траєкторій на об'єкт управління в якості керуючого сигналу подають деякий випадковий процес. Значення керуючих сигналів і реакцій у відповідь об'єкта протоколюють і на цій основі формують навчальну вибірку :

У ході навчання, нейронна мережа повинна вловити і запам'ятати залежність значень керуючого сигналу u(k 1) від наступного значення реакції об'єкта управління y(k), що знаходиться перед цим у стані S(k 1). При управлінні об'єктом, інверсний нейроемулятор підключається як контролер, отримуючи при цьому на вхід x(k) значення уставки r(k + 1) і стану об'єкта управління S(k), що надходить по каналу зворотного зв'язку:

Передбачається, що сформована при навчанні інверсна модель об'єкта управління є адекватною, отже сигнал управління, видається.
Спеціалізоване інверсне нейрокерування
Спеціалізоване інверсне нейрокерування (Specialised Inverse Neurocontrol) [19] [18] використовує методику навчання нейроконтролера в режимі он-лайн, використовуючи поточну помилку відхилення положення об'єкта від уставки e (k) = r (k) y (k). Схема підключення нейроконтролера така сама, як і методі узагальненого інверсного нейроуправления. На вхід мережі подається вектор x(k):

Нейронна мережа генерує керуючий вектор u(k), який переводить об'єкт управління положення y(k + 1). Далі обчислюється поточна помилка роботи нейроконтролера e (k) = r (k + 1) y (k + 1)

Обчислюється градієнт зміни ваг

Потім проводиться корекція ваг нейроконтролера за методом якнайшвидшого спуску або будь-яким іншим градієнтним методом.

Похідна є якобіан об'єкта управління, значення якого задається аналітично за заданою математичною моделлю об'єкта управління. Однак, на практиці, для отримання прийнятної якості управління часто досить обчислити лише знак якобіана. Ітерації корекції значень коефіцієнтів продовжуються до досягнення прийнятної якості управління.

Метод зворотного пропуску помилки через прямий нейроемулятор

Метод зворотного пропуску помилки через прямий нейроемулятор: ліворуч – схема навчання прямого нейроемулятора; справа – схема навчання нейроконтролера

Метод зворотного пропуску помилки через прямий нейроемулятор (Backpropagation Through Time, Model Reference Adaptive Control, Internal Model Control) [20] [21] [22] [8] заснований на ідеї застосування тандему з двох нейронних мереж, одна з яких виконує функцію управління контролером, а . Прямий нейроемулятор служить для обчислення градієнта помилки нейроконтролера в процесі навчання і далі не використовується. Можна сказати, що нейроконтролер і нейроемулятор є єдиною нейромережею, при цьому, при навчанні нейроконтролера ваги прямого нейроемулятора "заморожуються". Прямий нейроемулятор навчається першим. Для цього, на вхід об'єкта управління подається випадковий сигнал u, що змінює положення об'єкта управління y, і формується навчальна вибірка:

Навчання прямого нейроемулятора виконується в режимі оф-лайн. Прямий нейроемулятор вважається навченим, якщо при однакових значеннях на входах нейроемулятора та реального об'єкта, відмінність між значеннями їх виходів стає незначною. Після закінчення навчання прямого нейроемулятора проводиться навчання нейроконтролера. Навчання виконується в режимі он-лайн за такою самою схемою, як і у разі спеціалізованого інверсного нейроуправління. Спочатку (на такті k) на вхід нейроконтролера надходить бажане положення об'єкта управління наступного такту r(k + 1). Нейроконтролер генерує сигнал керування u(k), який надходить на входи об'єкта керування та нейроемулятора. В результаті керований об'єкт переходить у положення y(k + 1), а нейроемулятор генерує реакцію . Далі обчислюється помилка управління e(k) = r(k + 1) y(k + 1) і пропускається у зворотному напрямку за правилом зворотного розповсюдження. Вагові коефіцієнти зв'язків нейроемулятора при цьому не коригуються. Механізм зворотного проходження помилки через прямий нейроемулятор реалізує локальну інверсну модель у точці простору станів об'єкта управління. Пройшовши через нейроемулятор, помилка далі поширюється через нейроконтролер, але тепер її проходження супроводжується корекцією вагових коефіцієнтів нейроконтролера. Прямий нейроемулятор виконує функції додаткових шарів нейронної мережі нейроконтролера, у яких ваги зв'язків не коригуються.

Метод нейроуправління з еталонною моделлю

Нейроуправління з еталонною моделлю

Метод нейроуправління з еталонною моделлю (Model Reference Adaptive Control, Neural Adaptive Control) [22] [23] - варіант нейроуправління за методом зворотного пропуску помилки через прямий нейроемулятор, з додатково впровадженою в схему еталонною моделлю (Re. Це робиться з метою підвищення якості перехідного процесу: у разі коли перехід об'єкта в цільове положення за один такт неможливий, траєкторія руху і час здійснення перехідного процесу стають погано прогнозованими величинами і можуть призвести до нестійкості перехідного процесу. Для зменшення цієї невизначеності, вводиться еталонна модель, що є, як правило, стійкою лінійною динамічною системою першого або другого порядку. У ході навчання, еталонна модель на такті k отримує на вхід уставку r(k + 1) і генерує опорну траєкторію y'(k + 1), яка порівнюється зі становищем об'єкта управління y(k + 1) з метою отримати помилку управління e(k + 1), мінімізувати яку навчається нейроконтролер.

Метод нейромережевої фільтрації зовнішніх збурень

Схема методу нейромережевої фільтрації зовнішніх збурень

Метод нейромережевої фільтрації зовнішніх збурень (Adaptive Inverse Control - Linear and Nonlinear Adaptive Filtering, Internal Model Control) [24] служить для поліпшення якості роботи

контролера в ланцюзі управління. Спочатку ця схема була запропонована Б. Уідроу для використання спільно з нейроконтролерами, навченими за методом узагальненого інверсного нейроуправління [25]. У пізнішій роботі [26] ним були застосовані нейроконтролери, навчені методом зворотного поширення помилки через прямий нейроэмулятор. В принципі, нейромережеву фільтрацію помилок можна використовувати для підвищення якості роботи контролера будь-якого типу, не обов'язково нейромережевого. У цій схемі використовується дві попередньо навчені нейронні мережі: інверсний нейроемулятор, навчений так само, як це робиться в методі узагальненого інверсного нейроуправління і прямий нейроемулятор, навчений так само, як це робиться в методі зворотного поширення помилки через прямий нейроемулятор. Нехай на об'єкт управління надходить керуючий сигнал, що явився результатом підсумовування сигналу контролера і коригувального сигналу системи фільтрації зовнішніх збурень, обчисленого на попередньому такті. Сигнал спрямовується на прямий нейроемулятор об'єкта управління, а реакція прямого нейроемулятора порівнюється реальним положенням системи y(k). Різниця цих величин e(k) сприймається як небажане відхилення системи, викликане зовнішнім обуренням. Для придушення небажаного ефекту сигнал надходить на інверсний нейроемулятор, який розраховує коригуючий сигнал для коригування керуючого сигналу нейроконтролера на наступному такті. Для використання цього методу, об'єкт управління повинен мати обертову динаміку, а також необхідно мати адекватну математичну або імітаційну модель об'єкта управління для навчання прямого та інверсного нейроемуляторів.
Схема прогнозуючого модельного нейроуправління
(NN
Predictive Control, Model Predictive Control, Neural Generalized Predictive Control) [27] [28] мінімізує функціонал вартості інтегральної помилки Q(k), прогнозованої на L загальний функціонал вартості Q(k). Для прогнозування майбутньої поведінки системи та обчислення помилок використовується прямий нейроемулятор, навчений так само, як у методі зворотного поширення помилки через прямий нейроемулятор. Примітність розглянутого методу полягає в тому, що в ньому відсутня нейроконтролер, що навчається. Його місце займає оптимізаційний модуль, що працює в режимі реального часу, в якому може бути використаний, наприклад, сімплекс-метод [29] або квазі-Ньютонівський алгоритм [30].

Оптимізаційний модуль отримує на такті цільову траєкторію на L тактів вперед, а якщо її немає, то L раз дублює значення поточної уставки r(k + 1) і використовує це як цільову траєкторію. Далі, для вибору оптимального впливу, що управляє, обчислення відбуваються у внутрішньому циклі системи нейроуправління (його ітерації позначаються як j). За час одного такту управління оптимізаційний модуль подає на вхід нейроемулятора серію різних впливів , де t - глибина прогнозування, отримує різні варіанти поведінки системи, обчислює для них функцію вартості Q (k) і визначає найкращу стратегію управління . У результаті на об'єкт подається керуючий сигнал . На такому такті стратегія ST перераховується заново.
Адаптивні критики
Схема адаптивної критики: ліворуч – етап управління; справа – етап навчання Методи нейроуправління з урахуванням **адаптивної критики** (Adaptive Critics), які також відомі як наближене динамічне програмування (Approximated Dynamic Programming, ADP), останніми роками дуже популярні [31] [32] [33] [34]. Системи адаптивної критики вибирають керуючий сигнал з урахуванням мінімізації функціоналу оцінок помилок майбутнього з нескінченним горизонтом:

Тут – коефіцієнт забування, ,e(k) = r(k + 1) y(k + 1) – відхилення траєкторії об'єкта управління від уставки, обчислюване кожному такті роботи системи. Система включає два нейронні модулі: нейроконтролер і модуль критики (критик). Модуль критики виконує апроксимацію значень функціоналу вартості J(k), нейроконтроллер навчають мінімізувати функціонал вартості J(k).

У режимі управління об'єктом на вхід нейроконтролера надходить вектор , що викликає появу на його виході сигналу управління u(k), в результаті чого об'єкт управління переходить у положення y(k + 1). Далі проводиться обчислення значення поточної помилки керування e(k). Модуль критики, отримуючи на вході вектор, оцінює функції вартості J(k). На наступному такті процес

повторюється: обчислюються нові значення e(k+1) та J(k+1). Навчання системи нейроуправління відбувається в режимі он-лайн і складається з двох етапів: навчання модуля критики та навчання нейроконтролера. Спочатку розраховується помилка тимчасової різниці w(k) = e(k) + J(k + 1) J(k). Потім методом якнайшвидшого спуску виконується корекція ваги зв'язків для модуля критики wcritic:

Значення градієнта розраховується методом зворотного поширення помилки. Корекція ваги зв'язків нейроконтролера wcontrol проводиться аналогічно: Значення похідної знаходять шляхом зворотного поширення величини через модуль критики, а значення градієнта шляхом зворотного поширення помилки через модуль контролера. Корекція ваги продовжується, поки система не досягне необхідного рівня якості управління. Таким чином, на кожному кроці покращується закон управління шляхом навчання нейроконтролера (ітерація за стратегіями, Policy Iteration), а також підвищується здатність системи оцінювати ситуацію шляхом навчання критика (ітерація за значеннями, Value Iteration). Конкретна схема побудови системи адаптивної критики може відрізнятися від вищеописаної, що зветься евристичне динамічне програмування (Heuristic Dynamic Programming, HDP). У методі дуального евристичного програмування (Dual Heuristic Programming, DHP) модуль критики обчислює похідну функціонала глобальної вартості, а методі глобального дуального евристичного програмування (Global Dual Heuristic Programming, GHDP) критиком обчислюються як сам функціонал функції вартості J, і його похідна . Відомі модифікації методу, в яких модуль критики приймає рішення виключно на основі сигналу, що управляє. Їхні англомовні абревіатури мають приставку AD (Action Dependent): ADHDP, ADDHP, ADGDHP. У деяких версіях адаптивної критики модуль критики складається з двох частин: власне, модуля критики та прямого нейроемулятора. Останній видає передбачення поведінка об'єкта управління, на основі яких критик формує оцінку функції вартості J. Такі версії звуться засновані на моделі (model based).

Гібридне нейро-ПІД управління

Схема гібридного нейро-ПІД управління

Гібридне нейро-ПІД управління (NNPID Auto-tuning, Neuromorphic PID Self-tuning) [35] [36] дозволяє здійснювати самоналаштування ПІД-регулятора в режимі он-лайн шляхом використання нейронних мереж. Налаштування ПІД-регулятора виконується в режимі он-лайн, за помилкою керування e(k) = r(k + 1) y(k + 1). На такті k нейронна мережа отримує уставку r(k + 1) і генерує коефіцієнти управління ПІД-контролера K1 (пропорційний), K2 (інтегральний), K3 (диференціальний), які надходять на ПІД-контролер разом із значенням поточної помилки зворотного зв'язку e(k). У ході роботи, ПІД-контролер розраховує поточний керуючий сигнал u(k) за формулою:

u(k) = u(k 1) + K1(k)(e(k) e(k 1)) + K2(k) e(k) + K3(k)(e(k) 2e(k мен) на контролі)

Навчання нейромережі відбувається в режимі реального часу помилково зворотного зв'язку, методом якнайшвидшого спуску.

Тут - вектор виходів нейронної мережі, що надходить на ПІД-контролер.

Градієнти обчислюють шляхом зворотного поширення помилки. Якобіан об'єкт управління чи його знак знаходиться аналітично, на основі математичної моделі об'єкта управління.

Гібридне паралельне нейроуправління

Схема гібридного паралельного нейроуправління Методи гібридного **паралельного** нейроуправління (Parallel Neurocontrol, Stable Direct Adaptive Control, Additive Feedforward Control) [24] [27] передбачають паралельне використання нейроконтролерів та звичайних контролерів для управління динаміками. При цьому нейроконтролер і звичайний контролер, у ролі якого виступає, наприклад, ПІД-контролер, набувають однакових значень уставки. Можливі такі варіанти спільного підключення нормального контролера і нейроконтролера:

до об'єкта управління підключається стандартний контролер, після чого нейроконтролер навчається керувати вже замкнутою стандартним контролером системою. Після навчання нейроконтролера він підключається до системи, а керуючі сигнали обох контролерів

підсумовуються;

нейроконтроллер навчається керувати об'єктом управління, після навчання починає функціонувати у штатному режимі. Далі, керувати замкнутою нейроконтроллером системою налаштовується звичайний контролер. Після налаштування звичайного контролера він підключається до системи, керуючий сигнал обох контролерів підсумовується;

області дії звичайного контролера та нейроконтролера розмежовуються. Наприклад, у просторі станів об'єкта управління для нейроконтролера виділяється окрема область LS:

При цьому звичайний контролер розраховується на управління об'єктом поза цією областю простору стану. При паралельній роботі обох контролерів, керуючий сигнал надходить на об'єкт або від нейроконтролера, якщо поточний стан системи знаходиться в межах LS області, або, в іншому випадку, від звичайного контролера. Гібридне паралельне нейроуправління представляє компромісне рішення для впровадження нейроуправління в промисловість та переходу від звичайних контролерів до нейромережевих.
^

Вороновський Г. К., Генетичні алгоритми, штучні нейронні мережі, 1997

^ Werbos, PJ Backpropagation and neurocontrol: review and prospectus // International Joint Conference on Neural Networks, Vol **.** 1. - P. 209-216. - Washington, DC , USA, 18-22 Jun 1989

^ Gundy-Burlet K., Krishnakumar K., Limes G., Bryant D. Усунення з Intelligent Flight Control System для Simulated C-17 Aircraft // J. of Aerospace Computing, Information, and Communication. - 2004. - Vol. 1, N 12. - P. 526 - 542

^ Кондратьєв А.І., Тюменцев Ю.В. **Нейросітельне** адаптивне відмовостійке управління рухом маневреного літака // XII Всеросійська науково-технічна конференція "Нейроінформатика - 2010": Частина 2. - М.: НІЯУ МІФІ, 2010. - С. 262 - 273

. Нейрокомп'ютери в управлінні гелікоптерами // Штучний інтелект. - 2000. - № 3. - С. 290 - 298

^ D. Gu and H. Hu. Neural Predictive Control для Car-like Mobile Robot // International Journal of Robotics and Autonomous Systems, Vol. 39, No. 2-3, May, 2002

[Терехов В.А., Єфімов Д.В., Тюкін І.Ю. Нейросетельні системи управління: Навч. посібник для вузів. - М.: Вищ. школа 2002. - 183 с.]

^ **1 2** Danil V. Prokhorov. Toyota Prius HEV Neurocontrol and Diagnostics // Neural Networks. - 2008. - No. 21. - P. 458 - 465

^ Dias FM, Mota AM Comparison між Diferent Control Strategies using Neural Networks // 9th Mediterranean Conference on Control and Automation. – Dubrovnik, Croatia, 2001

^ Venayagamoorthy GK, Harley RG, Wunsch DC Implementation Adaptive Critic-based Neurocontrollers для Turbogenerators в Multimachine Power System, IEEE Transactions on Neural Networks. - 2003. - Vol. 14, Issue 5. – P. 1047 – 1064.

^ D'Emilia G., Marrab A., Natalea E. За допомогою neural networks for quick and accurate auto-tuning of PID controller // Robotics and Computer-Integrated Manufacturing. - 2007. - Vol. 23. – P. 170 – 179.

Змеу К.В., Марков Н.А., Шипітько І.А., Ноткін Б.С. Безмодельне прогнозуюче інверсне нейроуправління з еталонним перехідним процесом, що регенерується // Інтелектуальні системи. – 2009. – № 3. – С. 109 – 117.

^ Кузнєцов Б.І., Василець Т.Є., Варфоломєєв А.А. Синтез нейроконтролера з передбаченням для двомасової електромеханічної системи // Електротехніка та електромеханіка. - 2008. - Т. 3. - С. 27 - 32.

^ Д.А. Дзюба, О.М. Чорнодоб. Застосування методу контрольованого обурення для модифікації нейроконтролерів у реальному часі / / Математичні Машини та Системи. - 2010. - № 4. - С. 20 - 28.

[Widrow B., Smith FW Pattern-recognizing control systems // Proceedings of Computer and Information Sciences. - Washington, USA - 1964. - Vol. 12. - P. 288 - 317.]

^ Omidvar O., Elliott DL eds. Neural Systems for Control // Academic Press, New York, 1997. - 358 з .

^ Ronco E. Incremental Polynomial Controller Networks: Two Self-Organising Non-Linear Controllers // Ph.D. Disseration Thesis, Glasgow, 1997. - 207 p.

^ **1 2** [Омату С., Халід М., Юсоф Р. Нейроуправління та його додатки: пров. з англ. - М.: ІПРЖР, 2000. - 272 с.]

^ **1 2** Psaltis D., Sideris A., Yamamura AA Multilayered Neural Network Controller // IEEE Control Systems Magazine – 1988. – Vol. 8, Issue 2. - P. 17 - 21.

^ Werbos P. Backpropagation через час: що це робить і як зробити це // Proceedings of the IEEE. - October 1990. - Vol. 78, N. 10. – P. 1550 – 1560

^ [Jordan MI and Rumelhart DE Forwardmodels: Supervised learning with distal teacher // Cognitive Science – 1990. – Vol. 16. - P. 313 - 355.]

^ **1 2** [Narendra KS, Parthasarathy KK Identification and control of dynamical systems using neural networks // IEEE Transactions on Neural Networks. - 1990. - N 1. - P. 4 - 27.]

^ Venelinov Topalov, A. Kaynak. Online навчання в adaptive neurocontrol schemes with sliding mode algorithm // IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. – 2001. – V. 31. – I. 3. – P. 445 – 450

^ **1 2** Dias FM, Mota AM Comparison between Different Control Strategies using Neural Networks // 9th Mediterranean Conference on Control and Automation. – Dubrovnik, Croatia, 2001.

^ Widrow B., Adaptive Inverse Control // Процедури 2nd IFAC Workshop на Adaptive Systems in Control and Signal Processing – Lund, Sweden, July 1986. – P. 1 – 5.

^ Widrow B., Inline GL Процедури міжнародних робіт на Neural Networks для Identification, Control, Robotics, і Signal/Image Processing – 21 23 Aug 1996, Venice, Italy. – P. 30 – 38.

^ **1 2** Hagan MT, Demuth HB Neural мережі для контролю // Процеси з American Control Conference. - San Diego, USA, 1999. - Vol. 3. – P. 1642 – 1656.

^ [Rossiter JA Model-based Predictive Control: a Practical Approach // CRC Press, 2003. – 318 с.]

^ [Takahashi Y. , Japan, 25 - 29 October, 1993. - Vol. 3. – P. 1464 – 1468.]

^ Soloway D., Haley PJ Neural Generalized Predictive Control // Proceedings of IEEE International Symposium on Intelligent Control. – 15 – 18 September 1996. – P. 277 – 281.

^ Prokhorov D. and Wunsch D. Adaptive Critic Designs // IEEE Transactions on Neural Networks. - 1997. - Vol. 8, N 5. – P. 997 – 1007.

^ Venayagamoorthy GK, Harley RG, Wunsch DC. - 2003. - Vol. 14, Issue 5. - P. 1047 – 1064.

^ Ferrari S., Stengel RF Model-Based Adaptive Critic Designs // Learning and Approximated Dynamic Programming, J. Si, A. Barto, W. Powell, і D. Wunsch, Eds. New York: Wiley, 2004, Chapter. 3

^ Редько В.Г., Прохоров Д.В. Нейросетевые адаптивні критики // VI Всеросійська науково-технічна конференція “Нейроінформатика-2004". Збірник наукових праць . Частина 2. М .: МІФІ , 2004. - С. 77 - 84.

^ [D'Emilia G., Marrab A., Natalea E. Use neural networks for quick and accurate auto-tuning of PID controller // Robotics and Computer-Integrated Manufacturing. - 2007. - Vol. 23. – P. 170 – 179.]

^ [Saiful A., Omatu S. Neuromorphic self-tuning PID controller // Proceedings of IEEE International Conference on Neural Networks, San Francisco, USA, 1993. – P. 552 – 557]

Посилання

А.Н. Чорнодуб, Д.А. Дзюба. Огляд методів нейроуправління// Проблеми програмування. - 2011. - No 2. - С. 79-94.

Hagan MT, Demuth HB Neural мережі для контролю // Процедури американського контролю конференцій. - San Diego, USA, 1999. - Vol. 3. - P. 1642 - 1656.

Література

Сігеру Омату , Марзуки Халід , Рубія Юсоф Нейроуправління і його програми = Neuro-Control and its Applications - 2- е . - М .: ІПРЖР , 2000. - С. 272. - ISBN ISBN 5-93108-006-6.

В. А. Терехов, Д. В. Єфімов, І. Ю. Тюкін Нейромережні системи управління - 1-е. - Вища школа, 2002. - С. 184. - ISBN 5-06-004094-1.

Саймон Хайкін Нейронні мережі: повний курс = Neural Networks: A Comprehensive Foundation - 2-ге. - М.: "Вільямс", 2006. - С. 1104. - ISBN 0-13-273350-1.

Omidvar O., Elliott DL eds. Neural Systems for Control - New York: Academic Press, 1997. - З . 358. - ISBN 0-12-526430-5.

Категорії:

Інтелектуальна робототехніка

Теорія управління

Штучні нейронні мережі