

Навчальне видання

*Захарченко Микола Васильович
Горохов Сергій Михайлович
Балан Микола Макарович
Гаджієв Матін Магсуд-огли
Корчинський Володимир Вікторович
Ложковський Анатолій Григорович*

**МАТЕМАТИЧНІ ОСНОВИ ОПТИМІЗАЦІЇ
ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМ**

Підручник

Редактор Кодрул Л.А.
Комп'ютерна верстка Гардиман Ж.А.

Здано в набір 16.07.2010 Підписано до друку 31.08.2010
Формат 60x90/16 Зам. № 42
Тираж 300 прим. Обсяг 15,0 друк. арк.
Віддруковано на видавничому устаткуванні фірми RISO
у друкарні редакційно-видавничого центру ОНАЗ ім. О.С. Попова
м. Одеса, вул. Старопортофранківська, 61
Тел. 720-78-94
ОНАЗ, 2010

Міністерство освіти і науки України
Одеська Національна академія зв'язку ім. О.С. Попова

**Захарченко М.В., Горохов С.М., Балан М.М.,
Гаджієв М.М., Корчинський В.В., Ложковський А.Г.**

**Математичні основи оптимізації
телекомунікаційних систем**

За загальною редакцією академіка МАІ
д.т.н., проф. Захарченка М.В

**Затверджено Міністерством освіти і науки України
як підручник для студентів вищих навчальних закладів
(лист № 14/18-Г-784 від 25.03.2010)**

**Одеса
ОНАЗ ім. О.С. Попова
2010**

УДК 621.391:517.977
ББК 32.967.9
М-34

Рецензенти:

М.М. Климаш, д-т. техн. наук, професор (Національний університет «Львівська політехніка»);

М.Б. Копитчук, д-т. техн. наук, професор (Одеський національний політехнічний університет).

Математичні основи оптимізації телекомунікаційних систем: підручник. За загальною редакцією Захарченко М.В / Захарченко М.В., Горохов С.М., Балан М.М., Гаджієв М.М., Корчинський В.В., Ложковський А.Г. – Одеса: ОНАЗ ім. О.С. Попова, 2010. – 240 с.

ISBN 978-966-7598-53-2

У підручнику систематизовано викладено математичні основи методів визначення екстремумів функцій однієї та багатьох змінних, основи варіаційного числення та теорії графів. Розглянуті питання пошуку оптимального розв'язання задач, пов'язаних з ефективністю систем передавання інформації, корекцією лінійних спотворень каналів зв'язку, застосування оптимізації у задачах синтезу цифрових фільтрів. Наведені основні поняття теорії графів та розглянуті обчислювальні алгоритми пошуку оптимальних шляхів і максимального потоку у мережах зв'язку.

Надані теоретичні основи оптимізаційних підходів у системах зв'язку дозволяють забезпечити аналіз і синтез алгоритмів обробки сигналів у пристроях перетворення сигналів та складають основу математичних моделей об'єктів, які проектуються.

Підручник призначений для студентів вищих навчальних закладів для всіх спеціальностей за напрямками 6.050.901 «Радіотехніка», 6.050.902 «Радіотехніка, радіоелектронні апарати та зв'язок», 6.050.903 «Телекомунікації», 6.050.904 «Мережі та системи поштового зв'язку», інженерно-технічних працівників, які розробляють телекомунікаційні системи і займаються експлуатацією систем передачі даних і мереж зв'язку.

ББК 32.967.9

ISBN 978-966-7598-53-2

© Захарченко М.В., Горохов С.М., Балан М.М.,
Гаджієв М.М., Корчинський В.В., Ложковський А.Г., 2010

24 Князева Н.А. Алгоритмы поиска максимального потока в сети с ограничениями на ветвях и узлах / Н.А. Князева // Сети, узлы и распределение информации: труды учебных институтов связи. – Л.: ЛЭИС. – Вып. 27. – 1980. – С.8 – 13.

25 Михлин С.Г. Вариационные методы в математической физике / С.Г. Михлин – М.: Наука, 1970. – 512 с.

4 Тамм Ю.А. Адаптивная коррекция сигнала ПД / Ю.А. Тамм – М.: Связь. 1978.

5 Захарченко Н.В., Нудельман П.Я. Выбор узлов аппаратуры передачи данных: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман – Одесса: ОЭИС им. А.С. Попова, 1985. – Ч. 1.– 88 с.

6 Нудельман П. Я. Полиномные синтезаторы временных и частотных характеристик / П. Я. Нудельман – М.: Связь, 1975.

7 Нудельман П.Я. Обработка сигналов в линейных аналоговых и цифровых цепях: метод. руководство для научно-исследовательских работ студентов / П.Я. Нудельман – Одесса: ОЭИС им. А.С. Попова, 1987. – Ч. 1.

8 Френкс Л. Теория сигналов / Френкс Л. – М.: Советское радио, 1974.– 344 с.

9 Рабинер Л. Теория и применение цифровой обработки сигналов: монография / Рабинер Л., Гоулд Б. Пер. с англ. – М.: Мир. 1978.– 848 с.

10 Гольденберг Л.М. Цифровая обработка сигналов: справочник / Л.М. Гольденберг, Б.Д. Матюшкин, М.Н. Поляк – М.: Радио и связь, 1985.

11 Lucky R.W. Principles of data communication / R.W. Lucky, J. Salz, E.J. Weldon // Bell Telephone Laboratories. – MCORAW – HILL. – 1968.

12 Falconer D.D., Magee F.R. Evaluation of decision feedback-equalization and Viterbi algorithm detection for voiceband data transmission / D.D. Falconer, F.R. Magee // IEEE. Trans. of Commun. – 1976. – V. COM-24. – № 10. – Part 1, 2.

13 Kawas-Kaleh O. Double decision feedback equalizer / Kawas-Kaleh O. // Frequenc.– 1979. – V. 33, № 5.

14 Захарченко Н.В. Выбор узлов аппаратуры передачи данных: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман – Одесса: ОЭИС им. А.С. Попова, 1986. – Ч. 2.– 87 с.

15 Зюко А.Г. Теория передачи сигналов / А.Г. Зюко, Ю.Ф. Коробов – М.: Связь, 1972. - 279 с.

16 Финк Л.М. Теория передачи дискретных сообщений / Л.М. Финк – М.: Сов. радио, 1970. - 726 с.

17 Кисель В.Л. Синтез гармонических корректоров для высокоскоростных систем связи / В.Л. Кисель – М.: Связь, 1979.

18 Филипс Л. Методы анализа сетей / Л. Филипс, А. Гарсиа – Диас – М.: Мир, 1984. – 496 с.

19 Форд Л. Потоки в сетях / Л. Форд, Д. Фалкерсон – М.: Мир. – 1966. - 276 с.

20 Лазарев В.Г. Динамическое управление потоками информации в сетях связи / В.Г. Лазарев, Ю.В. Лазарев // – Радио и связь, 1983. – 216 с.

21 Теория сетей связи; под ред. В.Н. Рогинского. – М.: Радио и связь, 1981. – 323 с.

22 Пападимитриу Х. Комбинаторная оптимизация. Алгоритмы и сложность / Х. Пападимитриу, К. Стайглиц – М.: Мир, 1985. – 512 с.

23 Астафьева Н.А. Об одном алгоритме поиска кратчайших путей в сети связи / Н.А. Астафьева, Л.Е. Яшук // Труды учебных институтов. – Л.: ЛЭИС. – Вып. 57. – 1972. – С. 124-129.

ЗМІСТ

ВСТУП.....	6
1 ФУНКЦІЇ ОДНІЄЇ ЗМІННОЇ.....	9
1.1 Властивості функцій однієї змінної.....	9
1.2 Монотонні й унімодальні функції	11
1.3 Критерії оптимальності функції однієї змінної.....	12
1.4 Ідентифікація екстремуму функції однієї змінної.....	13
Контрольні запитання та задачі до розділу 1.....	17
Література до розділу 1.....	18
2 МЕТОДИ ПОШУКУ ЕКСТРЕМУМУ УСЕРЕДИНІ ЗАДАНОГО ІНТЕРВАЛУ.....	19
2.1 Визначення границь інтервалу.....	19
2.2 Етап зменшення інтервалу.....	20
2.3 Метод вилучення інтервалів.....	20
2.4 Метод розподілу інтервалу навпіл.....	21
2.5 Пошук за допомогою методу золотого перетину.....	22
2.6 Порівняння методів вилучення інтервалів.....	24
Контрольні запитання та задачі до розділу 2.....	26
Література до розділу 2.....	27
3 МЕТОДИ ПОШУКУ ЕКСТРЕМУМУ З ВИКОРИСТАННЯМ ПОХІДНИХ.....	28
3.1 Метод Ньютона-Рафсона.....	28
3.2 Метод середньої точки.....	29
3.3 Метод січних.....	31
Контрольні запитання та задачі до розділу 3.....	32
Література до розділу 3.....	32
4 МЕТОДИ ПОШУКУ ЕКСТРЕМУМУ ФУНКЦІЙ БАГАТЬОХ ЗМІННИХ.....	33
4.1 Необхідні умови екстремуму. Аналітичний метод знаходження екстремуму.....	33
4.2 Достатні умови екстремуму.....	35
4.3 Застосування необхідних і достатніх умов екстремуму у задачі корекції спотворень каналу зв'язку.....	36
4.4 Ітераційний підхід. Унімодальність і опуклість.....	38
4.5 Метод перетинів.....	40
Контрольні запитання та задачі до розділу 4.....	44
Література до розділу 4.....	45
5 ГРАДІЄНТНІ МЕТОДИ ПОШУКУ ЕКСТРЕМУМУ.....	46
5.1 Вектор-градієнт і його властивості	46
5.2 Метод найшвидшого підйому (спуску).....	49
5.3 Спрощені градієнтні процедури.....	51
5.4 Метод сполучених градієнтів.....	53
Контрольні запитання та задачі до розділу 5.....	55
Література до розділу 5.....	56

6. ОСНОВИ ВАРІАЦІЙНОГО ЧИСЛЕННЯ.....	57
6.1 Класична задача про брахістохрону і поняття функціонала.....	57
6.2 Варіація функціонала і його властивості.....	61
6.3 Рівняння Ейлера.....	65
6.4 Задачі з рухливими кінцями і задачі на умовний екстремум.....	74
6.5 Достатні умови екстремуму.....	80
Контрольні запитання та задачі до розділу 6.....	82
Література до розділу 6.....	83
7 ОСНОВИ ТЕОРІЇ ГРАФІВ.....	84
7.1 Основні поняття теорії графів.....	84
7.2 Модель мережі зв'язку.....	88
7.3 Матричне подання графів.....	89
Контрольні запитання та задачі до розділу 7.....	91
Література до розділу 7.....	93
8 АНАЛІЗ, СИНТЕЗ І ОПТИМІЗАЦІЯ МЕРЕЖ ЗВ'ЯЗКУ.....	94
8.1 Задачі аналізу, синтезу і оптимізації мереж зв'язку.....	94
8.2 Моделі систем розподілу інформації.....	96
8.3 Математична модель потоку вимог.....	99
8.4 Навантаження та його види.....	103
8.5 Характеристики якості обслуговування.....	104
8.6 Аналіз систем в умовах пуассонівського потоку вимог.....	107
8.6.1 Система із втратами $M/M/m$	109
8.6.2 Система із необмеженою чергою $M/M/m/\infty$	112
8.6.3 Система із обмеженою чергою $M/M/m/r$	117
8.6.4 Система із необмеженою чергою $M/D/m/\infty$	119
8.6.5 Система із необмеженою чергою $M/G/1/\infty$	122
8.7 Аналіз систем в умовах реального потоку вимог.....	127
8.7.1 Функція розподілу станів системи із втратами $H/D/m$	128
8.7.2 Імовірність втрат системи $H/D/m$	130
8.7.3 Система із необмеженою чергою $H/D/m/\infty$	132
8.7.4 Система із необмеженою чергою $fBM/D/1/\infty$	134
8.7.5 Система із необмеженою чергою $G/M/1/\infty$	137
8.7.6 Система із необмеженою чергою $G/D/1/\infty$	140
Контрольні запитання та задачі до розділу 8.....	142
Література до розділу 8.....	144
9 ПРИКЛАДИ РОЗВ'ЯЗАННЯ ЗАДАЧ З ОПТИМІЗАЦІЇ.....	145
9.1 Приклади розв'язання задач пошуку точки екстремуму усередині заданого інтервалу.....	145
9.1.1 Визначення границь інтервалу.....	145
9.1.2 Метод розподілу інтервалу навпіл.....	145
9.1.3 Метод золотого перетину.....	146
9.2 Приклади розв'язання задач з використанням похідних.....	147
9.2.1 Приклад розв'язання задач з використанням методу Ньютона-Рафсона.....	147
9.2.2 Приклад розв'язання задач з використанням методу січних.....	147

$S(00110) = 8$; заміна не відбувається;
 $S(00111) = 8$; заміна не відбувається;
 $S(01000) = 4$; $V_{21} = 4$; $V_{23} = 4$; $V_{24} = 4$; $V_{25} = 4$;
 $S(01001) = 5$; заміна не відбувається;
 $S(01010) = 11$; заміна не відбувається;
 $S(01011) = 2$; $V_{21} = 2$; $V_{23} = 2$; $V_{41} = 2$; $V_{43} = 2$; $V_{51} = 2$; $V_{53} = 2$;
 $S(01100) = 8$; заміна не відбувається.

Далі покажемо лише ті перерізи, які призводять до зміни міжполюсних потоків:

$S(01111) = 4$; $V_{31} = 4$;
 $S(10000) = 4$; $V_{12} = 4$; $V_{13} = 4$; $V_{14} = 4$; $V_{15} = 4$.
 $S(10111) = 4$; $V_{32} = 4$; $V_{42} = 4$; $V_{52} = 4$;
 $S(11011) = 2$; $V_{13} = 2$.

В результаті отримана матриця V максимальних потоків між усіма вершинами графа

$$V = \begin{bmatrix} \infty & 4 & 2 & 4 & 4 \\ 2 & \infty & 2 & 4 & 4 \\ 4 & 4 & \infty & 7 & 7 \\ 2 & 4 & 2 & \infty & 5 \\ 2 & 4 & 2 & 5 & \infty \end{bmatrix}. \quad (9.151)$$

Елементи головної діагоналі матриці позначені символом " ∞ ", однак практичного значення максимальний потік V_{ss} не має, тому замість " ∞ " можна наводити будь-які інші символи, наприклад, "-".

Розглянуті алгоритми, завдяки їх великій обчислювальній ефективності, можуть використовуватися при синтезі та аналізі мереж зв'язку великої розмірності.

Достоїнства мережних моделей обумовлюють їхнє широке застосування при проектуванні та удосконаленні великих та складних систем, пристроїв та мереж зв'язку.

Література до розділу 9

- 1 Балан М.М. Методи пошуку точки екстремуму всередині заданого інтервалу: метод. вказівки до проведення практичного заняття з дисципліни "Методи оптимізації" / М.М. Балан – Одеса: ОНАЗ ім. О.С. Попова, 2004. – 16 с.
- 2 Реклейтис Г. Оптимізація в техніці: В 2-х кн. – Кн.1 / Реклейтис Г., Рейвиндран А., Рэгдел К. – М.: Мир, 1986. – 352 с.
- 3 Захарченко Н.В. Оптимізація и моделирование систем связи: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, 1988. – Ч. 1.– 86 с.

Проілюструємо роботу описаного алгоритму на прикладі мережі, показаної на рис. 9.46.

Оскільки в графі є п'ять вершин, розрядність коду перерізу дорівнює також п'яти. Будемо нумерувати розряди коду зліва направо. Таким чином, перший розряд відповідає першій вершині, другий розряд – другій вершині і т. ін. Нагадаємо, що одиниці в коді перерізу відповідають витокам, нулі – стокам.

Тоді переріз 01010 розділяє такі потоки:

$$V_{21}, V_{23}, V_{25}, V_{41}, V_{43}, V_{45}.$$

При перевірці пропускних здатностей перерізів S , вписаних в порядку послідовного збільшення їх двійкових кодів, отримуємо такі значення міжполюсних потоків V_{st} :

$S(00000)$ – цей переріз не розглядається;

$S(00001)$ – в переріз входять ребра $(5, 1), (5, 2), (5, 3), (5, 4)$;

$$\text{величина перерізу } S(00001) = c(5, 1) + c(5, 2) + c(5, 3) + c(5, 4) = 0 + 0 + 0 + 5 = 5.$$

Переріз розділяє міжполюсні потоки з тими самими індексами, що й індекси ребер, які складають переріз, а саме: $V_{51}, V_{52}, V_{53}, V_{54}$. Через те, що їх значення (у початковій матриці V прийняті рівними нескінченності) більші величини знайденого перерізу $S(00001) = 5$, в матриці V здійснюється необхідна заміна

$$V = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ 5 & 5 & 5 & 5 & \infty \end{bmatrix}. \quad (9.149)$$

$S(00010)$ – в переріз входять ребра $(4, 1), (4, 2), (4, 3), (4, 5)$. Величина перерізу $S(00010) = c(4, 1) + c(4, 2) + c(4, 3) + c(4, 5) = 0 + 3 + 2 + 0 = 5$. Потоки, що розділяються перерізом – $V_{41}, V_{42}, V_{43}, V_{45}$.

Результат заміни величин потоків у матриці V

$$V = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ 5 & 5 & 5 & \infty & 5 \\ 5 & 5 & 5 & 5 & \infty \end{bmatrix}. \quad (9.150)$$

Будемо називати коди перерізів, їх величини й ті значення міжполюсних потоків, котрі замінюються в матриці V .

$S(00011) = 5$; заміна не відбувається;

$S(00100) = 7$; $V_{31} = 7$; $V_{32} = 7$; $V_{34} = 7$; $V_{35} = 7$;

$S(00101) = 12$; заміна не відбувається;

9.3 Приклади розв'язання екстремальних задач функцій багатьох змінних.....	148
9.3.1 Визначення критеріїв оптимальності функцій багатьох змінних.....	148
9.3.2 Пошук точки екстремуму функції двох змінних з використанням методу найшвидшого спуску	149
9.4 Корекція лінійних спотворень каналу зв'язку.....	151
9.4.1 Постановка задачі корекції.....	151
9.4.2 Середньоквадратична похибка коректування.....	153
9.4.3 Налаштування поліномних коректорів. Гармонійний коректор.....	158
9.4.4 Застосування оптимізації у задачах синтезу цифрових фільтрів.	161
9.4.5 Приклад розрахунку оптимальних значень коефіцієнтів гармонійного коректора методом покоординатного спуску.....	169
9.5 Задача максимізації пропускної спроможності каналу зв'язку.....	171
9.6 Екстремальні задачі, що пов'язані з поняттям ентропії.....	173
9.7 Задача пошуку імпульсу з мінімальною ефективною шириною спектру....	175
9.8 Лінійна корекція перехідних процесів.....	181
9.9 Методи пошуку шляхів у мережах.....	197
9.10 Алгоритми пошуку множини шляхів.....	197
9.11 Алгоритми пошуку найкоротших шляхів.....	204
9.12 Алгоритми пошуку максимального потоку.....	224
Література до розділу 9.....	238

ВСТУП

Характерною рисою рівня сучасної науки і техніки є широке застосування методів оптимізації при розв'язанні теоретичних і прикладних задач. Стосовно техніки зв'язку проблема оптимізації полягає не лише в організації зв'язку але і у організації його належним чином, наприклад, з максимально можливою в даних умовах швидкістю передавання інформації або з мінімальною ймовірністю помилки при заданій складності апаратури передачі, або з мінімальними габаритами, витратами, тощо.

Впродовж останніх десятиліть досягнуто значних успіхів у розробці нових варіантів модуляції і кодування при передаванні інформації, а також у високоефективній реалізації цих теоретичних результатів у розробці телекомунікаційної апаратури. Швидкість передавання інформації в сучасних передавальних пристроях значно зросла та потребує від розробників застосування в приймальному обладнанні оптимальних процедур подавлення «складних» завад, що утворюються сумісною дією гаусівського шуму, міжсимвольної завади, флуктуацій фази та низки інших чинників, що заважають.

У підручнику наводяться деякі, найбільш застосовані методи і прийоми відшукування екстремумів, зокрема таких, які спираються на численні процедури, що можуть бути реалізовані на ЕОМ. Підручник демонструє можливості методів оптимізації при проектуванні або адаптивному налаштуванні вузлів сучасної апаратури ТКС.

Існують, принаймні, три основні типи задач пошуку оптимальних рішень в теорії і практиці зв'язку:

1. Дослідницькі завдання, пов'язані з відшукуванням якнайкращих варіантів побудови пристроїв, систем і мереж зв'язку (наприклад, пошук алгоритмів оброблення сигналів, що максимізували відношення сигнал-шум, або пошук сигналів, що мінімізують займану смугу частот, та ін.).

2. Завдання проектування, в яких потрібно вибрати оптимальні параметри апаратури (наприклад, розрахувати фільтр так, щоб забезпечити задані вимоги до його вибіркової за мінімального порядку передавальної функції або вибрати розташування точок у сигнальному сузір'ї, за якого мінімальна з відстаней між точками максимальна, тощо).

3. Завдання побудови адаптивних пристроїв, які самооптимізуються та самоналагоджуються, а в умовах заводової ситуації, що змінюється, або навантаження, що змінюється, зуміють вибрати режим, який мінімізує втрати одного або іншого типу (налаштування коректора спотворень каналу у разі комутованої мережі зв'язку, маршрутизації пакетів у мережі зв'язку та ін.).

Дана класифікація є умовною і часто задачу, що має бути розв'язана, можна віднести відразу до двох або навіть до трьох зазначених вище типів. Проте характерним для всіх цих варіантів задач є те, що існує деяка залежність критерію якості роботи системи зв'язку від одного або декількох параметрів, що оптимізуються. Належним вибором аргументів слід забезпечити досягнення завданням критерієм максимального або мінімального значення. Ці задачі

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} \end{bmatrix}. \quad (9.147)$$

Перерізу, який розділяє, наприклад, вершини 1, 3, 5 і 2, 4, відповідають рядки 1, 3, 5 та стовпці 2 і 4 матриці C , на перерізі яких розміщені елементи $c_{12}, c_{14}, c_{32}, c_{52}, c_{54}$, що визначають набір ребер, які входять в даний переріз.

Якщо $c(x, y) > 0$, це означає, що ребро (x, y) існує і входить у переріз, при $c(x, y) = 0$ ребро відсутнє, тобто не входить у переріз. Обчислення значення пропускної здатності перерізу зводиться до додавання значень $c(x, y)$ ребер, що входять в переріз.

Розглянемо, як приклад, орієнтований граф, який показаний на рис. 9.46, матриця пропускних здатностей ребер графа

$$C = \begin{bmatrix} \infty & 0 & 0 & 1 & 3 \\ 0 & \infty & 0 & 0 & 4 \\ 4 & 1 & \infty & 2 & 0 \\ 0 & 3 & 2 & \infty & 0 \\ 0 & 0 & 0 & 5 & \infty \end{bmatrix}. \quad (9.148)$$

У наведеному прикладі величина пропускної здатності перерізу S , який відділяє вершини 1, 3, 5 від вершин 2, 4, складає $C(S) = c_{12} + c_{14} + c_{32} + c_{52} + c_{54} = 0 + 1 + 1 + 2 + 0 + 5 = 9$ ітд.

Розглянутий алгоритм дозволяє одночасно з обчисленням величин усіх перерізів визначити максимальні потоки між усіма парами вершин графа. Алгоритм ґрунтується на тому, що індекси усіх елементів матриці $c(x, y)$ збігаються не тільки з індексами усіх ребер (x, y) , які входять в переріз, але й з індексами усіх міжполюсних потоків V_{st} , розділених цим перерізом.

У результаті роботи алгоритму отримуємо матрицю V , елемент якої V_{st} визначає величину максимального потоку між вершинами s і t .

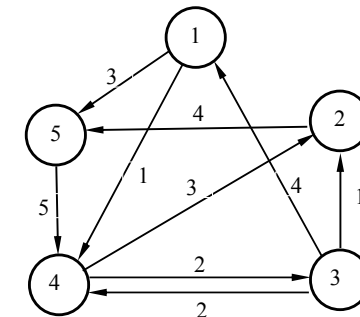


Рисунок 9.46 – Орієнтований граф (до матриці (9.148))

цьому мінімальний переріз проходить через ребра (1, 2), (2, 6), (6, 5) та (6, 7). Величина мінімального перерізу, утворена як сума пропускних здатностей ребер, що складають переріз, дорівнює величині максимального потоку:

$$c(1, 2) + c(2, 6) + c(6, 5) + c(6, 7) = 4 + 7 + 3 + 2 = V_{14}.$$

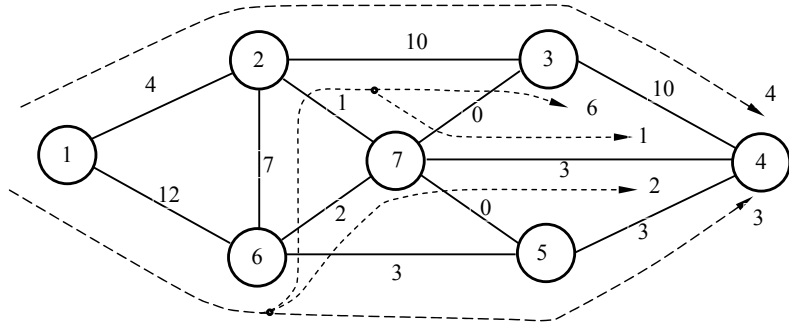


Рисунок 9.45 – Шляхи передачі максимального потоку

Матричний алгоритм, базується на пошуку множини перерізів, що розділяють задану пару вершин s та t , визначенні величин пропускних здатностей перерізів та виборі серед цих величин мінімальної.

Зручним способом подання перерізів є їх зображення у вигляді n -розрядного двійкового коду (n – число вершин графа), в якому кожній вершині відповідає свій розряд в послідовності 1, 2, 3, ..., n .

Дві підмножини вершин графа, утворюються перерізом, будемо інтерпретувати як підмножини витоків (типу s) та стоків (типу t). Першій з підмножин поставимо у відповідність поодинокі символи в коді перерізу, другій – нульові. Оскільки в неорієнтованій мережі підмножини витоків та стоків можна поміняти місцями (максимальний потік V_{st} дорівнює максимальному потоку V_{ts}), загальна кількість перерізів в такому графі у два рази менше, ніж в орієнтованому графі, у зв'язку з чим будь-який переріз в неорієнтованому графі може бути подано $(n - 1)$ -розрядним двійковим кодом. Беручи до уваги, що коди, які містять усі нулі чи усі одиниці, не є перерізами, в орієнтованому графі можливо $2^n - 2$ перерізів, в неорієнтованому – $2^{n-1} - 2$ перерізів.

Перелік ребер, які входять до будь-якого перерізу, легко отримати з матриці пропускних здатностей гілок C та коду цього перерізу. Очевидно, якщо рядкам матриці C надати значення витоків, а стовпцям – стоків, то до перерізу, заданого деяким двійковим кодом, входять ребра, визначені елементами матриці, що лежать на перерізах відповідних цьому кодові рядків і стовпців (елементами перерізу). Так, для графа, який містить п'ять вершин, матриця C в загальному випадку буде мати вид:

оптимізації функцій однієї або багатьох змінних переходить у задачі оптимізації функціоналів, якщо критерій виявляється залежним від аргументу складнішого вигляду, наприклад, сигналу що є іншою функцією – функцією часу). Тоді задача оптимізації функціонала – функції, аргументом якої є інша функція, може бути зведена до задачі оптимізації функції нескінченно великого числа змінних.

Розв'язання поставлених вище задач на екстремум можуть відшукуватися як аналітично, так і за допомогою високоефективних обчислювальних алгоритмів. Обидва підходи інтенсивно використовувалися впродовж останніх десятиліть і призвели до суттєвого прогресу в теорії і техніці зв'язку. Сучасні високошвидкісні системи передачі даних, принаймні, в частині, що стосується до пристроїв перетворення сигналів, пройшли не тільки стадію оптимізації в процесі їх дослідження і розробки, але і використовують методи оптимізації для адаптації до умов передавання дискретних сигналів, що змінюються.

Розглянуті питання теоретичного вивчення методів оптимізації та їх практичного застосування при пошуку оптимальних рішень у проектуванні, в інженерних додатках, налаштуванні й експлуатації телекомунікаційних систем.

Викладений матеріал по чисельних методах оптимізації функцій однієї і багатьох змінних, методах з використанням похідних, пошуку екстремуму градієнтними методами, елементах варіаційного числення, цифровій фільтрації пов'язаний з принциповими питаннями передавання дискретної інформації.

У книзі наведені основні поняття теорії графів, а також показані можливості цієї теорії по моделюванню мереж зв'язку та проведення мережного аналізу. Розглянуті ефективні обчислювальні алгоритми пошуку оптимальних шляхів та максимального потоку в графі.

Розглянута математична модель телекомунікаційної мережі, яка побудована на засадах теорії систем масового обслуговування та є основою розв'язання задач її оптимізації. При цьому за мету ставиться визначення структурних параметрів або алгоритмів функціонування мережі такими, для яких при заданих потоках, якості і дисципліні обслуговування вартість або обсяг мережі будуть мінімальні, або при заданих потоках, дисципліні обслуговування і вартості якісні показники функціонування мережі мають бути оптимальними. Розв'язання складних задач оптимізації потребує установлення залежностей і значень величин, що характеризують якість обслуговування, від характеристик і параметрів вхідного потоку вимог. Необхідне визначення структурних параметрів мережі при заданих потоках, дисципліні і якості обслуговування є предметом розв'язання задач синтезу телекомунікаційних мереж.

Застосування методів оптимізації є реально можливим лише в тих випадках, коли відповідна практична задача знаходить математичний опис. Оптимізаційні методи орієнтовані на реалізацію за допомогою ЕОМ. Детальний розгляд структури кожного з методів дозволяє для кожного з методів застосувати програмування на ЕОМ, використати обчислювальну техніку, провести багатоітераційні розрахунки та одержати ефективно й точно розв'язання задачі оптимізації з заданими вимогами.

По кожному з розділів розглянуті приклади розв'язання оптимізаційних задач та використання методів оптимізації в технічних застосуваннях, а саме при корекції лінійних спотворень каналу зв'язку, у задачах синтезу цифрових фільтрів, при пошуку оптимальних значень коефіцієнтів гармонійного коректора.

Метою підручника є досягнення студентами знань специфічної кваліфікації та умінь описувати й реалізовувати відповідні алгоритми аналогової і цифрової обробки сигналів.

Підручник написаний на основі дисциплін, що викладаються у ОНАЗ ім. О.С. Попова. Зміст відповідає вимогам державних загальноосвітніх стандартів у сферах математики і телекомунікаційних систем, та навчальній програмі дисципліни «Методи оптимізації» для студентів спеціальностей 7.092401 – Телекомунікаційні системи та мережі та 7.092402 – Інформаційні мережі зв'язку.

У написанні розділів 9.1 і 9.2 приймали участь доцент Ільїн Д.Ю. та інж. Рябуха О.М., розділу 9.3 – аспірант Іскендерзаде Ш.Г.

Підручник може бути корисним спеціалістам, що працюють у галузі телекомунікаційних систем та мереж зв'язку.

$\mu_{14} = 1-6-2-3-4$, яким передаються шість одиниць потоку, і $\mu_{14} = 1-6-2-7-4$, яким передається одна одиниця потоку.

На цьому процес знаходження шляхів припиняється, оскільки граф поділяється на дві незв'язні частини за перерізом, що дорівнює за величиною максимальному потоку V_{14} .

Результуючий максимальний потік

$$V_{14} = 0 + 4 + 3 + 2 + 6 + 1 = 16 \text{ одиниць.}$$

Результуюча скоректована матриця C'_p отримана після знаходження усіх шляхів

$$C'_p = \begin{bmatrix} \infty & 0 & 0 & 0 & 0 & 2 & 0 \\ 8 & \infty & 0 & 0 & 0 & 14 & 9 \\ 0 & 20 & \infty & 0 & 0 & 0 & 5 \\ 0 & 0 & 20 & \infty & 23 & 0 & 9 \\ 0 & 0 & 0 & 17 & \infty & 6 & 5 \\ 26 & 0 & 0 & 0 & 0 & \infty & 0 \\ 0 & 11 & 5 & 3 & 5 & 4 & \infty \end{bmatrix}. \quad (9.145)$$

Після віднімання результуючої матриці C_p з вихідної C отримуємо матрицю $C'' = C - C_p$

$$C'' = \begin{bmatrix} \infty & 4 & 0 & 0 & 0 & 12 & 0 \\ -4 & \infty & 10 & 0 & 0 & -7 & 1 \\ 0 & -10 & \infty & 10 & 0 & 0 & 0 \\ 0 & 0 & -10 & \infty & -3 & 0 & -3 \\ 0 & 0 & 0 & 3 & \infty & -3 & 0 \\ -12 & 7 & 0 & 0 & 3 & \infty & 2 \\ 0 & -1 & 0 & 3 & 0 & -2 & \infty \end{bmatrix}, \quad (9.146)$$

позитивні елементи якої $c''(x, y) > 0$ визначають використовувані ємності гілок (x, y) для передачі максимального потоку за шляхами множини $\{\mu_{14}\}$.

З (9.146) видно, що для передачі максимального потоку V_{14} використовуються ребра, величини дугових потоків яких дорівнюють:

$$f_{12} = 4 \text{ од.}; f_{16} = 12 \text{ од.}; f_{23} = 10 \text{ од.}; f_{27} = 1 \text{ од.};$$

$$f_{34} = 10 \text{ од.}; f_{54} = 3 \text{ од.}; f_{62} = 7 \text{ од.}; f_{65} = 3 \text{ од.};$$

$$f_{67} = 2 \text{ од.}; f_{74} = 3 \text{ од.}$$

Шляхи передачі максимального потоку із зазначенням числа одиниць, збільшуючих потік, показані на рис. 9.45. На рисунку числа біля ребер визначають величини дугових потоків $f(x, y)$. Із порівняння рис. 9.44 та 9.45 виходить, що ребра $(1, 2)$, $(2, 3)$, $(3, 4)$, $(2, 6)$, $(6, 5)$, $(6, 7)$ повністю зайняті максимальним потоком V_{14} , що передається; для них $f(x, y) = c(x, y)$. При

Задана мережа, що показана на рис. 9.44, з матрицею C пропускних здатностей гілок

$$C = \begin{bmatrix} \infty & 4 & 0 & 0 & 0 & 14 & 0 \\ 4 & \infty & 10 & 0 & 0 & 7 & 10 \\ 0 & 10 & \infty & 10 & 0 & 0 & 5 \\ 0 & 0 & 10 & \infty & 20 & 0 & 6 \\ 0 & 0 & 0 & 20 & \infty & 3 & 5 \\ 14 & 7 & 0 & 0 & 3 & \infty & 2 \\ 0 & 10 & 5 & 6 & 5 & 2 & \infty \end{bmatrix}. \quad (9.143)$$

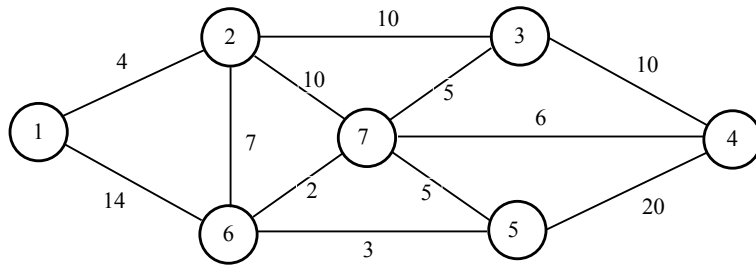


Рисунок 9.44 - Мережа графа (до табл.9.17)

Прийmemo витік $s = 1$, стік $t = 4$. Вихідне значення потоку $V_{14} = 0$.

На першому етапі визначається шлях $\mu_{14} = 1-2-3-4$, яким передаються 4 одиниці потоку. При цьому пропускні здатності гілок $(1, 2)$, $(2, 3)$, $(3, 4)$ зменшуються на чотири одиниці, пропускні здатності обернених гілок $(2, 1)$, $(3, 2)$, $(4, 3)$, збільшуються на таку саму величину. У результаті виконаних дій отримаємо скоректовану матрицю C'

$$C' = \begin{bmatrix} \infty & 0 & 0 & 0 & 0 & 14 & 0 \\ 8 & \infty & 6 & 0 & 0 & 7 & 10 \\ 0 & 14 & \infty & 6 & 0 & 0 & 5 \\ 0 & 0 & 14 & \infty & 20 & 0 & 6 \\ 0 & 0 & 0 & 20 & \infty & 3 & 5 \\ 14 & 7 & 0 & 3 & \infty & 2 \\ 0 & 10 & 5 & 6 & 5 & 2 & \infty \end{bmatrix}. \quad (9.144)$$

Значення потоку $V_{14} = 0 + 4 = 4$.

Аналогічно далі отримаємо шляхи: $\mu_{14} = 1-6-5-4$, яким передаються три одиниці потоку; $\mu_{14} = 1-6-7-4$, яким передаються дві одиниці потоку;

1 ФУНКЦІЇ ОДНІЄЇ ЗМІННОЇ

Задача оптимізації, у якій характеристична міра задана функцією однієї змінної, відноситься до найбільш простого типу оптимізаційних задач. Проте, аналіз задач такого типу займає центральне місце в оптимізаційних дослідженнях як теоретичної, так і практичної спрямованості. Це зв'язано не тільки з тим, що саме такі задачі звичайно розв'язуються в інженерній практиці, але і з тим, що одномірні методи оптимізації часто використовуються для аналізу підзадач, що виникають при реалізації ітеративних процедур, орієнтованих на рішення багатомірних задач оптимізації. Важливість теоретичних і прикладних оптимізаційних задач з однією керованою змінною обумовила розробку великого числа алгоритмів їхнього рішення. Класифікація методів рішення одномірних задач ґрунтується на різних припущеннях і допущеннях щодо природи і властивостей функції $f(x)$.

1.1 Властивості функцій однієї змінної

Відповідно до найбільш простого визначення, функція $f(x)$ являє собою правило, що дозволяє кожному значенню x поставити у відповідність єдине значення $y = f(x)$. У цьому випадку x зветься *незалежною змінною*, а y – *залежною змінною*. Розглянемо множину $S \in R$, де R – множина усіх дійсних чисел. Ми можемо визначити відповідність (або перетворення), за допомогою якого кожній точці $x \in S$ приписується єдине числове значення. Така відповідність називається скалярною функцією f , визначеною на множині S .

Коли множина $S = R$, ми маємо справу з *повсюди визначеною функцією* однієї змінної. Якщо S є деяка підмножина множини R , то функція f визначена в *обмеженій області*.

Наприклад,

$f(x) = x^3 + 2x^* - x + 3$ для усіх $x \in R$ є *повсюди визначена функція*, тоді як

функція $f(x) = x^3 = 2x^2 - x + 3$ для усіх $x \in S = \{x - 5 \leq x \leq 5\}$ *визначена в обмеженій області*.

В теорії оптимізації f називається *цільовою функцією*, а S – *припустимою областю*, множини точок що задовольняють обмеженням, або *областю припустимих значень* x .

Ряд фізичних процесів можна описати (або побудувати моделі цих процесів) за допомогою *безперервних функцій*, тобто функцій, що мають властивість безперервності в кожній точці x , що належить областям їхнього визначення (див. рис. 1.1).

Однак в інженерних додатках нерідкі і такі випадки, коли доводиться використовувати *розривні функції*. Наприклад, якщо ми будемо графік функції, змінна якої визначена у окремих границях, то в результаті можемо одержати кусково-безперервну криву, зображену на рис. 1.2. Вона описується розривною функцією, яка може приймати значення в широкому діапазоні.

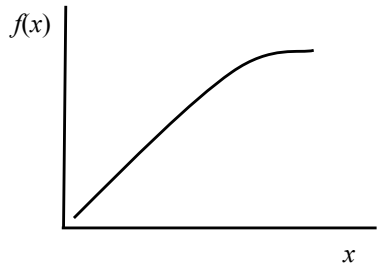


Рисунок 1.1 – Безперервна функція

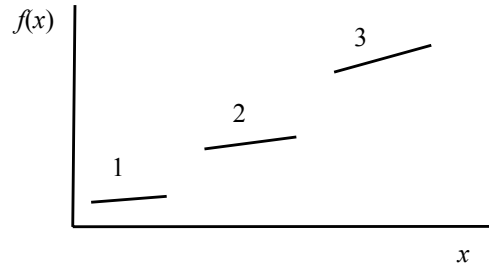


Рисунок 1.2 – Розривна функція

Зрозуміло, не завжди необхідно, щоб область припустимих значень незалежної змінної x містила всі дійсні числа з розглянутого інтервалу. Цілком можливі випадки, коли змінна приймає тільки *дискретні* значення. Наприклад, якщо ми будемо графік функції, що являє залежність показників від конкретно заданих проміжків часу, то природно обмежитися лише послідовністю точок, показаних на рис. 1.3, оскільки значення показника для кожного конкретного заданого проміжку часу є суворо фіксованим (кінцевим).

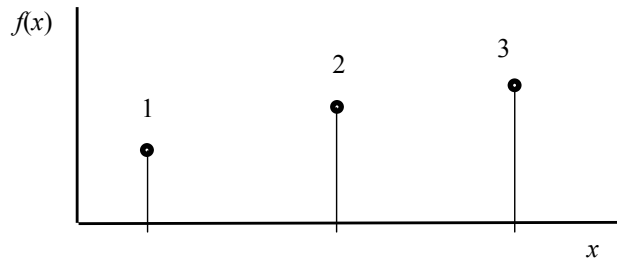


Рисунок 1.3 – Дискретна функція

Властивості безперервних функцій.

Безперервні функції мають наступні властивості:

- 1) сума або добуток безперервних функцій є безперервною функцією;
- 2) відношення двох безперервних функцій є функцією, безперервною в усіх точках, у яких знаменник відношення не обертається в нуль.

Очевидно, що в залежності від того, чи є досліджувана функція безперервною або розривною, а також у залежності від структури припустимої області для реалізації процедури пошуку точок екстремуму функції варто використовувати різні методи. Необхідно відзначити, що метод, ефективний при аналізі безперервних функцій, може виявитися неефективним при дослідженні розривних функцій, хоча зворотне не виключається.

На додаток до перерахованих вище властивостей можна також класифікувати функції відповідно до їх форми, що визначає топологічні властивості функцій у розглянутому інтервалі.

Подальше повторення операції А вже не дає можливості позначити стік t . Таким чином, потік на рис. 9.43 є максимальним. Мінімальний переріз складається з ребер (s, x) , (y, x) , (y, t) . Як видно з рис. 9.43, для ребер мінімального перерізу $c(x, y) = f(x, y)$.

Алгоритм, використовуючий пошук шляхів. Мережний алгоритм [23], що базується на пошуку послідовності найкоротших за кількістю проміжних вершин графа (найкоротших за рангом) шляхів між заданими витоком s та стоком t . Для кожного шляху μ_{st} визначається величина додаткового потоку P , після чого коректуються значення пропускних здатностей прямих дуг, що належать шляху, та відповідних зворотних дуг.

Алгоритм закінчує свою роботу, коли не існує жодного шляху, збільшуючого потік.

Для роботи алгоритму необхідно задати матрицю вихідних пропускних здатностей гілок мережі (ребер графа) C .

У результаті роботи алгоритму утворюється матриця C , що містить інформацію, необхідну для визначення дугових потоків $f(x, y)$. У вихідному стані матриця $C = C$.

Алгоритм може бути поданий послідовністю кроків.

1. Завдання початкового значення шуканого потоку: $V_{st} = 0$, матриці $C = C$.
2. Пошук чергового найкоротшого за рангом шляху.
3. Перевірка: шлях знайдено? Якщо так, то перехід до кроку 4, інакше – до кроку 8.
4. Визначення пропускної здатності шляху – величини додаткового потоку P , котрий може бути передано за знайденим шляхом:

$$P = \min_{x, y \in \mu_{st}} \{i(x, y), r(y, x)\}.$$
5. Збільшення шуканого потоку: $V_{st} = V_{st} + P$.
6. Корекція елементів матриці C : зменшення пропускних здатностей гілок (x, y) , що входять до шляху μ_{st} , на величину P , збільшення пропускних здатностей відповідних зворотних гілок (y, x) на ту саму величину P .
7. Перехід до кроку 2 алгоритму.
8. Віднімання отриманої матриці скоректованих пропускних здатностей гілок C з вихідної C , отримання матриці $C'' = C - C$.
9. Кінець роботи алгоритму.

У результаті виконання кроку 8 алгоритму визначається займання кожного ребра графа максимальним потоком, що передається, а також шляхів передачі максимального потоку.

Елементи матриці $c''(x, y) > 0$ визначають величини дугових потоків, а саме $f(x, y) = c''(x, y) > 0$. При цьому ребра графа, для котрих $f(x, y) = c(x, y)$ ($c(x, y)$ – значення пропускної здатності ребра (x, y)) в матриці C утворюють мінімальний переріз.

Розглянемо приклад, що ілюструє роботу алгоритму.

Операція В (зміна потоку). Стік t має позначку $(y^{\pm}, \varepsilon(t))$. Якщо ця позначка $(y^+, \varepsilon(t))$, то $f(y, t)$ змінюємо на $f(y, t) + \varepsilon(t)$; якщо ця позначка $(y^-, \varepsilon(t))$, то $f(t, y)$ змінюємо на $f(t, y) - \varepsilon(t)$. Потім у будь-якому з цих випадків переходимо до вершини y . Взагалі, якщо вершина y має позначку $(x^+, \varepsilon(y))$, то $f(x, y)$ змінюємо на $f(x, y) + \varepsilon(t)$, а якщо вона має позначку $(x^-, \varepsilon(y))$, то $f(y, x)$ змінюємо на $f(y, x) - \varepsilon(t)$ та переходимо до вершини x .

Коли досягнемо витoku s , зміна потоку припиняється, треба стерти всі старі позначки і знову перейти до операції А.

Процес розставляння позначок представляє, по суті, систематичний пошук шляху з s в t , збільшуючого потік.

Нехай дано граф, показаний на рис. 9.41, з пропускними здатностями ребер $c(x, y)$ та початковим потоком $f(x, y)$, що дорівнює одиниці, який проходить через вершини $\{s, x, y, t\}$. Числа $c(x, y)$ і $f(x, y)$ виписані по порядку біля кожного ребра.

Операцію А починаємо з того, що припишемо витoku s позначку $(-, \infty)$. Виходячи із s , приписуємо вершині y позначку $(s^+, \min(3, \infty)) = (s^+, 3)$. Це все позначки, котрі можна зробити з s . Єдина вершина, яку можна позначити з $y -$ це x . Вона може отримати позначку $(y^+, 1)$ або $(y^-, 1)$. Виберемо $(y^+, 1)$. Далі продовжуємо розставляння позначок від вершини x . При цьому позначається стік $t - (x^+, 1)$. Усі позначки показані на рис. 9.41. Це означає, що знайдено збільшуючий потік шлях, котрий можна визначити, повертаючись із стоку в напрямку, що вказується позначками. Вздовж цього шляху потік можна збільшити на $\varepsilon(t) = 1$. Цей шлях – $\{s, y, x, t\}$.

Новий потік величини 2 показано на рис. 9.42.

Стерши старі позначки, повторимо процес розставляння позначок. Оскільки $\varepsilon(t) = 1$, це означає, що вздовж знайденого шляху $s, (s, y), y, (x, y), x, (x, t), t$ потік можна збільшити на одиницю. Результуючий потік величини 3 показаний на рис. 9.43.

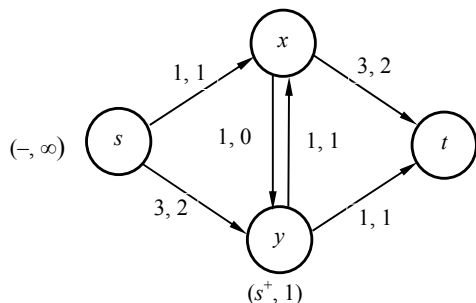


Рисунок 9.43 - Граф з результуючим потоком $f(x, y) = 3$

1.2 Монотонні й унімодальні функції

Функція $f(x)$ є *монотонною* (як при зростанні, так і при спаданні), якщо для двох довільних точок x_1 і x_2 , таких, що $x_1 \leq x_2$, виконується одне з наступних нерівностей:

$$f(x_1) \leq f(x_2) \quad (\text{монотонно зростаюча функція}),$$

$$f(x_1) \geq f(x_2) \quad (\text{монотонно спадаюча функція}).$$

На рис. 1.4 показаний графік монотонно зростаючої функції, а на рис. 1.5 – графік монотонно спадаючої функції. Помітимо, що монотонна функція не обов'язково повинна бути безперервною.

На рис. 1.6 показаний графік функції, що монотонно спадає при $x \leq 0$ і монотонно зростає при $x \geq 0$. Функція досягає свого мінімуму в точці $x = x^*$ (початку координат) і монотонна по обидві сторони від точки мінімуму. Такі функції називаються *унімодальними*.

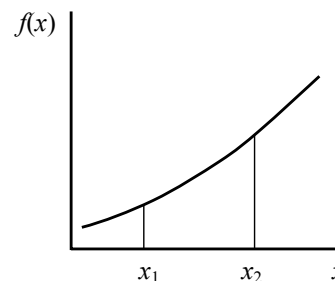


Рисунок 1.4 – Монотонно зростаюча функція

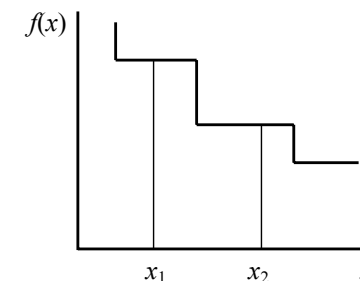


Рисунок 1.5 – Монотонно спадаюча функція

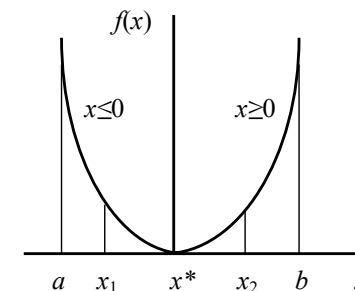


Рисунок 1.6 – Унімодальна функція

Визначення

Функція $f(x)$ є *унімодальною* на відрізку $a \leq x \leq b$ в тім і тільки тім випадку, якщо вона монотонна по обидві сторони від єдиної на розглянутому інтервалі екстремальної точки x^* [1]. Іншими словами, якщо x^* – єдина точка

мінімуму $f(x)$ на відрізку $a \leq x \leq b$, то $f(x)$ являється унімодальною на даному інтервалі тоді і тільки тоді, коли для точок x_1 і x_2 :

$$\text{з } x^* \leq x_1 \leq x_2 \text{ виходить, що } f(x^*) \leq f(x_1) \leq f(x_2),$$

$$\text{з } x^* \geq x_1 \geq x_2 \text{ виходить, що } f(x^*) \leq f(x_1) \leq f(x_2).$$

Як показано на рис. 1.7, унімодальна функція не обов'язково має бути безперервною. Унімодальність функцій є винятково важливою властивістю, що широко використовується в оптимізаційних дослідженнях.

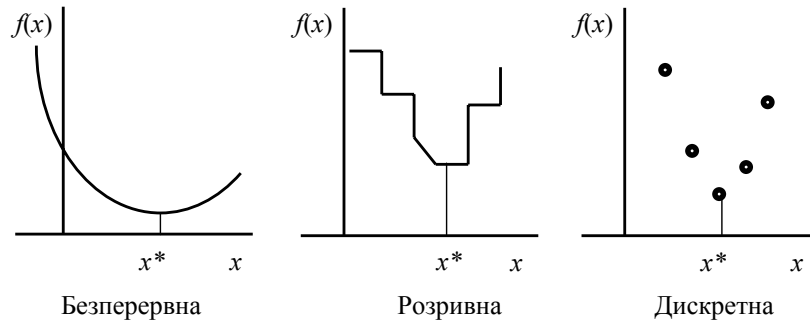


Рисунок 1.7 – Унімодальні функції

1.3 Критерії оптимальності функції однієї змінної

При аналізі оптимізаційних задач, як правило, виникають два загальних питання.

1. *Питання аналізу «у статистиці».* Як визначити, чи уявляє дана точка x^* оптимальне рішення задачі?

2. *Питання аналізу «у динаміці».* Якщо x^* не є точкою екстремуму, то яка послідовність дій приводить до одержання оптимального рішення?

У цьому розділі основна увага приділяється рішенню питання аналізу «у статистиці», а саме побудові множини *критеріїв оптимальності*, що дозволяє визначити, чи є дане рішення оптимальним.

Визначення

Функція $f(x)$, що визначена на множині S , досягає свого *глобального мінімуму* в точці $x^{**} \in S$ в тім і тільки тім випадку, якщо

$$f(x^{**}) \leq f(x) \text{ для усіх } x \in S.$$

Функція $f(x)$, що визначена на множині S , має *локальний мінімум (відносний мінімум)* у точці $x^* \in S$ в тім і тільки тім випадку, якщо $f(x^*) \leq f(x)$ для всіх x , віддалених від x^* на відстань, менше ε ,

$$x^* - \varepsilon < x < x^* + \varepsilon,$$

збільшувочими, та зменшення потоку в обернених ребрах, які є зменшувочими. Максимальне збільшення потоку за вибраним шляхом μ_{st} визначається як мінімум з таких двох величин:

$$\min\{i(x, y); (x, y) \in I - \text{пряме ребро}\};$$

$$\min\{r(x, y); (x, y) \in R - \text{обернене ребро}\}.$$

Для прикладу, показаного на рис. 9.40, максимальне збільшення потоку $P = \min\{i(s, a), r(b, a), r(c, b), i(c, t)\} = \min\{4, 5, 3, 6\} = 3$.

Кожен шлях із s в t будь-якого з трьох розглянутих типів, по якому можуть бути додатково послані одиниці потоку, називається збільшувочим шляхом або шляхом, збільшувочим потік із s в t .

Висновок: потік V із s в t максимальний тоді і тільки тоді, коли не існує шляху μ_{st} , який збільшує потік V .

Алгоритми пошуку максимального потоку базуються на теоремі Форда та Фалкерсона. Розділяються, як і алгоритми пошуку шляхів, на два класи – мережні та матричні.

Мережні алгоритми

Алгоритм розставляння позначок [18] може починатися з нульового потоку. Потім обчислення розвиваються у вигляді послідовності "розставляння позначок" (операція А), кожна з яких або призводить до потоку більшої величини (операція В), або ж закінчується висновком, що розглянутий потік є максимальним.

Вершинам графа приписуються позначки вигляду (x^+, ε) або (x^-, ε) , де $x \in N$ (N – множина вершин графа), а ε – натуральне число або ∞ .

Під час операції А кожна вершина знаходиться в одному з трьох станів: не позначена, позначена та переглянута або позначена і не переглянута.

Операція А (процес розставляння позначок). Витік s дістає позначку $(-, \varepsilon(s) = \infty)$. Після цього витік позначений та не переглянутий, а решта вершин не є позначеною.

Виберемо будь-яку позначену і не переглянуту вершину x . Нехай вона має позначку $(z^+, \varepsilon(x))$. Усім вершинам y , які не позначені та для котрих $f(x, y) < c(x, y)$, приписуємо позначку $(z^\pm, \varepsilon(x))$, де $\varepsilon(y) = \min(\varepsilon(x), c(x, y) - f(x, y))$. Такі вершини тепер позначені і не переглянуті. Усім вершинам y , які після цього не позначені та для яких $f(x, y) > 0$, приписуємо позначку $(x, \varepsilon(y))$, де $\varepsilon(y) = \min(\varepsilon(x), f(x, y))$.

Такі вершини y тепер позначені та не переглянуті. Вершина x після цього позначена та переглянута. Ця операція повторюється до тих пір, поки не буде позначеним та не переглянутим стік t , або ж до тих пір, поки не можна буде більше позначити жодну вершину, а стік залишиться непозначеним. В першому випадку переходимо до операції В, а у другому – максимальний потік визначений.

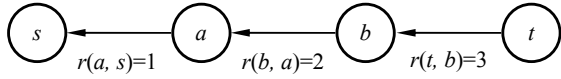


Рисунок 9.39 – Шлях, який складається із зменшуючих ребер

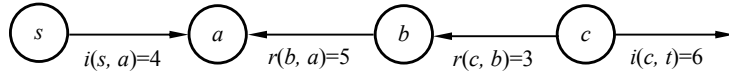


Рисунок 9.40 – Шлях, який складається із збільшуючих та зменшуючих ребер

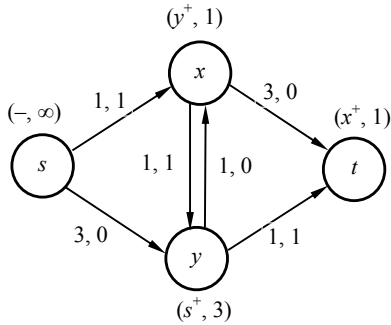


Рисунок 9.41 – Граф з початковим потоком $f(x, y) = 1$

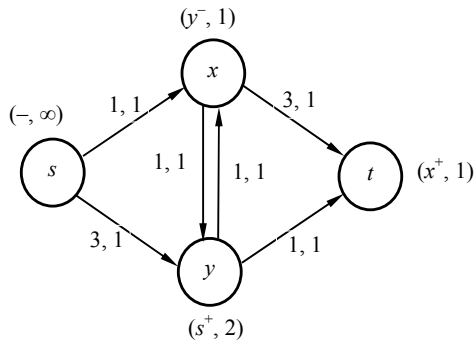


Рисунок 9.42 – Граф з новим потоком $f(x, y) = 2$

Якщо кожне пряме ребро належить до множини I , а обернене – до множини R , то вздовж розглядуваного шляху можна переслати додатковий потік із s в t . Це здійснюється шляхом збільшення потоку в прямих ребрах, які є

а саме, якщо існує $\varepsilon > 0$, таке, що для всіх x , що задовольняють умові $|x - x^*| < \varepsilon$, виконується нерівність $f(x^*) \leq f(x)$.

Зауваження

1. Аналогічні визначення глобального максимуму і локального максимуму можна одержати шляхом заміни знака нерівності на протилежний.
2. Якщо функція має властивість унімодальності, то локальний мінімум автоматично є глобальним мінімумом.
3. Якщо функція не є унімодальною, то можлива наявність декількох локальних екстремумів; при цьому глобальний мінімум можна визначити шляхом перебування всіх локальних екстремумів і вибору найменшого з них (див. рис. 1.8).

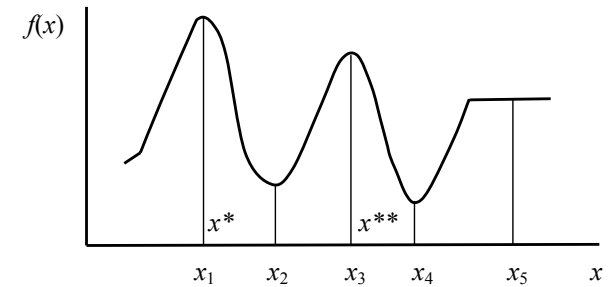


Рисунок 1.8 – Локальні і глобальні екстремуми

На рис. 1.8 x_1 – точка глобального максимуму, x_2 – точка локального мінімуму, x_3 – точка локального максимуму, x_4 – точка глобального мінімуму, а x_5 можна розглядати і як точку локального мінімуму, і як точку локального максимуму.

1.4 Ідентифікація екстремуму функції однієї змінної

Припустимо, що функція $f(x)$ однієї змінної x визначена на відкритому інтервалі (a, b) і n -кратно диференційована на цьому інтервалі. Якщо x^* – внутрішня точка інтервалу, то теорема Тейлора дозволяє записати зміну функції f при переході від точки x^* до точки $(x^* + \varepsilon)$ в наступному виді [2]:

$$f(x^* + \varepsilon) - f(x^*) = (\varepsilon) \frac{\partial f}{\partial x} \Big|_{x=x^*} + \frac{(\varepsilon^2)}{2!} \frac{\partial^2 f}{\partial x^2} \Big|_{x=x^*} + \dots + \frac{(\varepsilon^n)}{n!} \frac{\partial^n f}{\partial x^n} \Big|_{x=x^*} + o_{n+1}(\varepsilon), \tag{1.1}$$

де через $o_{n+1}(\varepsilon)$ позначена сума членів, у яких ступінь ε дорівнює $(n+1)$ і вище. Якщо x^* – локальний мінімум функції f на (a, b) , то за визначенням має

існувати ε -окіл точки x^* , такий, що для всіх x з цього окілу виконується нерівність

$$f(x) \geq f(x^*) \quad (1.2)$$

З нерівності (1.2) виходить, що

$$\varepsilon \frac{\partial f}{\partial x} \Big|_{x=x^*} + \frac{(\varepsilon^2)}{2!} \frac{\partial^2 f}{\partial x^2} \Big|_{x=x^*} + \dots + \frac{(\varepsilon^n)}{n!} \frac{\partial^n f}{\partial x^n} \Big|_{x=x^*} + o_{n+1}(\varepsilon) \geq 0. \quad (1.3)$$

При достатньо малому ε перший доданок домінує над іншими, а оскільки ε можна вибрати і позитивним, і від'ємним, то нерівність (1.3) буде виконуватися тільки у тому разі, якщо

$$\frac{\partial f}{\partial x} \Big|_{x=x^*} = 0. \quad (1.4)$$

Аналогічно розмірковуючи неважко установити, що нерівність (1.3) буде справедливою тільки тоді, коли

$$\frac{\partial^2 f}{\partial x^2} \Big|_{x=x^*} \geq 0. \quad (1.5)$$

Ця ж схема аналізу застосовна й у випадку визначення локального максимуму, з тією лише різницею, що знак нерівності (1.2) потрібно замінити на протилежний. Загальний результат можна сформулювати у виді наступної теореми.

Теорема 1.1

Необхідні умови того, що x^* є точкою локального мінімуму (максимуму) двічі диференційованої функції f на відкритому інтервалі (a, b) , виражаються наступними співвідношеннями:

- 1) $\frac{\partial f}{\partial x} \Big|_{x=x^*} = 0$,
- 2) $\frac{\partial^2 f}{\partial x^2} \Big|_{x=x^*} \geq 0$ (≤ 0).

Ці умови є необхідними, тобто у випадку, коли вони не виконуються, точка x^* не може бути точкою локального мінімуму (максимуму). З іншого боку, якщо ці умови виконуються, ми не маємо гарантії, що x^* є точкою локального мінімуму (максимуму). Розглянемо, наприклад, функцію $f(x) = x^3$, графік якої показаний на рис. 1.9. Ця функція задовольняє необхідним умовам наявності як локального мінімуму, так і локального максимуму на початку координат, однак не має ні максимуму, ні мінімуму при $x^* = 0$.

множини R . Це має місце в тому випадку, коли по ребру G вже протікає деякий потік. Відповідні ребра назвемо проміжними.

Позначимо через $i(x, y)$ максимальну величину, на яку може бути збільшений потік в ребрі (x, y) , через $r(x, y)$ – максимальну величину, на яку може бути зменшений потік в ребрі (x, y) . Очевидно,

$$i(x, y) = c(x, y) - f(x, y),$$

$$r(x, y) = f(x, y).$$

Припустимо, що ми хочемо переслати додаткову кількість одиниць потоку із витoku в стік. Можливі кілька способів розв'язання даної задачі (якщо вона існує).

Перший спосіб полягає у знаходженні шляху μ_{st} із витoku s в стік t , який складається із збільшуючих ребер. Оскільки $i(x, y)$ становить максимально можливе збільшення потоку в ребрі (x, y) , то величина додаткового потоку P по шляху μ_{st} складе:

$$P = \min_{x, y \in \mu_{st}} \{i(x, y)\}.$$

Для прикладу, показаного на рис. 9.38, по шляху μ_{st} можна переслати величину додаткового потоку

$$P = \min \{i(s, a), i(a, b), i(b, t)\} = \min \{3, 5, 2\} = 2.$$

Другий спосіб збільшення потоку із s в t полягає в знаходженні шляху μ_{ts} із t в s , який цілком складається із зменшуючих ребер. При цьому можна було б зменшити потік у кожному ребрі (x, y) , що призвело б до зменшення потоку із t в s , отже до збільшення чистого потоку із s в t . Через те що в кожному ребрі шляху μ_{ts} потік можна зменшити на величину $r(x, y)$, максимальне зменшення потоку на шляху μ_{ts} визначається величиною $\min \{r(x, y)\}; x, y \in \mu_{ts}$.

Для прикладу, показаного на рис. 9.39, по шляху μ_{ts} можна переслати назад максимум одну одиницю потоку:

$$P = \min \{r(t, b), r(b, a), r(a, s)\} = \min \{3, 2, 1\} = 1.$$

Універсальним є спосіб, який полягає в комбінації перших двох способів. При його використанні необхідно знайти шлях μ_{st} , ребра якого задовольняють таким умовам:

- всі прямі ребра шляху, направлені від s до t , належать множині I ;
- всі обернені ребра шляху, які мають напрямок від t до s , належать множині R .

Для прикладу розглянемо шлях μ_{st} , показаний на рис. 9.40. Прямими ребрами цього шляху є ребра (s, a) , (c, t) , оберненими – ребра (c, b) і (b, a) .

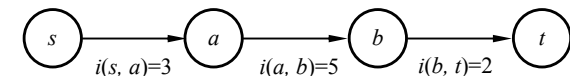


Рисунок 9.38 – Шлях, який складається із збільшуючих ребер

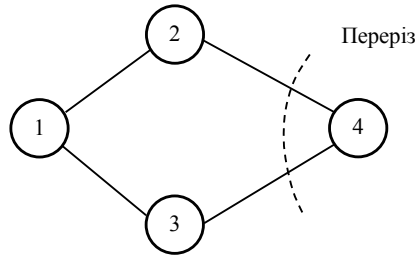


Рисунок 9.37 – Переріз графа

При розв'язанні задачі про максимальний потік інтерес становлять лише перерізи, які відокремлюють витік s від стоку t , для яких $s \in G'$, $t \in G''$.

Пропускною здатністю $C(S)$ або величиною перерізу S називається сума пропускних здатностей $c(x, y)$ усіх ребер, які складають переріз:

$$C(S) = \sum_{\substack{x \in G' \\ y \in G''}} c(x, y).$$

Пропускна здатність перерізу, показаного на рис. 9.37, дорівнює $c(2, 4) + c(3, 4)$ або $(c_{24} + c_{34})$.

Якщо переріз S розділяє вершини s і t , то величина максимального потоку з вершини s у вершину t має задовольняти умові $V \leq C(S)$. Один з найбільш важливих результатів теорії потоків у мережах сформульований теоремою про максимальний потік і мінімальний переріз, доведеною Фордом і Фалкерсоном [18]: *для будь-якої мережі з одним виток і одним стоком величина максимального потоку від витoku до стоку дорівнює величині максимального перерізу, який розділяє виток та стік.*

Ця теорема є основною в теорії потоків у мережах.

Нехай маємо граф G , в якому деяка кількість одиниць потоку проходить від витoku до стоку і для кожної одиниці потоку відомий маршрут руху. Ребра графа можна віднести до трьох різних категорій:

- ребра, в яких потік не може ні збільшуватися, ні зменшуватися (множину таких ребер позначимо через M);
- ребра, в яких потік може збільшуватися (множина I);
- ребра, в яких потік може зменшуватися (множина R).

Наприклад, ребра, що мають нульову пропускну здатність або значну вартість проходження потоку, мають належати до множини M ; ребра, в яких потік менший за пропускну здатність $f(x, y) < c(x, y)$, мають належати до множини I ; ребра, по яких уже проходить деякий потік, мають належати до множини R .

Ребра із множини I називаються збільшуваними, ребра із множини R – зменшуваними. Очевидно, що кожне ребро графа належить, принаймі, до однієї з множин M , I , R . Деякі ребра можуть належати як до множини I , так і до

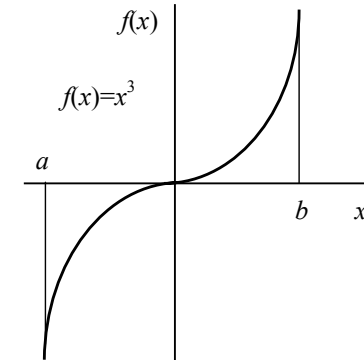


Рисунок 1.9 – Точка перегину

Для функції $f(x) = x^3$, при $n = 1$ $\left. \frac{\partial f}{\partial x} \right|_{x=0} = 3x^2 = 0$;

при $n = 2$ $\left. \frac{\partial^2 f}{\partial x^2} \right|_{x=0} = 6x = 0$;

при $n = 3$ $\left. \frac{\partial^3 f}{\partial x^3} \right|_{x=0} = 6$.

Визначення

Стационарною точкою називається точка x^* , у якій

$$\left. \frac{\partial f}{\partial x} \right|_{x=x^*} = 0.$$

Якщо стаціонарна точка не відповідає локальному екстремуму (мінімуму або максимуму), то вона є точкою перегину.

Для того, щоб провести розходження між випадками, коли стаціонарна точка відповідає локальному мінімуму, локальному максимуму або є точкою перегину, необхідно укласти достатні умови оптимальності.

Правила визначення точок локальних мінімумів і максимумів

Теорема 1.2

Нехай у точці вперше $(n - 1)$ похідні функції перетворюються в нуль, а похідна порядку n відмінна від нуля.

1. Якщо n – непарне, то x^* – точка перегину.
 2. Якщо n – парне, то x^* – точка локального екстремуму.
- Крім того:

- а) якщо ця похідна позитивна, то x^* – точка локального мінімуму;
 б) якщо ця похідна від'ємна, то x^* – точка локального максимуму.

Доказ

Твердження теореми можна довести за допомогою розвинення в ряд Тейлора, поданого рівністю (3.1). Оскільки порядок першої відмінної від нуля похідної дорівнює n , формулу (3.1) можна переписати в наступному виді:

$$f(x^* + \varepsilon) - f(x^*) = \frac{(\varepsilon^n)}{n!} \frac{\partial^n f}{\partial x^n} \Big|_{x=x^*} + o_{n+1}(\varepsilon). \quad (1.6)$$

Якщо n – непарне число, то права частина (1.6) може приймати як позитивні, так і від'ємні значення в залежності від того, чи є величина ε позитивною або від'ємною. Це означає, що в залежності від знака ε різниця $f(x^* + \varepsilon) - f(x^*)$ або позитивна, або від'ємна. Отже, функція не досягає в точці x^* свого мінімального або максимального значення, тобто x^* – точка перегину.

Далі розглянемо випадок, коли n – парне число. При цьому величина ε^n завжди позитивна, а знак правої частини (1.6) визначається першим доданком, якщо ε досить мала величина. Таким чином, якщо величина $\left(\frac{\partial^n f}{\partial x^n}\right) \Big|_{x=x^*}$ позитивна, то $f(x^* + \varepsilon) - f(x^*) > 0$ і точка x^* відповідає локальному мінімуму. Аналогічні міркування неважко провести також і для локального максимуму.

Для того, щоб застосувати теорему 1.2 до функції $f(x) = x^3$, графік якої показаний на рис. 1.2, раніше було підраховано:

$$\frac{\partial f}{\partial x} \Big|_{x=0} = 0, \quad \frac{\partial^2 f}{\partial x^2} \Big|_{x=0} = 0, \quad \frac{\partial^3 f}{\partial x^3} \Big|_{x=0} = 6.$$

Оскільки порядок першої відмінної від нуля похідної дорівнює 3 (непарне число), точка $x = 0$ є точкою перегину.

Зауваження

Вище передбачалося, що розглянута функція диференційована, або що її перша похідна існує і безперервна. Однак, якщо функція не є диференційованою в усіх точках області визначення, то навіть необхідна умова наявності екстремуму, що дозволяє ідентифікувати стаціонарні точки, може не виконуватися в точці екстремуму. Наприклад, розглянемо кусково-лінійну функцію

$$f(x) = \begin{cases} x & \text{при } x \leq 2, \\ 4 - x & \text{при } x \geq 2. \end{cases}$$

Ця функція безперервна у всіх точках дійсної осі, але недиференційована при $x = 2$. Функція досягає максимуму в точці $x = 2$, що не є стаціонарною відповідно до наданого вище визначення.

мінімізація вартості пересилання через певну систему певної кількості предметів з однієї точки в іншу, мінімізація часу перевезень в заданій системі.

Стосовно графів потік задає спосіб пересилання деяких об'єктів із однієї вершини графа в іншу по його ребрах (дугах). Вершина, з якої починається переміщення об'єктів, називається витокком; вершина, в якій закінчується переміщення об'єктів, називається стоком. Об'єкти, які переміщуються чи "протікають" із витокку в стік, називаються одиницями потоку.

Якщо кількість одиниць потоку, котра може проходити по ребру (x, y) обмежена, то ребро має обмежену пропускну здатність, котру прийнято позначати c_{xy} (або $c(x, y)$).

Припустимо, що є граф, в якому деяка кількість одиниць потоку проходить від витокку до стоку. Назвемо кількість одиниць, які проходять по ребру (x, y) потоком у даному ребрі (дузі), або *дуговим потоком*. Потік у ребрі (x, y) позначимо $f(x, y)$. Очевидно, $0 \leq f(x, y) \leq c(x, y)$.

Потоком із витокку s до стоку t в мережі називається множина невід'ємних чисел $f(x, y)$ (кожне із яких поставлене у відповідність деякому ребру графа), якщо ці числа задовольняють таким лінійним обмеженням [17, 21].

$$\sum_x f(x, y) - \sum_k f(y, k) = \begin{cases} -V, & \text{якщо } y = s; \\ 0, & \text{якщо } y \neq s, t; \\ V, & \text{якщо } y = t; \end{cases} \quad (9.141)$$

$$V \geq 0;$$

$$0 \leq f(x, y) \leq c(x, y) \text{ для усіх } x, y \quad (9.142)$$

Тут перша сума береться за ребрами, які ведуть до вершини y , друга – за ребрами, які ведуть із вершини y .

Число V називається величиною потоку.

Зазначимо, що обмеження (9.141) виражають той факт, що в кожному вершину, крім витокку та стоку, приходить стільки потоку, скільки з неї виходить (умова зберігання або неперервності потоку). Обмеження (9.142) показує, що потік (x, y) за дугою обмежений пропускну здатністю дуги $c(x, y)$.

Нехай V – величина потоку, який протікає з витокку s до стоку t по ребрах графа G . Якщо пропускну здатність ребер $c(x, y)$ графа G є кінцевою, то максимальна величина потоку буде обмежена величинами пропускну здатності ребер графа.

Максимальний потік визначається за допомогою одного з основних понять теорії графів – перерізу.

Переріз може бути визначений як множина ребер, вилучення яких із графа відокремило б деяку множину вершин від решти вершин, тобто вилучення яких із графа G поділяє його на два несполучених підграфи G' і G'' .

На рис. 9.37 показаний граф, в якому переріз, що складається із ребер (2, 4) та (3, 4), відокремлює вершину 4 від групи вершин 1, 2, 3. Тобто G' – це підграф, до якого входять вершини 1, 2, 3 та ребра (1, 2), (1, 3), (2, 3), G'' – підграф, що складається із вершини 4.

Процес триває доти, поки шість $(n - 1)$ вершин не дістануть позначку $*_j = 1$. Результуючі позначки записані останніми в стовпцях біля кожної вершини.

Отримані позначки містять всю необхідну інформацію про найкоротші шляхи з вершини 1. Так, на вершині 2 маємо результуючі позначки: $L_5 = 5$; $N_5 = 1$. Це означає, що довжина шляху $\mu_{12} - L_2 = 5$, шлях проходить через вершину 1, але $s = 1$, отже, шлях прямий, тобто $L_{12} = 5$; $\mu_{12} = \{1, 2\}$.

Для вершини 5: $L_5 = 31$; $N_5 = 6$. Довжина шляху $L_{15} = 31$, шлях проходить через вершину 6, але $s \neq 6$, отже, далі звертаємось до вершини 6 і визначаємо чергову вершину в шляху: $N_6 = 3$. Оскільки $3 \neq s$, звертаємось до вершини 3 і визначаємо $N_3 = 2$, далі $N_2 = 1$. Але $s = 1$, отже, шлях визначений. Таким чином, $L_{15} = 31$, $\mu_{15} = \{1, 2, 3, 6, 5\}$.

Аналогічно отримаємо решту шляхів:

$$L_{13} = 12, \mu_{13} = \{1, 2, 3\};$$

$$L_{14} = 17, \mu_{14} = \{1, 2, 4\};$$

$$L_{16} = 15, \mu_{16} = \{1, 2, 3, 6\};$$

$$L_{17} = 9, \mu_{17} = \{1, 7\}.$$

Розглянутий алгоритм може бути використаний для пошуку найкоротшого шляху між будь-якими двома заданими вершинами s і t . При цьому вершина s фіксується як витік, алгоритм працює у відповідності з поданою послідовністю кроків і закінчує роботу, як тільки за чергову, вибрану на кроці 2, вершину стане вершина t , тобто $i = t$.

Даний алгоритм може бути також використаний для пошуку шляхів мінімального рангу. Для цього досить усім ребрам графа надати вагу, що дорівнює одиниці: $I_{ij} = 1$. Тоді в результаті роботи алгоритму позначка L_j на вершинах графа $(j = \overline{1, n})$ визначить мінімальний ранг шляху з витіку у вершину j .

9.12 Алгоритм пошуку максимального потоку

Загальні відомості. Повідомлення, що надійшли до мережі від абонентів вузла s і призначені для абонентів вузла t , утворюють потік повідомлень, чи просто потік, між вузлами s і t .

Взагалі кажучи, потік визначає спосіб пересилання деяких об'єктів з одного пункту в інший. Потоки, наприклад, виникають при транспортуванні готової продукції від заводу до розподільного складу, при русі людей від місць проживання до місць роботи або при доставці листів від відправника до одержувача.

Незважаючи на різноманітність ситуацій, пов'язаних з потоками, в них виникає цілий ряд досить спільних проблем. Прикладами таких проблем могла б слугувати максимізація сумарного об'єму перевезень в певній системі з однієї точки в іншу;

Контрольні питання та задачі до розділу 1

1.1. Наведіть приклади функцій однієї змінної.

1.2. Сформулюйте умови досягнення функцією глобального мінімуму (максимуму).

1.3. Сформулюйте умови досягнення функцією строгого глобального мінімуму (максимуму).

1.4. Дайте визначення локального мінімуму (максимуму) функції однієї змінної.

1.5. Сформулюйте необхідні і достатні умови досягнення функцією локального екстремуму.

1.6. Яка послідовність дій при аналітичному дослідженні функції однієї змінної на екстремум?

1.7. Що таке точка перегину, і як її ідентифікувати?

1.8. Як перевірити, чи є функція опуклою або увігнутою?

1.9. У чому полягає властивість унімодальності функцій і в чому полягає важливе значення цієї властивості при рішенні задач оптимізації з однією змінною?

1.10. Нехай дана точка задовольняє достатнім умовам існування локального мінімуму. Як установити, чи є цей мінімум глобальним?

1.11. Задані наступні функції однієї змінної:

а) $f(x) = x^5 + x^4 - (x^3/3) + 2$;

б) $f(x) = (2x + 1)^2(x - 4)$.

Для кожної з заданих функцій знайдіть:

1) інтервал(и) зростання, спадання;

2) точки перегину (якщо такі є);

3) інтервал(и), у якому (у яких) функція увігнута, опукла;

4) локальні і глобальні максимуми (якщо такі є);

5) локальні і глобальні мінімуми (якщо такі є).

1.12. Установіть, які з наступних функцій є опуклими або увігнутими:

а) $f(x) = e^x$;

б) $f(x) = e^{-x}$;

в) $f(x) = 1/x^2$;

г) $f(x) = x + \log x$ при $x > 0$;

д) $f(x) = |x|$;

ж) $f(x) = x \log x$ при $x > 0$;

к) $f(x) = x^{2k}$, де k – ціле число;

л) $f(x) = x^{2k+1}$, де k – ціле число.

1.13. Розгляньте функцію

$$f(x) = x^3 - 12x + 3 \text{ в інтервалі } -4 \leq x \leq 4.$$

Знайдіть локальні мінімуми, локальні максимуми, глобальний мінімум і глобальний максимум f у заданому інтервалі.

1.14. Визначте області, у яких наступна функція опукла або увігнута:

$$f(x) = e^{-x^2}.$$

Знайдіть глобальний максимум і глобальний мінімум цієї функції.

1.15. Розгляньте задачу безумовної оптимізації з цільовою функцією однієї змінної $f(x)$. Користуючись наведеними в таблиці даними про значення похідних порядку 1, 2, 3, 4 у точках x_i ($i = 1, 2, \dots, 10$), ідентифікуйте кожну з точок (установіть, чи являється вона точкою максимуму, мінімуму, перегину або не є точкою екстремуму; укажіть випадки, коли не можна зробити означений висновок і т. ін.).

Таблиця 1.1 – Значення похідних порядку 1, 2, 3, 4 у точках x_i

x_i	$f'(x_i)$	$f''(x_i)$	$f'''(x_i)$	$f^{(4)}(x_i)$
x_1	0	+	Немає даних	Немає даних
x_2	0	0	+	Немає даних
x_3	0	-	Немає даних	Немає даних
x_4	-	-	Немає даних	Немає даних
x_5	0	0	-	Немає даних
x_6	0	0	0	-
x_7	0	0	0	0
x_8	0	0	0	+
x_9	+	+	Немає даних	Немає даних
x_{10}	0	-	+	-

Література до розділу 1

1 Захарченко Н.В. Оптимизация и моделирование систем связи: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, 1988. – Ч. 1.– 86 с.

2 Реклейтис Г. Оптимизация в технике: В 2-х кн. – Кн.1 / Реклейтис Г., Рейвиндран А., Рэгсдел К. – М.: Мир, 1986. – 352 с.

позначки: $i \quad L_1 = 0; N_1 = 0; *_1 = 0$. Решта вершин дістає позначки $L_j = \infty; N_j = 0; *_j = 0$ ($i = 1, n, j \neq s$).

Біля кожної вершини на рис. 9.36 подано позначки, отримані в процесі роботи алгоритму. Верхні позначки – початкові.

Спочатку усі вершини графа мають третю позначку $* = 0$, серед вершин вибирається $i = s = 1$, оскільки $L_1 = 0$, а для решти вершин $L_j = \infty$.

Із вершиною 1 мають безпосередні зв'язки вершини 2, 3, 7. Для них формуються нові позначки:

$$L_2^H = L_1 + I_{12} = 0 + 5 = 5 < L_2 = \infty,$$

$$L_3^H = L_1 + I_{13} = 0 + 15 = 15 < L_3 = \infty,$$

$$L_7^H = L_1 + I_{17} = 0 + 9 = 9 < L_7 = \infty.$$

Як бачимо, для усіх вершин нова позначка має замінити стару, при цьому $N_2 = 1, N_3 = 1, N_7 = 1$.

Нові позначки на вершинах 2, 3, 7 записані у другий рядок.

Вершині 1 присвоюється позначка $*_1 = 1$ і вершина 1 з усіх подальших дій вибуває.

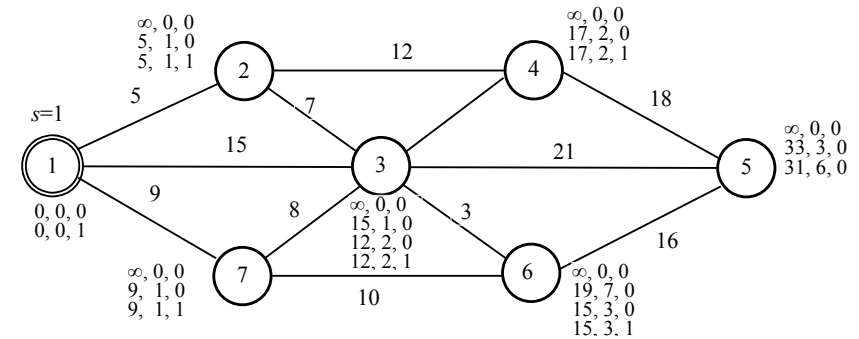


Рисунок 9.36 – Граф мережі з позначками

Далі вибирається вершина 2: $i = 2, L_2 = 5$. Від неї переглядаються вершини 3 і 4:

$$L_3^H = L_2 + I_{23} = 5 + 7 = 12 < L_3 = 15,$$

$$L_4^H = L_2 + I_{24} = 5 + 10 = 15 < L_4 = \infty.$$

В результаті на вершинах 3 і 4 записуються нові позначки:

$$L_3 = 12; N_3 = 2;$$

$$L_4 = 15; N_4 = 2.$$

Вершина 2 дістає позначку $*_2 = 1$ і з подальшого розгляду вибуває.

найкоротшого за заданим критерієм (довжина, вартість, час і т. ін.) шляху з вершини-витоку s у дану вершину j ; N_j – номер вершини, що є передкінцевою в найкоротшому шляху з s в j ; $*_j$ – службова позначка, яка формується в процесі пошуку найкоротших шляхів для збільшення швидкості збіжності алгоритму. Приймається $*_j = 0$, якщо найкоротший шлях μ_{sj} із s в j ще не визначено; $*_j = 1$, якщо найкоротший шлях μ_{sj} визначено.

Сутність алгоритму зводиться до виконання таких кроків.

1. Присвоєння початкових позначок усім вершинам графа:

$$L_s = 0; N_s = 0; *_s; (s - \text{вершина-витік});$$

$L_j = 0; N_j = 0; *_j; (i = \overline{1, n}, j \neq n)$. Тут номери передкінцевих вершин позначені нулями. Мається на увазі, що в нумерацію вершин графа число "0" не включено.

2. Серед усіх вершин графа $j (i = \overline{1, n})$, в яких найкоротший шлях ще не визначений, тобто у яких $*_j = 0$, вибирається i , яка має мінімальне значення L_j (при першому виборі завжди $i = s$).

3. Для усіх вершин q , суміжних вибраній i , у котрих $*_q = 0$, формується нова позначка $L_q^H = L_i + I_{iq}$. Тут L_i – позначка вибраної вершини i , I_{iq} – довжина (вартість, час і т. ін) ребра зв'язку вершини та перевірюваної вершини q .

Нова позначка L_q^H порівнюється з позначкою L_q , котра вже присвоєна вершині q .

Якщо $L_q^H < L_q$, відбувається заміна позначок L_q і N_q на вершині q :

$$L_q = L_q^H; N_q = i; *_q - \text{не змінюється.}$$

4. Після перевірки усіх вершин, суміжних до вершини i , вершині i присвоюється позначка $*_i = 1$.

5. Перевірка закінчення роботи алгоритму за умовою: $(n - 1)$ вершина дістала позначку $*_i = 1$. Якщо умова виконується, то алгоритм роботу закінчує, якщо ні, то здійснюється перехід до кроку 2 алгоритму.

В результаті роботи алгоритму на кожній вершині отримана інформація про довжину найкоротшого шляху з вершини s у відповідну вершину. Послідовність вершин, що складають шлях, визначається позначками N_j . Ця послідовність буде зворотною. Заміна її на пряму не становитиме труднощів.

Так, нехай необхідно отримати шлях μ_{sj} – із s в j . На вершині j визначається передкінцева вершина p – в шляху: $p = N_j$, далі на вершині p визначається чергова вершина $q = N_p$ і так, поки витік s не стане черговою визначеною вершиною.

Розглянемо граф, зображений на рис. 9.36, числа біля ребер якого відповідають значенням I_{ij} . Нехай за витік вибрано вершину $s = 1$. Початкові

2 МЕТОДИ ПОШУКУ ТОЧКИ ЕКСТРЕМУМУ УСЕРЕДИНИ ЗАДАНОГО ІНТЕРВАЛУ

Розглянемо одномірні методи пошуку точки екстремуму усередині заданого інтервалу в процесі виконання двох етапів: етапу визначення границь інтервалу й етапу зменшення інтервалу пошуку, у якому використовуються метод розподілу інтервалу навпіл і метод золотого перетину.

2.1 Визначення границь інтервалу

На цьому етапі спочатку вибирається вихідна точка, а потім на основі правила вилучення будується відносно широкий інтервал, що містить точку екстремуму. Звичайно пошук граничних точок такого інтервалу проводиться за допомогою евристичних методів пошуку, хоча в ряді випадків можна також використовувати методи екстраполяції [1, 2, 3]. Відповідно до одного з евристичних методів, що був запропонований Свенном, $(k+1)$ -а пробна точка визначається за рекурентною формулою

$$x_{k+1} = x_k + 2^k \Delta, \quad k = 0, 1, 2, \dots,$$

де x_k – довільно вибрана початкова точка; Δ – величина кроку, що підбирається деяким способом. Знак Δ визначається шляхом порівняння значень

$$f(x_0), f(x_0 + |\Delta|) \text{ і } f(x_0 - |\Delta|).$$

Якщо

$$f(x_0 - |\Delta|) \geq f(x_0) \geq f(x_0 + |\Delta|),$$

то, відповідно до припущення про унімодалність, точка мінімуму повинна розміщуватися правіше точки x_0 і величина Δ вибирається позитивною (рис.2.1). Якщо змінити знаки нерівностей на протилежні, то Δ варто вибирати негативною. Якщо виконується

$$f(x_0 - |\Delta|) \geq f(x_0) \leq f(x_0 + |\Delta|),$$

то точка мінімуму лежить між $(x_0 - |\Delta|)$ та $(x_0 + |\Delta|)$ і пошук граничних точок завершений.

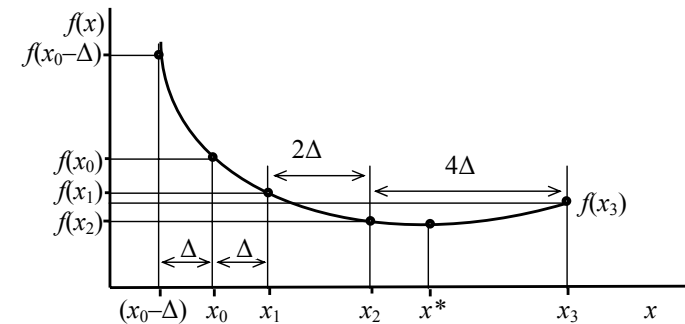


Рисунок 2.1 – Графічні ілюстрації до процедури пошуку граничних точок

Випадок, коли

$$f(x_0 - |\Delta|) \leq f(x_0) \geq f(x_0 + |\Delta|),$$

суперечить припущенню про унімодальність при пошуку точки мінімуму. Виконання цієї умови свідчить про те, що функція не є унімодальною.

Відзначимо, що ефективність пошуку граничних точок безпосередньо залежить від величини кроку Δ . Якщо Δ велике, то отримуємо грубі оцінки координат граничних точок, і побудований інтервал виявляється досить широким. З іншого боку, якщо Δ мале, для визначення граничних точок може знадобитися досить великий обсяг обчислень.

2.2 Етап зменшення інтервалу

Після того, як установлені границі інтервалу, що містить точку екстремуму, можна застосувати більш складну процедуру зменшення інтервалу пошуку з метою отримання уточнених оцінок координат екстремуму. Величина підінтервалу, що вилучається на кожному кроці, залежить від розміщення пробних точок x_1 і x_2 усередині інтервалу пошуку. Оскільки місцезнаходження точки екстремуму апріорі невідомо, доцільно припустити, що розміщення пробних точок повинне забезпечувати зменшення інтервалу в тому самому відношенні. Крім того, з метою підвищення ефективності алгоритму бажано, щоб зазначене відношення було максимальним. Подібну стратегію називають мінімаксною стратегією пошуку.

2.3 Метод вилучення інтервалів

Фактично всі одновимірні методи пошуку, які використовуються на практиці, засновано на припущенні, що досліджувана функція в припустимій області принаймні має властивість унімодальності. Корисність цієї властивості визначається тим фактом, що для унімодальної функції $f(x)$ порівняння значень $f(x)$ у двох різних точках інтервалу пошуку дозволяє визначити, у якому із заданих двома зазначеними точками підінтервалів точка екстремуму відсутня.

Теорема

Нехай функція $f(x)$ унімодальна на замкненому інтервалі $a \leq x \leq b$, а її мінімум досягається в точці x^* . Розглянемо точки x_1 і x_2 , які розміщені в інтервалі у такий спосіб, що $a < x_1 < x_2 < b$. Порівнюючи значення функції в точках x_1 і x_2 , можна зробити наступні висновки.

1. Якщо $f(x_1) > f(x_2)$, то точка мінімуму $f(x)$ не лежить в інтервалі (a, x_1) , тобто $x^* \in (x_1, b)$ (рис. 2.2, а).

2. Якщо $f(x_1) < f(x_2)$, то точка мінімуму $f(x)$ не лежить в інтервалі (x_2, b) , тобто $x^* \in (a, x_2)$ (рис. 2.2, б).

Вершини 2 і 3 безпосередньо пов'язані з вершиною 1, останньою із постійно позначених вершин. Для них:

$$L_2 = L_1 + c_{12} = 0 + 3 < \infty \quad L_3 = L_1 + c_{13} = 0 + 7 < \infty.$$

Тому вершинам 2 і 3 приписуються тимчасові позначки

$$L_2 = L_1 + c_{12} = 0 + 3 \quad L_3 = L_1 + c_{13} = 0 + 7.$$

Далі на кроці 2 вершина 2 дістає постійну позначку 3^n , від неї змінюються тимчасові позначки вершин 3 і 6, а саме:

$$L_3 = L_2 + c_{23} = 3 + 2 = 5 < 7, \quad L_6 = L_2 + c_{26} = 3 + 1 = 4 < \infty.$$

В результаті $L_3 = 7, L_6 = 4$.

На кроці 3 постійну позначку дістає вершина 6 – 4^n і т. ін.

У результаті виконання сьомого кроку вершині t приписується постійна позначка $L_8 = 7^n$. Отже, довжина найкоротшого шляху із s в t дорівнює 7. Цей шлях складається із ребер (i, j) , для кожного з яких різниця між значеннями постійних позначок його кінцевих вершин i та j дорівнює довжині цього ребра. Тобто умова, за якої вершини i та j належать до найкоротшого шляху, може бути записана:

$$L_j = L_i + c_{ij} \quad (9.140)$$

Співвідношення (9.140) можна використовувати рекурсивно, рухаючись від вершини t до вершини s . Визначивши вершину, що безпосередньо передує t в найкоротшому шляху, будемо повторювати дану процедуру до тих пір, поки не досягнемо вершини s .

Для розглядуваного прикладу найкоротший шлях μ_{st} утворюється послідовністю вершин:

$$\mu_{st} = \{s, 1, 6, 4, 5, t\}.$$

Для визначення передостанньої вершини 5 у шляху μ_{st} необхідно серед постійних позначок табл. 9.1 в її останньому рядку вибрати ту, для якої справедливе співвідношення (9.140):

$$L_1 = L_s + c_{s7}; \quad 7 = 7 + 0.$$

Далі серед вершин, суміжних вершині 5, шукаємо ту, для якої справедливий вираз (9.140):

$$L_5 = L_4 + c_{45}; \quad 7 = 6 + 1;$$

$$L_5 \neq L_3 + c_{34}; \quad 7 \neq 5 + 4.$$

Цією вершиною є вершина 4. Далі чинимо так само, поки не визначимо весь шлях.

Ефективний обчислювальний алгоритм, який дозволяє визначити найкоротші шляхи з фіксованої вершини-витоку s в решту вершин-стоків мережі поданий в роботі [23]. Алгоритм одночасно з пошуком найкоротших шляхів визначає і послідовність вершин, що складають шлях.

Алгоритм полягає в присвоєнні кожній вершині графа j ($j = \overline{1, n}$; n – кількість вершин) трьох позначок виду: $L_j, N_j, *j$. Тут L_j – довжина

що, якщо величина старої тимчасової позначки менша за другу з порівнюваних величин, то позначка залишається тією самою.

2. Серед тимчасових позначок вибрати ту, значення якої мінімальне, та оголосити її постійною позначкою. Якщо при цьому постійна позначка приписується вершині t , то алгоритм завершує роботу. У противному разі перейти до кроку 1.

Алгоритм може бути виконаний за допомогою таблиці розв'язання, в якій стовпці відповідають вершинам графа, стоки – крокам обчислювального процесу, а її елементи – постійним і тимчасовим позначкам.

Розглянемо граф, показаний на рис. 9.35. Вершина s тут є виток, t – стоком, а числа c_{ij} , приписані ребрам, відповідають їх довжині, вартості, часу і т. ін.

Робота алгоритму починається з того, що виток s приписується постійна позначка 0^n , а вершинам $1, 2, \dots, t$ – тимчасові позначки $L_j = c_{sj}$. Дані позначки записуються в табл. 9.1 як нульовий крок.

В табл. 9.1 на кроці 0 $L_j = 0$, тому що довжина ребра $c_{s1} = 0$, решта вершин мають позначки ∞ , тому що вони безпосередньо не зв'язані з вершиною s .

На кроці 1 вершина 1 одержує постійну позначку, тому що $L_1 = 0$ є мінімальною серед усіх тимчасових позначок.

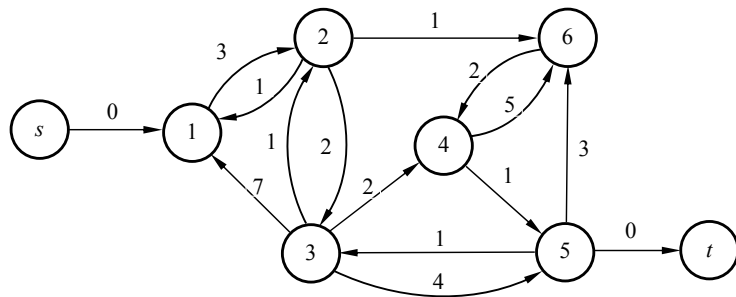
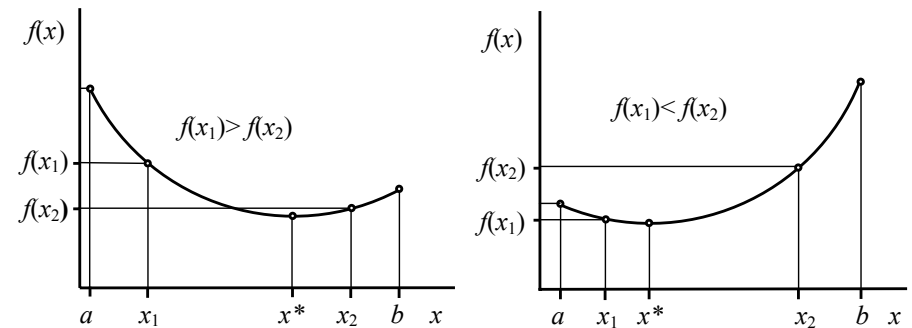


Рисунок 9.35 – Граф мережі (до задачі пошуку найкоротшого шляху)

Таблиця 9.1 – Дані позначки графа, поданого на рис. 9.35

Крок/вершина	s	1	2	3	4	5	6	7
0	0^n	0	∞	∞	∞	∞	∞	∞
1	0^n	0^n	3	7	∞	∞	∞	∞
2	0^n	0^n	3^n	5	∞	∞	4	∞
3	0^n	0^n	3^n	5	6	∞	4^n	∞
4	0^n	0^n	3^n	5^n	6	9	4^n	∞
5	0^n	0^n	3^n	5^n	6^n	7	4^n	∞
6	0^n	0^n	3^n	5^n	6^n	7^n	4^n	∞
7	0^n	0^n	3^n	5^n	6^n	7^n	4^n	7^n

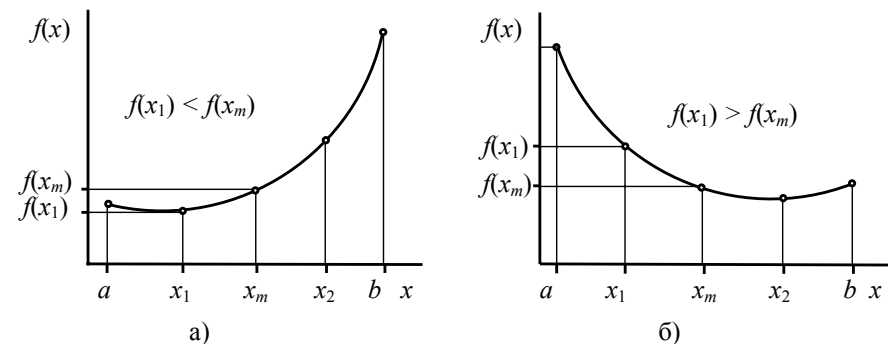


а) б)
Рисунок 2.2 – Графічні ілюстрації до теореми

Примітка. Якщо $f(x_1) = f(x_2)$, то можна вилучити обидва крайні інтервали (a, x_1) і (x_2, b) , при цьому точка мінімуму повинна розміщуватися в інтервалі (x_1, x_2) .

2.4 Метод розподілу інтервалу навпіл

Розглянутий метод дозволяє вилучати точно половину інтервалу на кожній ітерації. Іноді цей метод називають триточковим пошуком на рівних інтервалах, оскільки його реалізація заснована на виборі трьох пробних точок, які рівномірно розподілені в інтервалі пошуку. Нижче приводиться опис основних кроків пошукової процедури, орієнтованих на знаходження точки мінімуму функції $f(x)$ в інтервалі (a, b) (рис. 2.3)



а) б)
Рисунок 2.3 – Графічні ілюстрації до методу розподілу інтервалу навпіл

К р о к 1. Покласти $x_m = (a + b) / 2$ і $L = b - a$. Обчислити значення $f(x_m)$.

К р о к 2. Покласти $x_1 = a + L/4$ і $x_2 = b - L/4$. Помітимо, що точки x_1, x_m, x_2 поділяють інтервал (a, b) на чотири рівні частини. Обчислити значення $f(x_1)$ і $f(x_2)$.

К р о к 3. Порівняти $f(x_1)$ і $f(x_m)$.

1. Якщо $f(x_1) < f(x_m)$ (рис. 2.3, а), вилучити інтервал (x_m, b) , поклавши $b = x_m$. Середньою точкою нового інтервалу пошуку стає точка x_1 . Отже, необхідно покласти $x_m = x_1$. Перейти до кроку 5.

2. Якщо $f(x_1) \geq f(x_m)$ (рис. 2.3, б), перейти до кроку 4.

К р о к 4. Порівняти $f(x_2)$ і $f(x_m)$.

1. Якщо $f(x_2) < f(x_m)$, вилучити інтервал (a, x_m) , поклавши $a = x_m$. Тому що середньою точкою нового інтервалу стає точка x_2 покласти $x_m = x_2$. Перейти до кроку 5.

2. Якщо $f(x_2) \geq f(x_m)$, вилучити інтервали (a, x_1) і (x_2, b) . Покласти $a = x_1$ і $b = x_2$. Відзначимо, що x_m продовжує залишатися середньою точкою нового інтервалу. Перейти до кроку 5.

К р о к 5. Обчислити $L = b - a$. Якщо величина $|L|$ мала, закінчити пошук. У протилежному випадку повернутися до кроку 2.

Зауваження

1. На кожній ітерації алгоритму вилучається точно половина інтервалу пошуку.

2. Середня точка послідовно отриманих інтервалів завжди збігається з однією з пробних точок x_1, x_2 або x_m , знайдених на попередній ітерації. Отже, на кожній ітерації потрібно не більше двох обчислень значення функції.

3. Якщо проведено n обчислень значення функції, то довжина отриманого інтервалу складає $(1/2)^n$ величини вихідного інтервалу.

4. З усіх методів пошуку на рівних інтервалах (двоточковий, триточковий, чотириточковий і т. ін.) триточковий пошук, або метод розподілу інтервалу навпіл, відрізняється найбільшою ефективністю [2, 3].

2.5 Пошук за допомогою методу золотого перетину

З проведеного вище розгляду методів вилучення інтервалів і мінімаксних стратегій пошуку можна зробити наступні висновки.

1. Якщо кількість пробних точок приймається рівною двом, то їх варто розміщувати на однакових відстанях від середини інтервалу.

2. Відповідно до загальної мінімаксної стратегії пробні точки повинні розміщуватися в інтервалі за симетричною схемою таким чином, щоб відношення довжини підінтервалу, що вилучається, до величини інтервалу пошуку залишалось постійним.

3. На кожній ітерації процедури пошуку повинне обчислюватися тільки одне значення функції в отримуваній точці.

Керуючись цими висновками, розглянемо симетричне розміщення двох пробних точок на початковому інтервалі одиничної довжини, яке показано на рис. 2.4. (Вибір одиничного інтервалу зумовлений розуміннями зручності).

Пробні точки віддалені від граничних точок інтервалу на відстані τ . При такому симетричному розміщенні точок довжина інтервалу, що залишається після вилучення, завжди дорівнює τ незалежно від того, яке зі значень функції в пробних точках виявляється меншим.

$$\sum_j f_{ij} - \sum_j f_{jt} = -1; \quad (9.139)$$

$$f_{ij} \geq 0.$$

Тут f_{ij} – величина потоку, який протікає по ребру (i, j) .

Згідно з рівністю (9.137), одиниця потоку витікає із витоку s (тут j – суміжні з витоком s вершини).

Рівність (9.138) гарантує збереження даної одиниці потоку при протіканні по графу $((i, j)$ – будь-яке ребро графа).

Згідно з рівністю (9.139), одиниця потоку втікає в стік t (тут j – суміжні стоку t вершини).

За найкоротший шлях може бути взята послідовність суміжних ребер (i, j) , для яких $f_{ij} = 1$.

Для розв'язання поставленої задачі було розроблено спеціальний метод, відомий під назвою алгоритму Дейкстри [17, 21].

Алгоритм базується на приписуванні вершинам тимчасових або постійних позначок. Спочатку кожній вершині графа, за винятком витоку, приписується тимчасова позначка, відповідна довжині найкоротшого шляху, який веде із витоку в дану вершину. Витоку приписується постійна позначка, значення якої дорівнює нулю.

Кожній вершині, в яку не можна потрапити безпосередньо із витоку, приписується тимчасова позначка ∞ , а решті вершин – тимчасові позначки c_{sj} , $j \neq s$. Якщо визначено, що вершина належить до найкоротшого шляху, її позначка стає постійною. Алгоритм базується на простому твердженні: частиною найкоротшого шляху є найкоротший шлях.

Алгоритм починає працювати при $j = s$. Потім величина послідовно збільшується на одиницю й при $j = t$ алгоритм завершує роботу.

Для заданої вершини j через L_j будемо позначати оцінку довжини найкоротшого шляху із витоку s у вершину j . Якщо ця оцінка не може бути поліпшена, тоді відповідні значення назвемо постійною позначкою і будемо позначати L_j^n , в протилежному разі вершина має тимчасову помітку.

Спочатку постійна позначка приписується тільки витоку. Кожна інша позначка є тимчасовою та її величина дорівнює довжині ребра, яке веде із витоку у відповідну вершину. Для визначення найближчої до витоку вершини вибирається тимчасова позначка з мінімальним значенням і оголошується постійною. Далі, доти, поки стоку не буде приписана постійна позначка, необхідно виконувати дві процедури.

1. Розглянути решту вершин з тимчасовою позначкою. Порівняти величину кожної тимчасової позначки з сумою величини останньої з постійних позначок і довжини ребра, що веде із відповідної постійно позначеної вершини в розглядувану вершину. Мінімальна з двох порівнюваних величин визначається як нова тимчасова позначка розглядуваної вершини. Відзначимо,

Для пошуку за матрицею V' найкоротших шляхів необхідно виконати такі дії: нехай визначається шлях в деяку вершину $j - \mu_{ij}$. У стовпці j матриці V' відшукується чергова одиниця, фіксується номер рядка k , який визначить вершину k , що входить до шляху, потім переглядається стовпець з номером k , знаходиться в ньому чергова одиниця, фіксується номер рядка p , тобто визначається номер чергової вершини p , яка входить до шляху μ_{ij} , і поки черговий номер обраної вершини не стане однаковим з вихідним i .

Таким чином, утворена послідовність вершин подається у зворотному порядку. Подати послідовність в прямому порядку не становитиме великих труднощів.

Так, за матрицею V' отримаємо такі шляхи мінімального рангу з вершини 1 в решту вершин графа:

- 1-2
- 1-3
- 1-3-4
- 1-2-5, 1-3-5,
- 1-3-4-6, 1-2-5-6, 1-3-5-6.

Мережні алгоритми пошуку найкоротших шляхів

Задача пошуку найкоротшого шляху (шляху мінімальної вартості) з певного витоків s до певного стоку t може бути сформульована як задача лінійного програмування (ЛП).

Будемо вважати, як і раніше, що кожному ребру (i, j) графа поставлено у відповідність певне число c_{ij} , назване узагальненою вартістю ребра (як така вартість може використовуватись і значення I_{ij} - довжини ребра, t_{ij} - часу передачі інформації по ребру і т. ін.). Фіктивним чи "безоплатним" ребрам приписується вартість $c_{ij} = 0$, а кожній парі вершин (i, j) , що не мають прямого зв'язку, приписується вартість $c_{ij} = \infty$.

Задача знаходження найкоротшого шляху із витоків s до стоку t є задачею знаходження в заданому графі такого шляху, для якого вартість проходження одиниці потоку мінімальна.

Математично ця задача може бути записана як така задача ЛП: мінімізувати

$$\sum_i \sum_j c_{ij} f_{ij} \quad (9.136)$$

за умови, що

$$\sum_j f_{si} - \sum_j f_{js} = 1; \quad (9.137)$$

$$\sum_j f_{ij} - \sum_j f_{ji} = 0; \quad i \neq s, i \neq t; \quad (9.138)$$

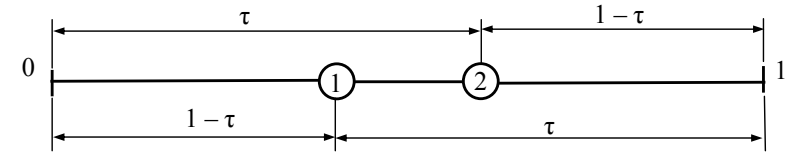


Рисунок 2.4 – Пошук за допомогою методу золотого перетину

Припустимо, що вилучається правий підінтервал. На рис. 2.5 показано, що підінтервал довжини τ , що залишився, містить одну пробну точку, яка розміщена на відстані $(1 - \tau)$ від лівої граничної точки.

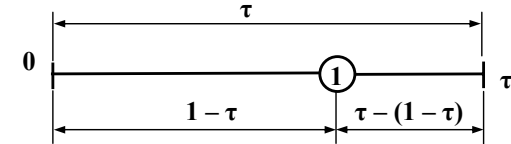


Рисунок 2.5 – Інтервали, отримані методом золотого перетину

Для того щоб симетрія пошукового зразка зберігалася, відстань $(1 - \tau)$ повинна складати τ -у частину довжини інтервалу (яка дорівнює τ). При такому виборі τ наступна пробна точка розміщується на відстані, рівній τ -й частині довжини інтервалу, від правої граничної точки інтервалу (рис. 2.6).

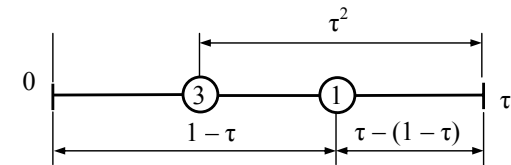


Рисунок 2.6 – Симетрія золотого перетину інтервалу

Отже, на першому кроці пошуку (рис.2.4) для точки 1 виконується співвідношення $\tau / 1 = (1 - \tau) / \tau$. Звідси виходить: $1 - \tau = \tau^2$.

Розв'яжемо це квадратне рівняння та отримаємо

$$\tau = (-1 \pm \sqrt{5})/2,$$

отже, позитивне рішення $\tau = 0,61803\dots$.

Перевіримо співвідношення отриманих відрізків (рис. 2.4) на наступному кроці пошуку, коли довжина інтервалу $(1, 2)$ дорівнює $\tau - (1 - \tau)$:

$$\frac{\tau - (1 - \tau)}{1 - \tau} = \frac{1 - \tau}{\tau},$$

$$\tau^2 - \tau(1 - \tau) = (1 - \tau)^2.$$

Поділимо обидві частини рівняння на $(1 - \tau)$.

Після перетворень остаточно отримуємо раніше записану умову симетрії пошукового зразка $\tau^2 = 1 - \tau$.

Таким чином, при виборі τ відповідно до умови $1 - \tau = \tau^2$ симетрія пошукового зразка, що показаний на рис. 2.4, зберігається при переході до зменшеного інтервалу, показаного на рис. 2.6.

Схема пошуку, за якою пробні точки ділять інтервал у цьому відношенні, відома за назвою пошуку за допомогою методу золотого перетину. Зауважимо, що після перших двох обчислень значень функції кожне наступне обчислення дозволяє вилучити підінтервал, величина якого складає $(1 - \tau)$ -у частку від довжини інтервалу пошуку.

Отже, якщо початковий інтервал має одиничну довжину, то величина інтервалу, який отриманий в результаті n обчислень значень функції, дорівнює τ^{n-1} . Можна показати, що пошук за допомогою методу золотого перетину є асимптотично найбільш ефективним способом реалізації мінімаксної стратегії пошуку.

У загальному випадку якщо права і ліва граничні точки інтервалу невизначеності (позначимо їх через XR і XL) відомі, то координати всіх наступних пробних точок, отримуваних згідно з методом золотого перетину, можна розрахувати за формулою:

$$w = XR - \tau^n \text{ або } w = XL - \tau^n,$$

у залежності від того, який підінтервал був вилучений на попередній ітерації – лівий чи правий. У наведених вище формулах через τ^n позначена n -на ступінь τ , де n – кількість обчислень значень функції.

Пошук за допомогою методу золотого перетину може бути закінчений виходячи з заданої кількості обчислень значень функції (і, отже, величини інтервалу невизначеності), або по досягненні відносної точності шуканого значення функції. Найкращим є використання обох критеріїв одночасно.

2.6 Порівняння методів вилучення інтервалів

Нижче проводиться порівняння відносних ефективностей розглянутих методів вилучення інтервалів. Позначимо довжину вихідного інтервалу невизначеності через L , а довжину інтервалу, отриманого в результаті n обчислень значень функції, – через L_n . Як показник ефективності того чи іншого методу вилучення інтервалів уведемо в розгляд характеристику відносного зменшення початкового інтервалу $K_L(n) = L_n / L$.

Нагадаємо, що при використанні методу розподілу інтервалу навпіл і методу золотого перетину довжина отриманого інтервалу складає $L(0,5)^{n/2}$ і $L(0,618)^{n-1}$, відповідно. Отже, відносне зменшення інтервалу після n обчислень значень функції дорівнює

$$K_L(n) = L(0,5)^{n/2} \text{ – для методу розподілу інтервалу навпіл;}$$

$$K_L(n) = L(0,618)^{n-1} \text{ – для методу золотого перетину.}$$

Для порівняння розглянемо також метод рівномірного пошуку (рис. 2.7), відповідно до якого оцінювання функції проводиться в n рівновіддалених одна від одної точках, при цьому інтервал L ділиться на $(n + 1)$ рівних інтервалів

$$B = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}. \quad (9.133)$$

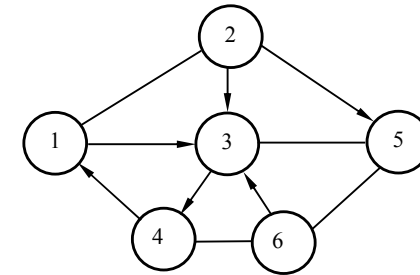


Рисунок 9.34 – Граф мережі (до алгоритму, побудованого на застосуванні операцій бульової алгебри).

В матриці \hat{A} елементи головної діагоналі $\beta_j = 0$.

Після першого етапу для вершини-витоку 1 в результаті перетворення рядків 2 і 3, суміжних вершині-витоку 1, матриця B' буде мати вид

$$B' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}. \quad (9.134)$$

Результуюча матриця B'

$$B' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (9.135)$$

відповідні ребрам графа, які входять хоча б у один найкоротший за транзитністю (рангом) шлях від фіксованої вершини-витоку в усі інші вершини-стоки графа. Перетворення матриці B в матрицю B' здійснюється виконанням операцій логічного додавання, логічного множення, заперечення рівнозначності, інверсії рядків матриці B .

Перетворення матриці B в B' завжди можливе, оскільки вихідна мережа подається кінцевим графом. Кількість етапів, необхідних для перетворення матриці, визначається максимальним значенням рангу в усіх найкоротших шляхах від вершини-витоку до решти вершин мережі. Максимально можлива кількість етапів, необхідних для перетворення матриці B , буде мати місце в разі подання мережі графом-деревом з двома кінцевими вершинами та становитиме $(n - 1)$ для графа з n вершинами.

"Етапом" назвемо процес вибору з вершин графа таких вершин, які є суміжними вже вибраним, тобто які мають з вибраними безпосередній зв'язок. Так, перший етап полягає у виборі вершин, суміжних вершині-витоку, другий – у виборі усіх вершин, суміжних вибраним на першому етапі, і т. ін., до тих пір, поки усі вершини графа не будуть вибрані.

Вихідними для роботи алгоритму є: матриця B і три допоміжних двійкових слова довжиною n – нагромадження H , гасіння Γ й ознаки закінчення формування матриці B' – ПФ.

Алгоритм полягає у виконанні кроків:

Крок 1. Запис вихідного стану Γ – нуль в розряді з номером вершини-витоку, одиниці – в усіх інших розрядах.

Крок 2. Запис рядка вершини-витоку з матриці B у матрицю B' , у H і ПФ.

Крок 3. Логічне множення інверсії H і Γ , результат залишається в Γ : $\Gamma = \bar{H} \& \Gamma$. Запис нулів в усі розряди H .

Крок 4. Визначення номера чергової вершини k у відповідності із значенням одиниці в k -му розряді ПФ.

Крок 5. Логічне множення k -го рядка матриці B і Γ . Результат розміщується в k -му ряду матриці B' : $B'_k = B_k \& \Gamma$.

Крок 6. Логічне додавання k -го рядка матриці B' і H , результат розміщується в H : $H = H \vee B'_k$.

Крок 7. Перевірка перегляду усіх рядків матриці B відповідно до змісту ПФ. Перехід до кроку 8 по закінченні перегляду, інакше – повернення до кроку 4.

Крок 8. Запис змісту H в ПФ; $H \rightarrow \text{ПФ}$.

Крок 9. Перевірка закінчення формування матриці B' . Якщо ПФ = 0, то це кінець алгоритму, інакше – повернення до кроку 3.

Для мережі, показаної на рис. 9.34, матриця $B = [b_{ij}]$ має вид:

довжини $L/(n + 1)$. Нехай x^* – точка, у якій спостерігається мінімум функції $f(x)$. Тоді точка істинного мінімуму $f(x)$, виявляється, міститься в інтервалі $\{[x^* - L/(n + 1)], [x^* + L/(n + 1)]\}$,

отже, $L_n = 2L/(n+1)$.

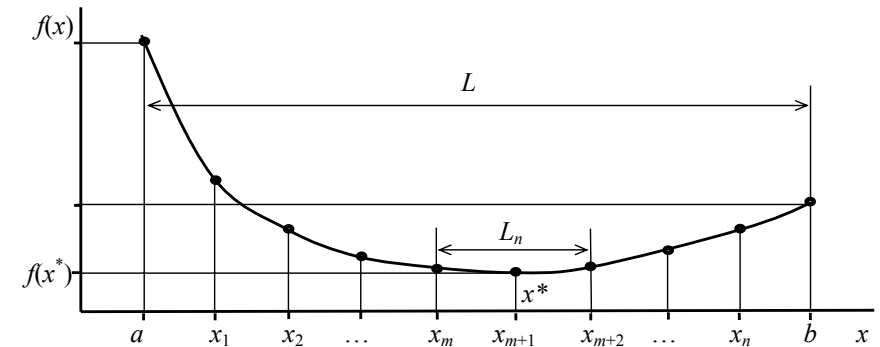


Рисунок 2.7 – Графічні ілюстрації до методу рівномірного пошуку

Отже, для методу рівномірного пошуку

$$K_L(n) = 2/(n+1).$$

У табл. 2.1 наведені значення $K_L(n)$, що відповідають обраним n , для трьох методів пошуку. З таблиці виходить, що пошук за допомогою методу золотого перетину забезпечує найбільше відносне зменшення вихідного інтервалу при однаковій кількості обчислень значень функції.

Таблиця 2.1 – Величини відносного зменшення інтервалу

Метод пошуку	Кількість обчислень значень функцій				
	$n = 2$	$n = 5$	$n = 10$	$n = 15$	$n = 20$
Метод розподілу інтервалу навпіл	0,5	0,177	0,031	0,006	0,0009
Метод золотого перетину	0,618	0,146	0,013	0,001	0,0001
Метод рівномірного пошуку	0,667	0,333	0,182	0,125	0,095

З іншого боку, можна також порівняти кількість обчислень значення функції, що необхідні для досягнення заданої величини відносного зменшення інтервалу або заданого ступеня точності. Якщо величина $K_L(n) = E$ задана, то значення n обчислюється за наступними формулами:

– для методу розподілу інтервалу навпіл
 $n = 2 \ln(E)/\ln(0,5)$;

– для методу золотого перетину
 $n = l + [\ln(E)/\ln(0,618)]$;

– для методу рівномірного пошуку
 $n = (2/E) - 1$.

У табл. 2.2 наведені дані про кількість обчислень значень функції, яка необхідна для визначення координати точки мінімуму з заданою точністю.

Таблиця 2.2 – Необхідна кількість обчислень значень функції

Метод пошуку	Задана точність			
	$E = 0,1$	$E = 0,05$	$E = 0,01$	$E = 0,001$
Метод розподілу інтервалу навпіл	7	9	14	20
Метод золотого перетину	5	8	11	16
Метод рівномірного пошуку	19	36	199	1999

Слід зазначити, що метод золотого перетину виявляється більш ефективним порівняно з іншими двома методами, оскільки він вимагає найменшого числа оцінювань значення функції для досягнення однієї і тієї самої заданої точності.

Контрольні запитання та задачі до розділу 2

2.1. Який порядок встановлення границь інтервалу невизначеності, що містить точку екстремуму?

2.2. У чому сутність пошуку екстремуму методом вилучення інтервалів? З яких етапів він складається?

2.3. У чому сутність і порядок скорочення інтервалу невизначеності методом розподілу інтервалу навпіл? Те ж, щодо методу золотого перетину?

2.4. В чому полягає ефективність методів вилучення інтервалів?

2.5. У чому сутність і порядок скорочення інтервалу невизначеності методом золотого перетину?

2.6. Визначте границі інтервалу функції $f(x) = (N \times 10 - 2x)^2$ при заданій початковій точці $x_0 = \frac{N \times 10}{M}$ і величині кроку $\Delta = \frac{N \times 10}{3 \times M}$.

N – остання цифра номера студентського квитка; якщо вона дорівнює нулю прийняти $N = 15$.

Якщо передостання цифра номера студентського квитка непарна, прийняти $M = 5$.

Якщо передостання цифра номера студентського квитка парна або нуль, прийняти $M = 4$.

У розрахунках отримані результати округлювати до трьох значущих цифр.

2.7. Визначте методом розподілу інтервалу навпіл мінімальні границі інтервалу, в якому знаходиться точка мінімуму функції $f(x)$.

Зробіть триразову процедуру зменшення інтервалу пошуку з метою отримання мінімальних границь інтервалу, в якому знаходиться точка мінімуму функції $f(x)$.

$$D = \begin{bmatrix} 0 & 7 & 5 & 3 & 9 & 13 & 8 & \infty \\ 7 & 0 & 2 & 4 & 6 & 10 & 8 & \infty \\ 5 & 2 & 0 & 2 & 4 & 8 & 6 & \infty \\ 3 & 4 & 2 & 0 & 6 & 10 & 5 & \infty \\ 9 & 6 & 4 & 6 & 0 & 10 & 10 & \infty \\ 13 & 10 & 8 & 10 & 10 & 0 & 7 & \infty \\ 8 & 8 & 6 & 5 & 10 & 7 & 0 & \infty \\ 18 & 15 & 13 & 15 & 9 & 12 & 10 & 0 \end{bmatrix} \quad (9.131)$$

і матрицю Γ

$$\Gamma = \begin{bmatrix} 1 & 4 & 4 & 4 & 4 & 4 & 4 & 8 \\ 4 & 2 & 3 & 3 & 3 & 3 & 3 & 8 \\ 4 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 3 & 3 & 4 & 3 & 3 & 7 & 8 \\ 4 & 3 & 3 & 3 & 5 & 6 & 3 & 8 \\ 4 & 3 & 3 & 3 & 5 & 6 & 7 & 8 \\ 4 & 3 & 3 & 4 & 3 & 6 & 7 & 8 \\ 5 & 5 & 5 & 5 & 5 & 5 & 7 & 8 \end{bmatrix} \quad (9.132)$$

Рішення, отримане на ітерації 5, не поліпшується на наступних ітераціях. Отже, оптимальне рішення: $D = D^5$, $\Gamma = \Gamma^5$.

Нехай потрібно визначити найкоротший шлях d_{15} . Довжина шляху $d_{15} = d_{15}^5 = 9$. Для знаходження відповідної послідовності вершин, що складають шлях, звернемося до матриці Γ . Оскільки $\gamma_5 = \gamma_{15}^5 = 4$, то вершина 4 є першою проміжною в найкоротшому шляху з вершини 1 до вершини 5.

Потім знаходимо наступну вершину на шляху до вершини 5, звертаючись до елемента $\gamma_{45} = 3$, далі $\gamma_{35} = 5$, отже послідовність вершин шляху μ_{15} визначена:

$$\mu_{15} = \{1, 4, 3, 5\}.$$

Аналогічно визначається решта найкоротших шляхів. Наприклад, довжина шляху $\mu_{82} = d_{82} = 15$, шлях складається із вершин:

$$\mu_{82} = \{8, 5, 3, 2\}.$$

Алгоритм, побудований на застосуванні операцій бульової алгебри.

Алгоритм пошуку найкоротших за рангом (транзитності) шляхів базується на застосуванні операцій бульової алгебри (логічних операцій) до матриці сполучення графа [22]. Пошук найкоротших шляхів в алгоритмі здійснюється за матрицею $V' = [\beta'_{ij}]$, яка знайдена шляхом перетворення вихідної матриці зв'язності $V = [\beta_{ij}]$. У матриці V' ненульовими стають лише елементи,

того, не розглядаються елементи головної діагоналі. В результаті необхідно дослідити елементи $d_{13}^1, d_{14}^1, d_{15}^1, d_{31}^1, d_{34}^1, d_{41}^1, d_{43}^1, d_{45}^1, d_{51}^1, d_{53}^1, d_{54}^1$.

Нові оцінки отримують лише такі елементи:

$$d_{13}^2 = \min(d_{13}^1; d_{12}^1 + d_{23}^1) = \min(\infty; 9 + 2) = 11;$$

$$d_{15}^2 = \min(d_{15}^1; d_{12}^1 + d_{25}^1) = \min(\infty; 9 + 7) = 16;$$

$$d_{31}^2 = \min(d_{31}^1; d_{32}^1 + d_{21}^1) = \min(\infty; 2 + 9) = 11;$$

$$d_{45}^2 = \min(d_{45}^1; d_{42}^1 + d_{25}^1) = \min(\infty; 12 + 7) = 19;$$

$$d_{51}^2 = \min(d_{51}^1; d_{52}^1 + d_{21}^1) = \min(\infty; 7 + 9) = 16;$$

$$d_{54}^2 = \min(d_{54}^1; d_{52}^1 + d_{24}^1) = \min(\infty; 7 + 12) = 19.$$

$$\text{Отже } \gamma_{13}^2 = \gamma_{15}^2 = \gamma_{31}^2 = \gamma_{45}^2 = \gamma_{51}^2 = \gamma_{54}^2 = 2.$$

Отримаємо матрицю D^2

$$D^2 = \begin{bmatrix} 0 & 9 & 11 & 3 & 16 & \infty & \infty & \infty \\ 9 & 0 & 2 & 12 & 7 & \infty & \infty & \infty \\ 11 & 2 & 0 & 2 & 4 & 8 & 6 & \infty \\ 3 & 12 & 2 & 0 & 19 & \infty & 5 & \infty \\ 16 & 7 & 4 & 19 & 0 & 10 & \infty & \infty \\ \infty & \infty & 9 & \infty & 10 & 0 & 7 & \infty \\ \infty & \infty & 6 & 5 & \infty & 7 & 0 & \infty \\ \infty & \infty & \infty & \infty & 9 & 12 & 10 & 0 \end{bmatrix} \quad (9.129)$$

і матрицю Γ^2

$$\Gamma^2 = \begin{bmatrix} 1 & 2 & 3 & 4 & 2 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 1 & 3 & 4 & 2 & 6 & 7 & 8 \\ 2 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix} \quad (9.130)$$

Виконуючи аналогічні обчислення на ітераціях 3, 4, 5, 11, 7, 8, отримаємо результат – матриця D

Задано функцію $f(x) = (N \times 10 - 2x)^2$ в інтервалі $\frac{N \times 10}{M} \leq x \leq K \times N \times 10$.

N – остання цифра номера студентського квитка; якщо вона дорівнює нулю, прийняти $N = 15$.

Якщо передостання цифра номера студентського квитка непарна, прийняти $K = 2$, а $M = 5$.

Якщо передостання цифра номера студентського квитка парна або нуль, прийняти $K = 3$, а $M = 4$.

2.8. Визначте методом золотого перетину мінімальні границі інтервалу, в якому знаходиться точка мінімуму функції $f(x)$.

Задані ті самі умови і вихідні дані, що і в завданні 2.2.

2.9. За рахунок чого забезпечується ефективність визначення границь інтервалу за методом Свенна?

2.10. Знайдіть точку мінімуму функції

$$f(x) = (10x^3 + 3x^2 + x + 5)^2.$$

Задано: початкова точка $x = 2$ і довжина кроку $\Delta = 0,5$;

а) визначте границі, у яких знаходиться точка екстремуму;

б) виконайте три ітерації та знайдіть інтервал невизначеності за методом вилучення інтервалів та за методом золотого перетину.

2.11. Реалізуйте процедуру одномірного пошуку точки екстремуму функції

$$f(x) = 3x^2 + 12/x^3 - 5 \text{ в інтервалі } 0,5 \leq x \leq 2,5,$$

використовуючи: а) метод розподілу інтервалу навпіл; б) метод золотого перетину.

У кожному випадку проведіть по три ітерації. Порівняйте результуючі інтервали пошуку, що отримані за допомогою перерахованих вище методів.

2.12. Користуючись кожним з методів одномірного пошуку, мінімізуйте наступні функції з точністю до одного знака після коми:

а) $f(x) = 3x^4 + (x - 1)^2$ в інтервалі $[0, 4]$;

б) $f(x) = (4x)(\sin x)$ в інтервалі $[0, \pi]$;

в) $f(x) = 2(x - 3)^2 + e^{0,5x^2}$ в інтервалі $[0, 100]$.

Література до розділу 2

1 Захарченко Н.В. Оптимизация и моделирование систем связи: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ООИС им. А.С. Попова, 1988. – Ч. 1. – 86 с.

2 Реклейтис Г. Оптимизация в технике: В 2-х кн. – Кн.1 / Реклейтис Г., Рейвиндран А., Рэгсдел К. – М.: Мир, 1986. – 352 с.

3 Балан М.М. Методи пошуку точки екстремуму всередині заданого інтервалу: метод. вказівки до проведення практичного заняття з дисципліни “Методи оптимізації” / М.М. Балан – Одеса: ОНАЗ ім. О.С. Попова, 2004. – 16 с.

3 МЕТОДИ ПОШУКУ ТОЧКИ ЕКСТРЕМУМУ З ВИКОРИСТАННЯМ ПОХІДНИХ

Загальні положення

Усі розглянуті в попередніх розділах методи пошуку ґрунтуються на припущеннях про унімодальність й у ряді випадків про безперервність досліджуваної цільової функції. Доцільно припустити, що, якщо на додаток до умови безперервності ввести вимогу диференційованості функції, то ефективність пошукових процедур можна істотно підвищити. Нагадаємо, що у розд. 1 установлено необхідну умову існування локального мінімуму функції в деякій точці x^* , відповідно до якого перша похідна функції в точці x^* повинна перетворюватися в нуль, тобто

$$f'(x^*) = df/dx|_{x=x^*} = 0.$$

Якщо функція $f(x)$ містить члени, що включають x у третій і більш високих ступенях, то безпосереднє вирішення аналітичного розв'язання рівняння $f'(x) = 0$ може виявитися скрутним. У таких випадках використовуються наближені методи послідовного пошуку стаціонарної точки функції f . Насамперед розглянемо класичну пошукову схему, орієнтовану на перебування кореня нелінійного рівняння. Ця схема була розроблена Ньютоном і пізніше уточнена Рафсоном [1].

3.1 Метод Ньютона-Рафсона

У рамках схеми Ньютона-Рафсона:

1) передбачається, що функція f двічі диференційована.

2) робота алгоритму починається в точці x_1 , що являє початкове наближення (або початкову оцінку) координати стаціонарної точки, або кореня рівняння $f'(x) = 0$. Потім будується лінійна апроксимація функції $f'(x)$ в точці x_1 , а точка, у якій апроксимуюча лінійна функція перетворюється в нуль, приймається як наступне наближення. Якщо точка x_k прийнята як поточне наближення до стаціонарної точки, то лінійна функція, що апроксимує функцію $f'(x)$ в точці x_k , записується у виді

$$\tilde{f}'(x; x_k) = f'(x_k) + f''(x_k)(x - x_k). \quad (3.1)$$

У точці $x_k = x_{k+1}$ $\tilde{f}'(x; x_k) = 0$

Прирівняємо праву частину рівняння (3.1) нулеві, одержимо наступне наближення:

$$x_{k+1} = x_k - [f'(x_k) / f''(x_k)].$$

Рис. 3.1 ілюструє основні кроки реалізації методу Ньютона-Рафсона.

У залежності від вибору початкової точки і виду функції алгоритм може як сходитися до істинної стаціонарної точки, так і розходитися, що показано на

змін. Елементи матриці Γ^1 при цьому: $\gamma_{24}^1 = 1$ і $\gamma_{42}^1 = 1$; ($q = 1$), решта елементів Γ^1 дорівнює відповідним елементам матриці Γ^0 .

	1	2	3	4	5	6	7	8	
1		9		3					Базовий рядок
2	9	0		∞					
3									
4	3	∞		0					
5									
6									
7									
8									

Базовий стовпець

Рисунок 9.33 – Елемент матриці D

Отримаємо матрицю D^1

$$D^1 = \begin{bmatrix} 0 & 9 & \infty & 3 & \infty & \infty & \infty & \infty \\ 9 & 0 & 2 & \infty & 7 & \infty & \infty & \infty \\ \infty & 2 & 0 & 2 & 4 & 8 & 6 & \infty \\ 3 & 12 & 2 & 0 & \infty & \infty & 5 & \infty \\ \infty & 7 & 4 & \infty & 0 & 10 & \infty & \infty \\ \infty & \infty & 8 & \infty & 10 & 0 & 7 & \infty \\ \infty & \infty & 6 & 5 & \infty & 7 & 0 & \infty \\ \infty & \infty & \infty & \infty & 9 & 12 & 10 & 0 \end{bmatrix} \quad (9.127)$$

і матрицю Γ^1

$$\Gamma^1 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix} \quad (9.128)$$

Ітерація 2. Визначаємо вершину 2 як базову й вилучаємо з матриці D^1 другий рядок і другий стовпець. Оскільки у другому рядку в стовпцях 11, 7 і 8 містяться ∞ , відповідні стовпці не розглядаються. Через те що у другому стовпці в рядках 11, 7 і 8 містяться ∞ , відповідні рядки не розглядаються. Крім

та матриці Γ^0

$$\Gamma^0 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix} \quad (9.126)$$

Ітерація 1. Вершина $k = 1$ визначається як базова, отже в матриці D^0 вилучаємо перший рядок і перший стовпець. Оскільки в базовому рядку у стовпцях 3, 5, 6, 7, 8 містяться елементи, що дорівнюють ∞ , то елементи цих стовпців можна не досліджувати відповідно до виразу (9.123).

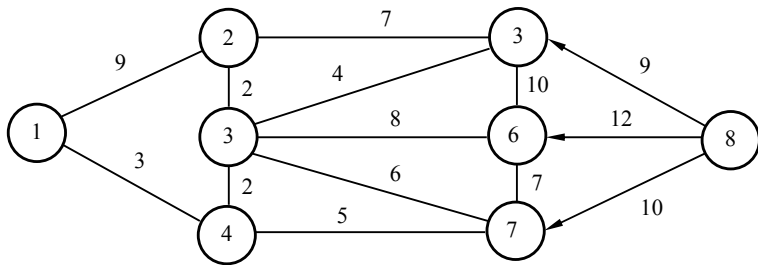


Рисунок 9.32 – Граф мережі (до пояснення алгоритму Флойда)

У базовому першому стовпці у рядках 3, 5, 6, 7, 8 знаходяться елементи, що дорівнюють ∞ , тому згідно з виразом (9.123) елементи цих рядків можна не досліджувати. Для того щоб визначити чи призведе використання вершини 1 до більш коротких шляхів, ніж ті, котрі записані в D^0 , необхідно дослідити за допомогою тримісної операції (9.123) лише елементи матриці D^0 , зображені на рис. 9.33, а саме елементи $d_{24}^0, d_{42}^0, d_{44}^0$. Оскільки елементи головної діагоналі ($i = j$) можна не розглядати, застосування операції (9.123) дає такі результати:

$$d_{24}^1 = \min(d_{24}^0; d_{21}^0 + d_{14}^0) = \min(\infty; 9 + 3) = 12;$$

$$d_{42}^1 = \min(d_{42}^0; d_{41}^0 + d_{12}^0) = \min(\infty; 3 + 9) = 12.$$

Через те що $d_{24}^1 = 12 < d_{24}^0 = \infty$ і $d_{42}^1 = 12 < d_{42}^0 = \infty$, зазначені елементи заносяться в матрицю D^1 , решта елементів матриці D^0 переноситься в D^1 без

рис. 3.2. Якщо початкова точка розташована правіше x_0 , то одержані в наслідок послідовних наближень точки віддаляються від стаціонарної точки x_1 .

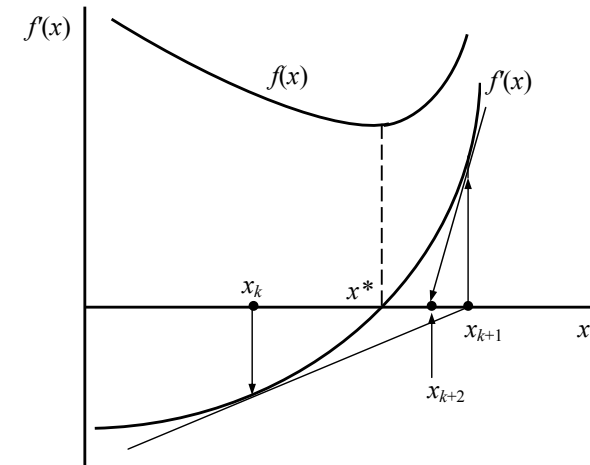


Рисунок 3.1 – Метод Ньютона-Рафсона (збіжність)

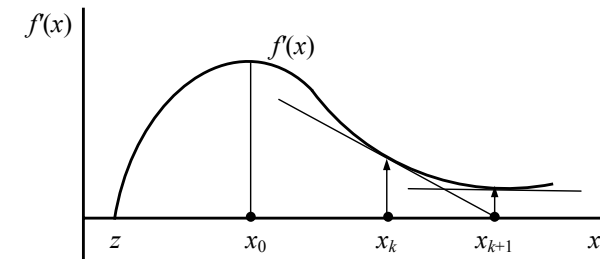


Рисунок 3.2 – Метод Ньютона-Рафсона (відсутність збіжності)

3.2 Метод середньої точки

Якщо функція $f(x)$ унімодальна у заданому інтервалі пошуку, то точкою екстремуму є точка, у якій $f'(x) = 0$. Якщо при цьому є можливість обчислювати як значення функції, так і її похідну, то для знаходження кореня рівняння $f'(x) = 0$ можна скористатися ефективним алгоритмом вилучення інтервалів, на кожній ітерації якого розглядається лише одна спробна точка. Наприклад, якщо в точці x_1 виконується нерівність $f'(x_1) > 0$, то з урахуванням припущення про унімодальність природно стверджувати, що точка мінімуму не

може знаходитися правіше x_1 і інтервал $x \geq x_1$ можна вилучити. Приведені міркування лежать в основі логічної структури методу середньої точки, що іноді називають *пошуком Больцано* [1].

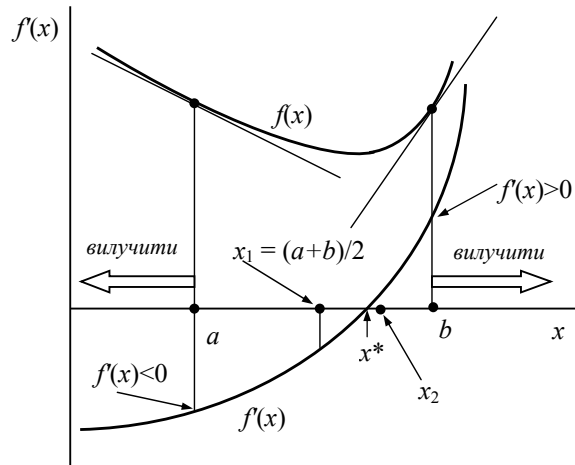


Рисунок 3.3 – Метод середньої точки

Визначимо дві точки a і b таким чином, що $f'(a) < 0$ і $f'(b) > 0$. Стационарна точка розташована між a і b . Обчислимо значення похідної функції в середній точці розглянутого інтервалу $x_1 = (a + b)/2$. Якщо $f'(x_1) > 0$, то інтервал (x_1, b) слід вилучити з інтервалу пошуку. З іншого боку, якщо $f'(x_1) < 0$ (див. рис. 3.3), то можна вилучити інтервал (a, x_1) . Нижче дається формалізований опис кроків алгоритму.

Нехай маємо обмежений інтервал $a \leq x \leq b$ і задану заздалегідь установлену величину припустимого відхилення – параметр ϵ .

Крок 1. Початковий інтервал (a, b) ; при цьому $f'(a) < 0$ і $f'(b) > 0$.

Крок 2. Обчислити $x_1 = (a + b)/2$ і $f'(x_1)$.

Крок 3. Якщо $|f'(x_1)| < \epsilon$ – пошук закінчено, можна визначити точку x_1 точкою екстремуму. У протилежному випадку ($|f'(x_1)| > \epsilon$), якщо $f'(x_1) > 0$, покласти $b = x_1$ і перейти до кроку 2. Якщо $f'(x_1) < 0$, покласти $a = x_1$ і перейти до кроку 2.

Слід зазначити, що логічна структура пошуку відповідно до викладеного методу вилучення інтервалів заснована лише на дослідженні знака похідної незалежно від значень, які ця похідна приймає. У наступному підрозділі розглядається метод січних, при реалізації якого враховуються як знак похідної, так і її значення.

Алгоритм починає роботу при $\tilde{A}^0 = [\gamma_{ij}^0]$, де $\gamma_{ij}^0 = j$. На q -й ітерації вершина γ_{ij}^q може бути отримана із співвідношення:

$$\gamma_{ij}^q = \begin{cases} k, & \text{якщо } d_{ij}^{q-1} > d_{ik}^{q-1} > d_{kj}^{q-1}, \\ \gamma_{ij}^{q-1} & \text{у протилежному випадку.} \end{cases}$$

Після побудови матриць D^0 та Γ^0 слід для кожного $k = 1, 2, \dots, n$ використовувати для обчислень елементи матриці D^{q-1} та з урахуванням $(q - 1)$ -ї ітерації виконати таку процедуру.

Крок 1. Вилучити елементи k -го рядка і k -го стовпця. Назвемо ці множини елементів базовим рядком і базовим стовпцем.

Крок 2. Для кожного елемента d_{ij}^{q-1} ($j \neq k$) матриці D^{q-1} , починаючи з першого, розміщеного в лівому верхньому куті матриці, перевірити виконання нерівності

$$d_{ik}^{q-1} + d_{kj}^{q-1} \geq d_{ij}^{q-1}. \quad (9.124)$$

Якщо нерівність (9.124) виконується, тоді вибрати нові значення i та j й знову виконати крок 2. Якщо нерівність (9.120) не виконується, то елементу d_{ij}^q матриці D^q присвоїти значення $d_{ij}^q = d_{ik}^{q-1} + d_{kj}^{q-1}$, а елементу γ_{ij}^q матриці \tilde{A}^q присвоїти значення $\gamma_{ij}^q = k$.

Після перевірки нерівності (9.124) для усіх елементів (i, j) знову перейти до виконання кроку 1 при $q = q + 1$ до $q = n$.

У відповідності з виразом (9.123) при отриманні D^q з D^{q-1} тримісну операцію, яка дозволяє встановити факт приналежності базового вузла k до найкоротшого шляху, слід виконувати лише з елементами матриці D^{q-1} , що не належать ні до базового рядка, ні до базового стовпця.

Для ілюстрації роботи алгоритму Флойда визначимо найкоротші шляхи між усіма парами вершин графа (рис. 9.32).

Початкові значення елементів матриці D^0

$$D^0 = \begin{bmatrix} 0 & 9 & \infty & 3 & \infty & \infty & \infty & \infty \\ 9 & 0 & 2 & \infty & 7 & \infty & \infty & \infty \\ \infty & 2 & 0 & 2 & 4 & 8 & 6 & \infty \\ 3 & \infty & 2 & 0 & \infty & \infty & 5 & \infty \\ \infty & 7 & 4 & \infty & 0 & 10 & \infty & \infty \\ \infty & \infty & 8 & \infty & 10 & 0 & 7 & \infty \\ \infty & \infty & 6 & 5 & \infty & 7 & 0 & \infty \\ \infty & \infty & \infty & \infty & 9 & 12 & 10 & 0 \end{bmatrix} \quad (9.125)$$

Рис. 9.31 ілюструє виконання процедури порівняння (9.122) для пари вершин i та j .

Алгоритм Флойда дозволяє розв'язати задачу отримання найкоротших шляхів між усіма вершинами графа для графа з n вершин за n ітерацій.

Позначимо символом d_{ij}^q оцінку довжин найкоротшого шляху з i в j , отриману на q -й ітерації.

Спосіб отримання на q -й ітерації величини d_{ij}^q , може бути описаний за допомогою такої тримісної операції (операції трикутника):

$$d_{ij}^q = \min(d_{ij}^{q-1}, d_{ik}^{q-1} + d_{kj}^{q-1}), \quad i \neq k \neq j. \quad (9.123)$$

Якщо операцію (9.123) виконувати з даною парою вершин i та j й усіма вершинами ($i \neq k \neq j$) у порядку зростання номерів k , то значення d_{ij}^n , що отримане на останній ітерації, буде дорівнювати довжині найкоротшого шляху з вершини i у вершину j .

На кожній ітерації алгоритму будуються дві матриці. Перша з них, яка називається матрицею довжин найкоротших шляхів, містить поточні оцінки довжин найкоротших шляхів, тобто на q -й ітерації дана матриця визначається як $D^q = [d_{ij}^q]$.

Алгоритм починає роботу при $D^0 = [d_{ij}^0] = [l_{ij}]$. Потім, виконуючи тримісну операцію (9.123) з усіма елементами матриці D^0 , отримаємо D^1 і т. ін, і так доти, поки для кожної пари вершин не буде виконаний критерій оптимальності (9.122).

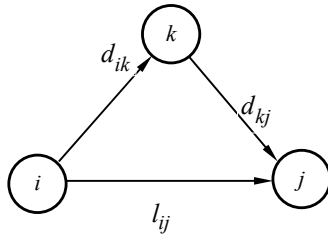


Рисунок 9.118 – Ілюстрація виконання процедури порівняння

Друга матриця, яка називається матрицею маршрутів (позначимо її Γ), використовується для знаходження проміжних вершин (якщо такі є) найкоротших шляхів.

На q -й ітерації вона визначається як $\tilde{A}^q = [\gamma_{ij}^q]$, де γ_{ij}^q – перша проміжна вершина найкоротшого шляху з i в j , яка вибирається з вершин множини $\{1, 2, \dots, k\}$ ($i \neq k \neq j$).

3.3 Метод січних

Метод січних є комбінацією методу Ньютона-Рафсона і загальної схеми вилучення інтервалів та орієнтований на знаходження кореня рівняння $f(x) = 0$ у інтервалі (a, b) , якщо, зрозуміло, такий корінь існує [1, 2].

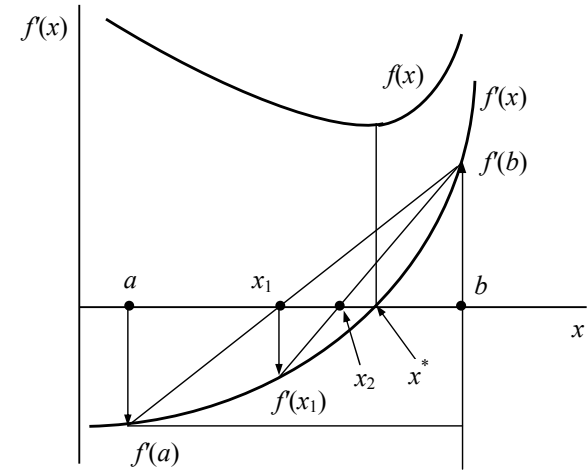


Рис. 3.4 Метод січних

Припустимо, що в процесі пошуку стаціонарної точки функції $f(x)$ в інтервалі (a, b) виявлені дві точки a і b , у яких знаки похідної різні. У цьому випадку алгоритм методу січних дозволяє апроксимувати функцію $f'(x)$ «січною прямою» (прямою лінією, що з'єднає дві точки) і знайти точку, у якій січна графіка $f'(x)$ перетинає вісь абсцис (рис. 3.4). Таким чином, наступне наближення до стаціонарної точки x^* визначається із співвідношення

$$\frac{b-a}{f'(b)-f'(a)} = \frac{b-x_1}{f'(b)}$$

за формулою

$$x_1 = b - \frac{(b-a)f'(b)}{f'(b)-f'(a)}$$

Якщо $|f'(x_1)| \leq \varepsilon$, пошук слід закінчити. У протилежному випадку необхідно вибрати одну з точок a або b таким чином, щоб знаки похідної в цій точці і точці x_1 були різні, а потім повторити основний крок алгоритму. Наприклад, у ситуації, що показана на рис. 3.4, у якості двох наступних точок мають бути обрані точки x_1 і b . Легко бачити, що на відміну від методу середньої точки метод січних заснований на дослідженні не тільки знака, але і значень похідної у спробних точках і тому в ряді випадків дозволяє вилучити більш половини інтервалу пошуку (див. рис. 3.4).

Контрольні питання та задачі до розділу 3

3.1 Що є спільним у різних методах пошуку екстремуму з використанням похідних?

3.2 В чому полягає відмінність різних методів пошуку екстремуму з використанням похідних?

3.3 Знайдіть дійсні корені рівняння (з точністю до одного знака після коми)

$$f(x) = 3000 - 100x^2 - 4x^5 - 6x^6 = 0,$$

використовуючи:

- а) метод Ньютона-Рафсона;
- б) метод середньої точки;
- в) метод січних.

3.4 Реалізуйте процедуру одномірного пошуку точки екстремуму функції

$$f(x) = 3x^2 + 12/x^3 - 5 \text{ в інтервалі } 0,5 \leq x \leq 2,5,$$

використовуючи методи пошуку з використанням похідних.

У кожному випадку проведіть по три ітерації. Порівняйте результуючі інтервали пошуку, що отримані за допомогою перерахованих вище методів.

3.5 У результаті експериментів встановлено, що траєкторія руху космічного тіла може бути описана наступним рівнянням:

$$f(x) = 4x^3 - 3x^2 + 2x + e^{x/2}.$$

Знайдіть корінь рівняння $f(x) = 0$ за допомогою кожного з методів із використанням похідних.

Література до розділу 3

1 Реклейтис Г. Оптимизация в технике: В 2-х кн. – Кн.1 / Реклейтис Г., Рейвиндран А., Рэгсдел К. – М.: Мир, 1986. – 352 с.

2 Захарченко Н.В. Оптимизация и моделирование систем связи: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, 1988. – Ч. 1.– 86 с.

$$R = \begin{bmatrix} 0 & 1 & \infty & 1 & \infty \\ 1 & 0 & 1 & 1 & 1 \\ \infty & 1 & 0 & \infty & 1 \\ 1 & 1 & \infty & 0 & 1 \\ \infty & 1 & 1 & 1 & 0 \end{bmatrix}. \quad (9.121)$$

Усі правила, розглянуті для отримання матриці L_{opt} , справедливі й для матриці R_{opt} , а саме $R_{\text{opt}} = R^{(R_{\text{max}})}$ і, крім того, якщо для певного степеня q справедливе $R^q = R^{(q-1)}$, то з цього виходить, що $R_{\text{opt}} = R^q = R^{(q-1)}$.

Зазначимо також, що при піднесенні матриці $R^{(1)}$ до будь-якого степеня q операцію $R^{(q)} = R^{(q-1)}R^{(1)}$ слід застосовувати лише до тих елементів матриці $R^{(q-1)}$, які ще не визначені, тобто $r_{ij}^{(q-1)} = \infty$.

Як тільки при піднесенні до чергового степеня q матриці $R^{(1)}$ усі елементи $r_{ij}^{(q)} \neq \infty$ ($i, j = \overline{1, n}$, граф G – зв'язний), процес обчислення слід припинити.

Отримана матриця $R^q = R_{\text{opt}}$.

Слід відзначити, що розглянутий метод дозволяє визначити довжини (ранги, вартості і т. ін) найкоротших шляхів між усіма вершинами графа, але не вказує на ті гілки, що входять до найкоротших шляхів.

Алгоритм Флойда [17, 21] дає можливість отримати як матрицю довжин найкоротших відстаней $L_{\text{opt}} = D$, так і маршрутну матрицю, що визначає вершини (гілки) графа, які складають шляхи. Спочатку за довжину найкоротшого шляху між довільними вершинами i та j приймається довжина гілки I_{ij} , що з'єднує ці вершини. Потім послідовно перевіряються проміжні вершини, що розміщені між i та j . Якщо довжина шляху, який проходить через певну проміжну вершину k , менша за попереднє (змінне) значення, яке було позначене, наприклад, d_{ij} , то змінній d_{ij} присвоюється нове значення.

Зазначимо, що вихідне значення $d_{ij} = I_{ij}$. Дана процедура повторюється для будь-яких пар вершин, поки не будуть отримані всі значення матриці $L_{\text{opt}} = D$.

Для будь-яких трьох різних вершин j, k, l сформульовані умови можуть бути записані у вигляді нерівності

$$d_{ik} + d_{kj} \geq d_{ij} \quad (i \neq k \neq j), \quad (9.122)$$

оскільки інакше найкоротший шлях з вершини i у вершину j повинен мати вершину k і тоді величина d_{ij} не дорівнювала б довжині найкоротшого шляху.

$$L^{(2)} = \begin{bmatrix} 0 & 50 & 90 & 80 & 105 \\ 50 & 0 & 40 & 90 & 100 \\ 90 & 40 & 0 & 85 & 60 \\ 80 & 90 & 85 & 0 & 25 \\ 105 & 100 & 60 & 25 & 0 \end{bmatrix}.$$

Через те, що $L^{(2)} \neq L^{(1)}$ і показник степеня "2" $< R_{\max} = 4$, процес обчислення L_{opt} слід продовжити, обчисливши $L^{(3)} = L^{(2)}L^{(1)}$. Елемент матриці $I_{12}^{(3)} = \min_k (I_{ik}^{(2)} + I_{k2}^{(2)}) = \min(I_{11}^{(2)} + I_{12}^{(2)} + I_{12}^{(2)} + I_{22}^{(2)} + I_{13}^{(2)} + I_{32}^{(2)} + I_{14}^{(2)} + I_{42}^{(2)} + I_{15}^{(2)} + I_{52}^{(2)}) = \min(0+50, 50+0, 90+40, 80+90, 105+130) = 50$.

Обчисливши решту елементів матриці $L^{(3)}$, отримаємо

$$L^{(3)} = \begin{bmatrix} 0 & 50 & 90 & 80 & 105 \\ 50 & 0 & 40 & 90 & 100 \\ 90 & 40 & 0 & 85 & 60 \\ 80 & 90 & 85 & 0 & 25 \\ 105 & 100 & 60 & 25 & 0 \end{bmatrix}.$$

Із порівняння $L^{(3)}$ та $L^{(2)}$ виходить, що $L^{(3)} = L^{(2)}$. Таким чином, $L_{\text{opt}} = L^{(3)} = L^{(2)}$.

Матрицю L_{opt} називають також дистанційною і позначають $D = [d_{ij}]$ [19].

Процедуру (9.119) можна застосувати для отримання шляхів мінімальної вартості, часу передачі інформації і подібних адитивних критеріїв оптимальності шляхів, надаючи елементам I_{ij} матриці L відповідне розуміння вартості, часу передачі інформації по гілці (i, j) та інших характеристик.

Для отримання матриці шляхів мінімального рангу R_{opt} , між усіма вершинами графа для вихідного графа необхідно побудувати матрицю безпосередніх зв'язків, або матрицю шляхів першого рангу $R = [r_{ij}]$. Матриця R будується аналогічно матриці L , однак замість довжини гілки зв'язку у відповідній позиції записується "1":

$$r_{ij} = \begin{cases} 0, & \text{якщо } i = j; \\ 1, & \text{якщо між } i \text{ та } j \text{ є безпосередній зв'язок}; \\ \infty, & \text{якщо між } i \text{ та } j \text{ зв'язку немає.} \end{cases}$$

Для графа (див. рис. 9.29) матриця R має вид:

4 МЕТОДИ ПОШУКУ ЕКСТРЕМУМУ ФУНКЦІЙ БАГАТЬОХ ЗМІННИХ

4.1 Необхідні умови екстремуму. Аналітичний метод знаходження екстремуму

Нехай f – функція $n + 1$ змінних $f(x_0, x_1, \dots, x_n)$. Її область визначення x породжується сукупністю областей визначення x_0, x_1, \dots, x_n , по кожній з $n + 1$ змінних; при цьому $x_k \in X_k$. Для визначеності будемо покладати, що x співпадає з дійсною віссю, а саме, $-\infty < x_k < \infty$, $k = 0, 1, \dots, n$ (рис. 4.1).

Тут і надалі у тексті для наочності, будемо користуватися рисунками функції (поверхні) $f(x_0, x_1, \dots, x_n)$.

Однак можливості наочного геометричного сприйняття багатомірних поверхонь, як відомо, достатньо обмежені. Тому у всіх розглянутих далі випадках будуть надані рисунки тільки двовимірних поверхонь.

Для функції $n + 1$ змінних поняття глобального і локального максимумів (мінімумів) вводяться аналогічно поданим раніше визначенням для функції однієї змінної. Варто тільки мати на увазі, що тепер екстремальна точка x , характеризується набором координат $x_0^*, x_1^*, \dots, x_n^*$, (див. рис. 4.1, де $n = 1$), – тому поняття ε -окилу виглядає складніше. Тут і надалі для стислості точку $(n + 1)$ -мірного простору з координатами x_0, x_1, \dots, x_n будемо позначати x . Таким чином, на відміну від функції однієї змінної $f(x)$, що розглядалася в попередньому розділі, функція $f(x)$ залежить від $n+1$ змінних.

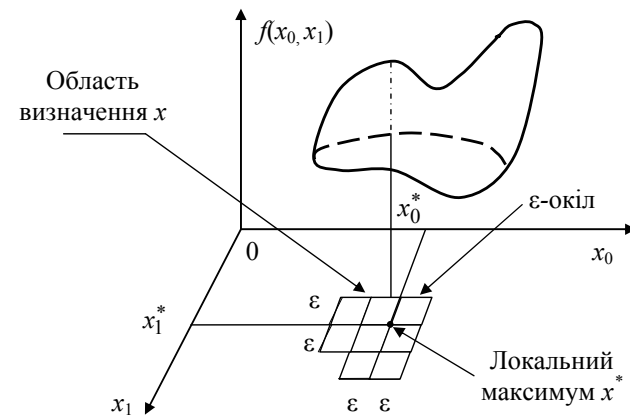


Рисунок 4.1 – ε -окил точки x^* функції двох змінних

Визначення 4.1. Множина точок x з координатами x_0, x_1, \dots, x_n , що задовольняють системі нерівностей

$$\begin{cases} |x_0 - x_0^*| \leq \varepsilon, \\ |x_1 - x_1^*| \leq \varepsilon, \\ \dots\dots\dots \\ |x_n - x_n^*| \leq \varepsilon, \end{cases} \quad (4.1)$$

утворює ε -окіл точки x з координатами x_0, x_1, \dots, x_n . Тому, наприклад, локальний екстремум визначається у такий спосіб (порівняємо з визначенням локального екстремуму функції однієї змінної у підрозділі 1.3).

4.2. Точка x з координатами x_0, x_1, \dots, x_n доставляє локальний максимум функції $f(x_0, x_1, \dots, x_n)$, якщо при деякому достатньо малому числі $\varepsilon > 0$ для всіх точок, що не збігаються зі згаданою точкою і задовольняють умові (4.1), виконується нерівність

$$f(x_0^*, x_1^*, \dots, x_n^*) \geq f(x_0, x_1, \dots, x_n), \quad (4.2)$$

Якщо дана нерівність є строгою, то говорять про строгий локальний максимум. Щоб дати визначення локальному мінімумові, досить у співвідношенні (4.2) змінити знак нерівності на знак протилежного розуміння (\leq).

Екстремум вважається глобальним, якщо нерівність (4.2) (або йому протилежна) виконується для всіх точок області визначення, а не тільки для тих, що належать ε -окилу (4.1).

Легко зрозуміти, що, якщо точка з координатами x_0, x_1, \dots, x_n є точкою локального екстремуму, то в ній буде досягтися локальний екстремум і за довільно обраною координатою x_j , якщо всі інші координати зафіксувати і покласти $x_k = x_k^*$, $k \neq j$. Отже, в екстремальній точці x^* з координатами $x_0^*, x_1^*, \dots, x_n^*$ має виконуватися необхідна умова екстремуму (теорема 3.1) стосовно до кожної незалежної змінної, узятій окремо. Отже, справедлива наступна теорема.

Теорема 4.1. Для того, щоб у точці x з координатами $x_0^*, x_1^*, \dots, x_n^*$ функція $f(x_0, x_1, \dots, x_n)$ мала локальний екстремум, необхідно, щоб усі її частинні похідні перетворювалися у цій точці в нуль [1]:

$$\left. \begin{aligned} \frac{\partial f(x_0, x_1, \dots, x_n)}{\partial x_j} &= 0, \quad j=0, 1, \dots, n \\ x_0 &= x_0^* \\ x_1 &= x_1^* \\ \dots & \\ x_n &= x_n^* \end{aligned} \right\} \quad (4.3)$$

За наявності декількох паралельних одно- і двотранзитних шляхів (рис. 9.30, б) для визначення довжини найкоротшого шляху необхідно операцію додавання у виразі (9.118) замінити на операцію вибору мінімуму із довжин усіх одно- та двотранзитних шляхів, тобто замість виразу (9.118) будемо мати

$$I_{ij}^{(2)} = \min_k (I_{ik}^{(1)} + I_{kj}^{(1)}) = \min[(I_{i1}^{(1)} + I_{1j}^{(1)}), (I_{i2}^{(1)} + I_{2j}^{(1)}), \dots, (I_{in}^{(1)} + I_{nj}^{(1)})]. \quad (9.119)$$

Таким чином, елемент $I_{ij}^{(2)}$ матриці $L^{(2)}$ дорівнює довжині найкоротшого шляху із i в j серед усіх одно- і двотранзитних шляхів.

При піднесенні матриці $L^{(1)}$ до степеня q з використанням операції (9.119) отримаємо матрицю $L^{(q)} = L^{(q-1)}L^{(1)}$, елемент якої

$$I_{ij}^{(q)} = \min_k (I_{ik}^{(q-1)} + I_{kj}^{(1)})$$

визначає довжину найкоротшого шляху між вершинами i та j серед усіх шляхів першого, другого, q -го рангу.

Отже, піднесення матриці $L^{(1)}$ до степеня максимального рангу R_{\max} призводить до отримання матриці найкоротших за довжиною шляхів між усіма парами графа, або матриці оптимальних шляхів

$$L_{\text{opt}} = L^{(R_{\max})}.$$

Якщо при піднесенні матриці $L^{(1)}$ до певного степеня q виявиться, що

$$L^{(q)} = L^{(q-1)}, \quad (9.120)$$

процес обчислень слід припинити, тому що при рівності (9.120) завжди виконується рівняння $L^{(q)} = L^{(q+1)}$.

Таким чином, $L_{\text{opt}} = L^{(q)} = L^{(q-1)}$. Для розглянутого прикладу (9.117) обчислимо $L^{(2)}$:

$$\begin{aligned} I_{12}^{(2)} &= \min(I_{11}^{(1)} + I_{12}^{(1)} + I_{12}^{(1)} + I_{22}^{(1)} + I_{13}^{(1)} + I_{32}^{(1)} + I_{14}^{(1)} + I_{42}^{(1)} + I_{15}^{(1)} + I_{52}^{(1)}) = \\ &= \min(50, 50, \infty, 170, \infty) = 50; \end{aligned}$$

$$\begin{aligned} I_{13}^{(2)} &= \min(I_{11}^{(1)} + I_{13}^{(1)} + I_{12}^{(1)} + I_{23}^{(1)} + I_{13}^{(1)} + I_{33}^{(1)} + I_{14}^{(1)} + I_{43}^{(1)} + I_{15}^{(1)} + I_{53}^{(1)}) = \\ &= \min(\infty, 90, \infty, \infty, \infty) = 90; \end{aligned}$$

$$\begin{aligned} I_{14}^{(2)} &= \min(I_{11}^{(1)} + I_{14}^{(1)} + I_{12}^{(1)} + I_{24}^{(1)} + I_{13}^{(1)} + I_{34}^{(1)} + I_{14}^{(1)} + I_{44}^{(1)} + I_{15}^{(1)} + I_{54}^{(1)}) = \\ &= \min(80, 140, \infty, 80, \infty) = 80 \quad \text{і т.ін.} \end{aligned}$$

Обчисливши решту елементів матриці $L^{(2)}$, одержимо:

Тоді матриця довжин L буде мати вид

$$L = \begin{bmatrix} 0 & 50 & \infty & 80 & \infty \\ 50 & 0 & 40 & 90 & 130 \\ \infty & 40 & 0 & \infty & 60 \\ 80 & 90 & \infty & 0 & 25 \\ \infty & 130 & 60 & 25 & 0 \end{bmatrix} \quad (9.117)$$

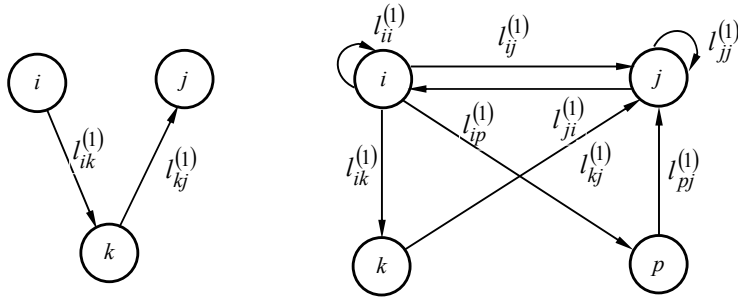
По суті, матриця L – це матриця відстаней безпосередніх зв'язків, тобто шляхів першого рангу. Позначимо матрицю $L^{(1)}$. Піднесемо матрицю $L^{(1)}$ до квадрата: $L^{(2)} = L^{(1)}L^{(1)}$, ij -й елемент матриці $L^{(2)}$ визначається за правилом

$$I_{ij}^{(2)} = \sum_{k=1}^n I_{ik}^{(1)} I_{kj}^{(1)} = I_{i1}^{(1)} I_{1j}^{(1)} + I_{i2}^{(1)} I_{2j}^{(1)} + \dots + I_{in}^{(1)} I_{nj}^{(1)}. \quad (9.118)$$

Інтерпретуючи тепер множення як послідовне, а додавання як паралельне з'єднання гілок (за аналогією зі степенуванням структурної матриці B , підрозділ 9.10) легко зрозуміти, що добуток $I_{ik}^{(1)} I_{kj}^{(1)}$ відповідає двотранзитному шляху (шляху другого рангу), який проходить із вершини i у вершину j через проміжну вершину k (рис. 9.30, а), а сума, наприклад, чотирьох добутоків

$$I_{ii}^{(1)} I_{ij}^{(1)} + I_{ij}^{(1)} I_{jj}^{(1)} + I_{ik}^{(1)} I_{kj}^{(1)} + I_{ip}^{(1)} I_{pj}^{(1)}$$

– чотирьом двотранзитним шляхам (рис. 9.30, б). Зауважимо, що добутки $I_{ii}^{(1)} I_{ij}^{(1)}$ і $I_{ij}^{(1)} I_{jj}^{(1)}$ фактично відповідають одотранзитним шляхам (тобто шляхам першого рангу, які мають лише одну гілку).



а) б)
Рисунок 9.30 – Ілюстрація степенування матриці L

Оскільки довжина шляху, який складається із декількох гілок, визначається сумою довжин гілок, тоді слід при множенні матриці $L^{(1)}$ операцію множення у виразі 9.118 замінити на операцію додавання, тобто замість $I_{ik}^{(1)} I_{kj}^{(1)}$ будемо мати $I_{ik}^{(1)} + I_{kj}^{(1)}$.

Цю умову, що називають умовою стаціонарності, часто записують у векторній формі:

$$\text{grad } f(x^*) = 0, \text{ або } \nabla f(x^*) = 0, \quad (4.4)$$

де символи $\text{grad } f(x)$ або $\nabla f(x)$ означають градієнт $f(x)$. Надамо визначення цього вектора.

Визначення 4.3. $(n + 1)$ -мірний вектор з координатами

$$\left[\frac{\partial f}{\partial x_0}, \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right] \quad (4.5)$$

називається градієнтом функції $f(x_0, x_1, \dots, x_n)$.

Точки $(n + 1)$ -мірного простору, що задовольняють умові (4.3) (або (4.4)), як і в випадку функції однієї змінної, називають стаціонарними. Зрозуміло, стаціонарна точка не обов'язково є екстремальною.

4.2 Достатні умови екстремуму

Приведемо без доказу достатню умову екстремуму функції багатьох змінних [1, 2, 3].

Теорема 4.2. Для того, щоб двічі безупинно диференційована функція $(n + 1)$ змінних $f(x)$ мала в стаціонарній точці локальний максимум (мінімум), достатньо, щоб матриця її других похідних (матриця Гессе)

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_0 \partial x_0} & \frac{\partial^2 f}{\partial x_0 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_0 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_0} & \frac{\partial^2 f}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_0} & \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} \quad (4.6)$$

була від'ємно- (позитивно-) визначеною, тобто щоб вираз

$$\sum_{k=0}^n \sum_{l=0}^n \frac{\partial^2 f}{\partial x_k \partial x_l} \xi_k \xi_l \quad (4.7)$$

був менше (більше) нуля при довільно обраних, дійсних числах $\xi_0, \xi_1, \dots, \xi_n$, таких, що

$$\xi_0^2 + \xi_1^2 + \dots + \xi_n^2 > 0.$$

Перевірка знаковизначеності матриці може бути проведена, наприклад, за допомогою критерію Сильвестра (27). Відповідно до цього критерію, необхідною і достатньою умовою позитивно-визначеності матриці

$$\begin{bmatrix} \alpha_{00} & \alpha_{01} & \dots & \alpha_{0n} \\ \alpha_{10} & \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n0} & \alpha_{n1} & \dots & \alpha_{nn} \end{bmatrix}, \quad (4.8)$$

а саме, виконання нерівності $\sum_{k=0}^n \sum_{i=0}^n \alpha_{ki} \xi_k \xi_i > 0$ при довільно обраних значеннях $\xi_0, \xi_1, \dots, \xi_n$ ($\xi_0^2 + \xi_1^2 + \dots + \xi_n^2 > 0$), що визначає в стаціонарній точці локальний мінімум, є виконання $n + 1$ нерівностей

$$\alpha_{00} > 0, \quad \begin{vmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{vmatrix} > 0, \quad \dots, \quad \begin{vmatrix} \alpha_{00} & \alpha_{01} & \dots & \alpha_{0n} \\ \alpha_{10} & \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n0} & \alpha_{n1} & \dots & \alpha_{nn} \end{vmatrix} > 0. \quad (4.9)$$

Необхідною і достатньою умовою від'ємно-визначеності матриці (4.8), що визначає в стаціонарній точці локальний максимум, є виконання нерівностей

$$(-1) \alpha_{00} > 0, \quad \begin{vmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{vmatrix} > 0, \quad \dots, \quad (-1)^{n+1} \begin{vmatrix} \alpha_{00} & \alpha_{01} & \dots & \alpha_{0n} \\ \alpha_{10} & \alpha_{11} & \dots & \alpha_{1n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n0} & \alpha_{n1} & \dots & \alpha_{nn} \end{vmatrix} > 0. \quad (4.10)$$

Визначення 4.4. Стаціонарну точку функції, багатьох змінних, що не є точкою екстремуму, називають сідловою, якщо в околі цієї точки (рис. 4.2) виконуються нерівності

$$f(x_0^*, x_1) \leq f(x_0^*, x_1^*) \leq f(x_0, x_1^*)$$

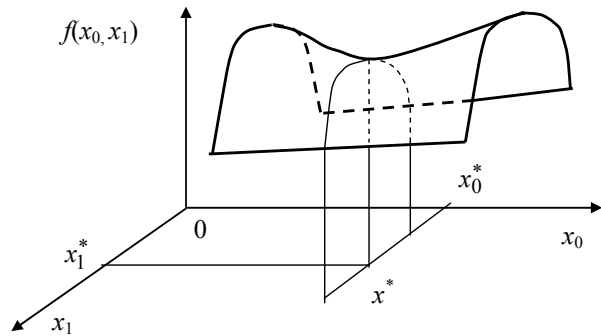


Рисунок 4.2 – Сідлова точка функції двох змінних

4.3 Застосування необхідних і достатніх умов екстремуму у задачі корекції спотворень каналу зв'язку

Приклад. Розглянемо застосування необхідних і достатніх умов екстремуму до задачі корекції спотворень каналу зв'язку. У випадку коректора з двома базисними функціями середньоквадратична похибка, як було раніш показано, приймає вигляд

$$D^2(g_0, g_1) = d - 2b_0g_0 - 2b_1g_1 + 2r_{01}g_0g_1 + r_{00}g_0^2 + r_{11}g_1^2. \quad (4.11)$$

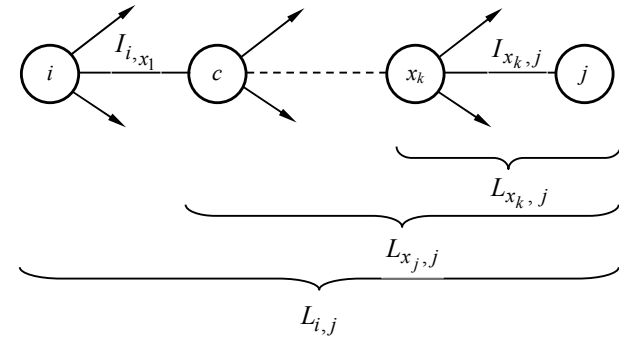


Рисунок 9.28 – Частини найкоротшого шляху

Легко зрозуміти, що для знаходження найкоротшого шляху від вершини x_0 до вершини x_m необхідно переглянути усі можливі шляхи й обрати той, що має екстремальну вагу.

Найбільш широко використовуються шляхи мінімальних довжини та рангу.

Нині існує ряд методів, які дозволяють упорядкувати процедуру визначення довжини, рангу та інших характеристик шляхів. Умовно ці методи можна поділити на дві групи – матричні та мережні або індексні.

Матричні алгоритми пошуку найкоротших шляхів

Піднесення матриці L . Матричний метод визначення довжин найкоротших шляхів, що дозволяє знайти довжини найкоротших шляхів між усіма вершинами графа одночасно, базується на застосуванні операцій над матрицями довжин L [19, 20].

Нехай мережа зв'язку задана у виді графа, зображеного на рис. 9.29, цифри на ребрах якого відповідають довжинам гілок.

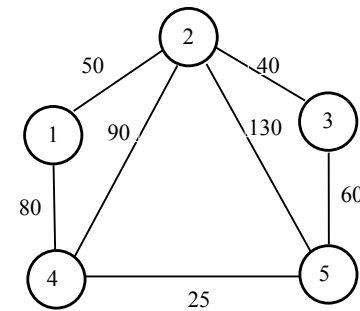


Рисунок 9.29 – Граф мережі (до пояснення степенування матриці L)

Природно, з дерева можна отримати множину шляхів із фіксованої вершини в будь-яку вершину графа послідовним переглядом ярусів дерева. Так,

$$m_{14} = h + \overline{ed} + \overline{abc} + \overline{agd} + \overline{efc} + \overline{abfd} + \overline{agfc};$$

$$m_{15} = e + \overline{ag} + \overline{hd} + \overline{abf} + \overline{hcf} + \overline{abcd} + \overline{hcbg}$$

і т. ін.

9.11 Алгоритми пошуку найкоротших шляхів

Загальні відомості

Розподіл каналів і потоків інформації на мережі зв'язку проводиться з урахуванням вагової характеристики шляху. Для оцінки цієї характеристики використовують різні критерії, наприклад довжину шляху, кількість транзитних ділянок (ранг) шляху, якість тракту передачі, надійність передачі інформації і т. ін.

Найпоширенішими методами розподілу є методи, які використовують "найкоротші" шляхи. Термін "найкоротший шлях" є досить умовним, тому що під найкоротшими розуміють шляхи, найкоротші за довжиною, з мінімальною кількістю транзитних ділянок, з максимальною пропускнуою здатністю, шляхи мінімальної вартості, максимальною надійністю і т. ін.

У графі, який являє собою мережу, виділено певну вершину i , що називається виток. Кожному шляху (i, \dots, j) , який з'єднує вершину-виток i з вершиною-стоком j , ставиться у відповідність величина $\omega(i, \dots, j)$, що називається вагою шляху.

Таким чином, на множині всіх шляхів графа $G = (X, A)$, що з'єднують виток i зі стоком j , визначено вагову функцію $W(i, j)$ вершини j , яка приймає значення $\omega(i, \dots, j)$ на шляху $\mu_{ij} = (i, \dots, j)$.

Шлях, на якому $W(i, j)$ досягає екстремального значення, називається екстремальним (найкоротшим або оптимальним) шляхом із i в j .

Усі методи вибору найкоротших шляхів, що розвинені в теорії потоків, базуються на цілком очевидному твердженні про те, що, якщо найкоротший шлях m_{ij} від довільної вершини i до вершини j проходить через проміжні вершини x_1, \dots, x_k (рис. 9.28), тоді найкоротші шляхи $\mu_{x_1j}, \dots, \mu_{x_kj}$ від вершин x_1, \dots, x_k до вершини j , відповідно, є частинами найкоротшого шляху μ_{ij} від i до j [19].

Якщо довжина шляху $\mu_{x_1, j}$ дорівнює $L_{x_1, j}$, то довжина шляху із i до j $L_{ij} = I_{i, x_1} + L_{x_1, j}$. Тут I_{i, x_1} – довжина гілки сполучення вершин i та x_1

Через те, що шлях μ_{ij} є найкоротшим,

$$L_{ij} = \min_{k=x_1, \dots, x_n} (I_{i, x_k} + I_{x_k, j}),$$

де n – кількість вершин графа.

Запишемо необхідну умову екстремуму: $D^2(g_0, g_1)$:

$$\left. \begin{aligned} \frac{\partial D^2}{\partial g_0} &= -2b_0 + 2r_{01}g_1 + 2r_{00}g_0 = 0, \\ \frac{\partial D^2}{\partial g_1} &= -2b_1 + 2r_{01}g_0 + 2r_{11}g_1 = 0, \end{aligned} \right\}$$

або, що те ж саме,

$$\left. \begin{aligned} r_{00}g_0 + r_{01}g_1 &= b_0, \\ r_{01}g_0 + r_{11}g_1 &= b_1, \end{aligned} \right\} \quad (4.12)$$

Як видно, мінімізація середньоквадратичної похибки (4.11) приводить до системи з двох лінійних рівнянь із двома невідомими. Рішення її має вид:

$$g_0 = \frac{b_0r_{11} - b_1r_{01}}{r_{00}r_{11} - r_{01}^2}, \quad g_1 = \frac{b_1r_{00} - b_0r_{01}}{r_{00}r_{11} - r_{01}^2} \quad (4.13)$$

Чи доставляє отримане рішення мінімум $D^2(g_0, g_1)$?

Знайдемо другі частинні похідні функції $D^2(g_0, g_1)$:

$$\frac{\partial [D^2]^2}{\partial g_0 \partial g_0} = 2r_{00}, \quad \frac{\partial [D^2]^2}{\partial g_1 \partial g_1} = 2r_{11},$$

$$\frac{\partial [D^2]^2}{\partial g_0 \partial g_1} = 2r_{01}, \quad \frac{\partial [D^2]^2}{\partial g_1 \partial g_0} = 2r_{01},$$

і надамо відповідну матрицю Гессе (4.6)

$$\begin{bmatrix} 2r_{00} & 2r_{01} \\ 2r_{01} & 2r_{11} \end{bmatrix}.$$

Перевіримо виконання нерівностей (4.9).

З урахуванням позначення (2.11,б), можемо записати

$$r_{00} = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |\Phi_k(\omega)|^2 d\omega > 0,$$

крім того,

$$\begin{vmatrix} 2r_{00} & 2r_{01} \\ 2r_{01} & 2r_{11} \end{vmatrix} = 4(r_{00}r_{11} - r_{01}^2) > 0 \quad (4.14)$$

і, згідно з позначенням і нерівністю Коші-Буняковського, що, як відомо [5], має вид

$$\left(\int_{-\Omega}^{\Omega} |\Phi_0(\omega)|^2 d\omega \right)^{\frac{1}{2}} \cdot \left(\int_{-\Omega}^{\Omega} |\Phi_1(\omega)|^2 d\omega \right)^{\frac{1}{2}} \geq \left| \int_{-\Omega}^{\Omega} \Phi_0(\omega)\Phi_1(\omega) d\omega \right|, \quad (4.15)$$

для довільних комплексних функцій $\Phi_0(\omega)$ і $\Phi_1(\omega)$, причому знак переходить у знак рівності лише для функцій $\Phi_0(\omega)$ і $\Phi_1(\omega)$, які співпадають з точністю до постійного множника (що не може мати місця при використанні гармонійного коректора).

Таким чином, наша матриця є позитивно визначеною і знайдене рішення насправді реалізує мінімум $D^2(g_0, g_1)$.

Зауваження 4.1. Нерівність (4.14) гарантує існування рішення системи лінійних рівнянь (4.12) при нерівній нулю її правій частині.

4.4 Ітераційний підхід. Унімодальність і опуклість

Для мінімізації функції $D^2(g_0, g_1, \dots, g_n)$, що характеризує точність коректування каналу зв'язку, а також для розв'язування багатьох інших задач оптимізації функцій декількох змінних, на практиці застосовують чисельні методи, що за допомогою ряду ітерацій дозволяють з достатньою точністю наблизитися до мінімуму цієї функції. З подібною ситуацією ми зустрічалися у розділі 3, де ітераційний підхід використовувався стосовно до функцій однієї змінної.

Апаратна реалізація такого підходу базується на добре відомих процедурах оптимізації, при виконанні яких напрямки обертання «ручок» коректора (зміни параметрів g_0, g_1, \dots, g_n) визначаються значеннями поверхні $D^2(g_0, g_1, \dots, g_n)$ і її похідних.

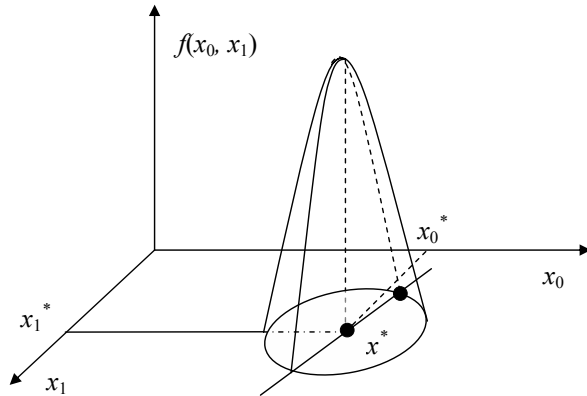


Рисунок 4.3 – Поверхня функції двох змінних x_0 і x_1

Про методи знаходження екстремумів такого типу і піде мова. Важливо відзначити, що процедури послідовних наближень до екстремуму вимагають, щоб досліджувана поверхня мала деякі загальні для всіх точок області визначення (глобальні) властивості.

Визначення 4.5. Функцію $n + 1$ змінних x_0, x_1, \dots, x_n будемо називати унімодальною, якщо уздовж прямої лінії, проведеної з довільної точки в області визначення цієї функції в точку екстремуму $(x_0^*, x_1^*, \dots, x_n^*)$ отримана одномірна функція задовольняє введеної вище умові унімодальності (визначення 3.4).

Сказане пояснює рис. 4.3.

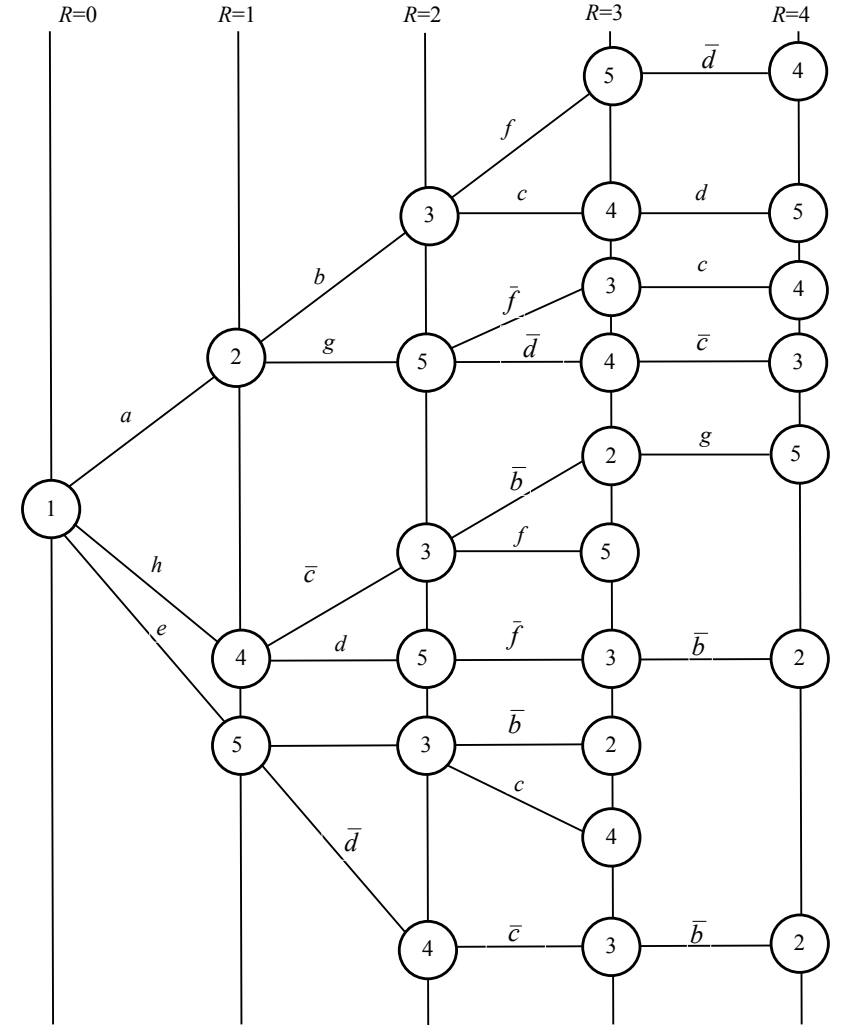


Рис. 9.27 – Дерево шляхів з вершинами витoku 1

Як бачимо, в дереві є три шляхи першого рангу ($R = 1$): a, h, e ; шість шляхів другого рангу ($R = 2$): ab, ag, hc, hd, ef, ed ; на третьому ярусі ($R = 3$) записано шляхи третього рангу: $abf, abc, agf, agd, hc\bar{b}, hc\bar{f}, hdf, ef\bar{b}, ef\bar{c}, ed\bar{c}$. Нарешті, шляхи четвертого рангу ($R = 4$): $abf\bar{d}, abcd, ag\bar{f}c, ag\bar{d}c, hc\bar{b}g, hd\bar{f}b, ed\bar{c}b$.

Оскільки $R_{\max} = 4$ для даного графа мережі, матриця $B^{(4)}$ буде вихідною в будь-якому разі, навіть якщо $B^{(4)} \neq B^{(3)}$.

В отриманій вихідній матриці $B^{(R_{\max})}$, як зазначалось раніше, кожний елемент $B_{ij}^{(R_{\max})} = m_{ij}$, тобто визначає множину всіх шляхів з вершини i у вершину j .

Так,

$$b_{12}^{(R_{\max})} = m_{12} = a + \bar{h}\bar{c}\bar{b} + e\bar{f}\bar{b} + \bar{e}\bar{d}\bar{c}\bar{b} + \bar{h}\bar{d}\bar{f}\bar{b}$$

$$b_{13}^{(R_{\max})} = m_{13} = ab + \bar{h}\bar{c} + e\bar{f} + ag\bar{f} + \bar{h}\bar{d}\bar{f} + \bar{e}\bar{d}\bar{c} + ag\bar{d}\bar{c}$$

$$b_{14}^{(R_{\max})} = m_{14} = h + \bar{e}\bar{d} + abc + e\bar{f}\bar{c} + ag\bar{d} + ag\bar{f}\bar{c} + ab\bar{f}\bar{d}$$

і т. ін.

Мережний алгоритм пошуку множини шляхів.

Мережні методи визначення множини шляхів між заданими вузлами мережі є графічним еквівалентом матричних методів.

Визначення множини шляхів базується на побудові дерева шляхів із фіксованої вершини-витоку (кореня дерева) до решти вершин-стоків графа.

Для побудови дерева шляхів необхідно перш за все відзначити "яруси" дерева. На ярусі $R = 0$ міститься вершина-витік – "корінь" дерева. На ярусі $R = 1$ розміщуються вершини, суміжні до вершини-витоку, тобто вершини, шляхи в які з вершини-витоку мають ранг, що дорівнює одиниці. На ярусі $R = 2$ розміщуються вершини, суміжні до вершин, розміщених на ярусі $R = 1$ і т. ін.

При запису вершини на ярус $R = 2$ і наступні яруси необхідно слідкувати за тим, щоб утворювані шляхи рангів 2, 3 були простими (самонеперерізними), тобто щоб жодна вершина на шляху не повторювалась більше одного разу.

Максимальне значення ярису (рангу шляху) дорівнює $R_{\max} = n - 1$ (n – кількість вершин графа).

Дерево шляхів містить усі шляхи з фіксованої вершини – витоку в усі інші вершини. При цьому на ярусі 1 містяться усі шляхи першого рангу, на ярусі 2 – усі шляхи другого рангу і т. ін.

Таким чином, на деякому k -му ярусі міститься інформація про всі шляхи k -го рангу з фіксованої вершини-витоку графа в решту вершин.

Для графа (рис. 9.26) побудуємо дерево шляхів з вершини 1. Дана вершина є коренем дерева і розміщується на нульовому ярусі. Значення рангу шляху тут $R = 0$. На перший ярус ($R = 1$) розміщуються вершини 2, 4, 5, які мають безпосередній зв'язок з вершиною 1.

Далі, на другий ярус від вершини 2 розміщуються вершини, які пов'язані з вершиною 2, а саме 3 і 5. Вершина 1 виключається з розгляду, оскільки шлях у вершину 2 пройшов через вершину 1. Від вершини 4 на другий ярус записуються вершини 3 та 5, від вершини 5 – вершини 4 і 3. Аналогічно записуються вершини на решту ярусів дерева.

Побудоване дерево для вершини-витоку 1 показано на рис. 9.27.

Часто при дослідженні функцій багатьох змінних спираються на властивість опуклості (ця властивість є більш сильною: з опуклості завжди випливає унімодальність) – вона звичайно гарантує збіжність процесу послідовних наближень до екстремуму.

Визначення 4.6

Поверхня $f(x_0, x_1, \dots, x_n)$ називається опуклою, якщо відрізок прямої лінії, що з'єднує будь-які дві точки, розташовані на поверхні, що лежить цілком, не вважаючи кінців, нижче цієї поверхні (рис. 4.4,а). Якщо відрізок, що з'єднує згадані точки, лежить цілком вище поверхні, то багатомірну функцію називають опуклою вниз (рис. 4.4,б).

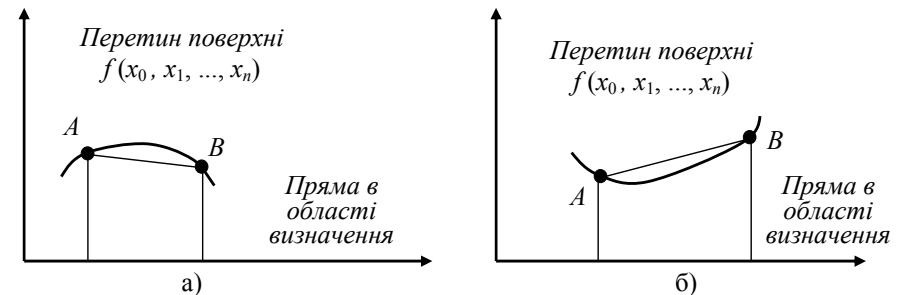


Рисунок 4.4 – Визначення типу опуклості багатомірної функції

Доведено, що середньоквадратична похибка коректування D^2 є опуклою униз функцією.

За своєю суттю опуклість вимагає того ж, що й унімодальність, але для будь-якого «вертикального» перетину поверхні, і не тільки для перетину, що проходить через точку екстремуму. Приведемо вид ліній рівня, що одержані у результаті горизонтальних перетинів опуклої (а) і унімодальної (б) поверхонь (рис. 4.5), (порівняємо перетини А, В, С, D, Е на обох рисунках: у перетині Е рис. 4.5, б умова опуклості порушується). На рис. 4.5 $c = 1, 2, 3, 4, 5$.

Зауваження. Поверхнею рівня функції $f(x_0, x_1, \dots, x_n)$ називають множину точок області x з координатами x_0, x_1, \dots, x_n , що задовольняють рівнянню

$$f(x_0, x_1, \dots, x_n) = c, \quad c = \text{const.}$$

У випадку двовимірних поверхонь говорять про лінії рівня

$$f(x_0, x_1) = c, \quad c = \text{const.}$$

Лінії рівня дозволяють наочно передавати характер зміни поверхні, не користуючись аксонометрією, і застосовуються в різних прикладних питаннях, наприклад, у топографії.

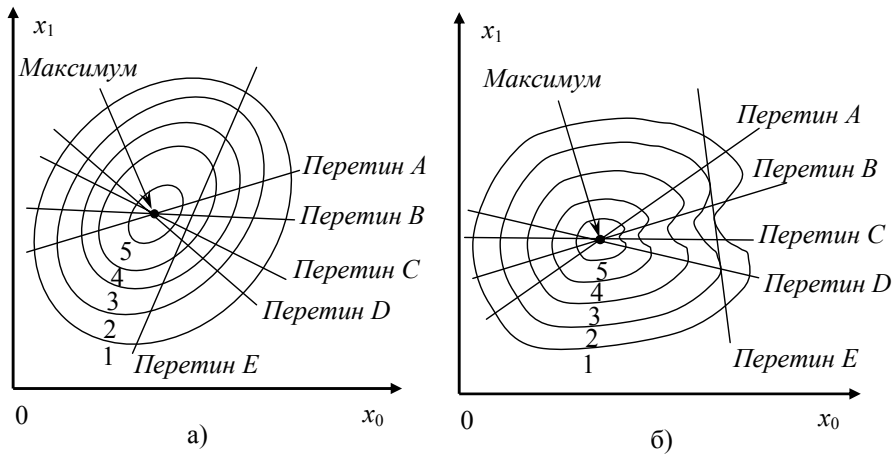


Рисунок 4.5 – Визначення рівнів опуклості багатомірної функції

4.5 Метод перетинів

Очевидно, найбільш простим з відомих методів багатомірної оптимізації, тобто відшукування екстремумів цільових функцій декількох змінних $f(x_0, x_1, \dots, x_n)$ є метод перетинів або метод почергового пошуку по кожній змінній x_0, x_1, \dots, x_n . Сутність його при відшуванні, наприклад, мінімуму полягає в тому, що кожну з незалежних змінних змінюють доти, поки досліджувана, функція – поверхня $f(x_0, x_1, \dots, x_n)$ не перестане убувати, потім переходять до іншої змінної, поки не будуть перебрані всі змінні, після чого повертаються до тієї змінної, з якої починався пошук, відкриваючи тим самим другу ітерацію процесу, і т. і.

Рис. 4.6 показує, що успіх застосування даної процедури оптимізації істотно залежить від виду поверхні $f(x_0, x_1, \dots, x_n)$ і навіть в опуклих уніз функціях метод перетинів не завжди дозволяє знайти точку мінімуму. На рис. 4.6 поверхня представлена своїми лініями рівня (у випадках а) і б) – еліптичними).

Ясно, що якість методу перетинів істотно залежить від виду ліній рівня. Оскільки кожна ланка траєкторії пошуку паралельна одній з координатних осей і загальна траєкторія руху в області визначення утворює «сходи», метод перетинів є найбільш ефективний у тому випадку, коли лінії рівня уявляють собою кола або еліпси, головні осі яких паралельні координатним осям. (рис. 4.6,а). Якщо ж головні осі еліпсів повернені відносно координатних осей, то приходить багато разів змінювати напрямок пошуку, перш ніж мінімум буде досягнуто (рис. 4.6,б). Цей метод зовсім не придатний при оптимізації поверхні зі зломом (або, як часто говорять, з вузьким яром), тому що не можна просунути вниз по зламу з точки А (рис. 4.6,в) не маючи можливості

Якщо при піднесенні до деякого степеня q виявиться, що $B^{(q)} = B^{(q-1)}$, обчислення слід припинити. При цьому $B^{(q)} = B^{(q-1)} = B^{(R_{\max})}$.

Покажемо процес отримання матриці $B^{(2)}$ для графа (див. рис. 9.26) матриці B (9.113).

Для отримання елемента $b_{12}^{(2)}$ слід використати елементи першого рядка і другого стовпця матриці $B^{(1)}$:

$$b_{12}^{(2)} = \frac{1 \quad a \quad 0 \quad h \quad e}{a + a + 0 + 0 + 0 = a} \begin{matrix} \text{– I рядок } B^{(1)} \\ \text{– II стовпець } B^{(1)} \end{matrix}$$

Аналогічно отримуємо решту елементів матриці $B^{(2)}$:

$$b_{13}^{(2)} = \frac{1 \quad a \quad 0 \quad h \quad e}{ab + h\bar{c} + e\bar{f}} \quad b_{14}^{(2)} = \frac{1 \quad a \quad 0 \quad h \quad e}{h + e\bar{d}}$$

$$b_{15}^{(2)} = \frac{1 \quad a \quad 0 \quad h \quad e}{e + ag + hd} \quad b_{21}^{(2)} = \frac{1 \quad \bar{a} \quad 0 \quad 0 \quad \bar{e}}{\bar{a} + g\bar{e}}$$

$$b_{23}^{(2)} = \frac{0 \quad b \quad 1 \quad \bar{c} \quad \bar{f}}{b + g\bar{f}} \quad b_{24}^{(2)} = \frac{h \quad 0 \quad c \quad 1 \quad \bar{d}}{\bar{a}h + bc + g\bar{d}}$$

і т. ін.

Отримані результати зведемо в матрицю

$$B^2 = \begin{bmatrix} 1 & a & ab + h\bar{c} + e\bar{f} & h + e\bar{d} & e + ag + hd \\ \bar{a} + g\bar{e} & 1 & b + g\bar{f} & \bar{a}h + bc + g\bar{d} & g + \bar{a}e + b\bar{f} \\ \bar{b}\bar{a} + \bar{f}\bar{e} & \bar{b} & 1 & c + f\bar{d} & f + \bar{b}g + c\bar{d} \\ d\bar{e} & \bar{c}\bar{b} & \bar{c} + d\bar{f} & 1 & d + \bar{c}f \\ \bar{e} & \bar{e}a + \bar{f}\bar{b} & \bar{f} + \bar{d}\bar{c} & \bar{d} + \bar{e}h + \bar{f}c & 1 \end{bmatrix} \quad (9.116)$$

З порівняння $B^{(2)}$ (9.116) і $B^{(1)}$ (9.113) випливає, що отримана матриця $B^{(2)}$ не є вихідною, тому що $B^{(2)} \neq B^{(1)}$, і процес обчислень слід продовжити, а саме:

- обчислити $B^{(3)} = B^{(2)} \cdot B^{(1)}$;
- якщо $B^{(3)} = B^{(2)}$, $\hat{0}$ $B^{(3)} = B^{(R_{\max})}$, результат отримано.

Якщо $B^{(3)} \neq B^{(2)}$, продовжити розрахунок, тобто знайти $B^{(4)}$.

$$\begin{aligned}
 m_{14} &= a \begin{vmatrix} b & 0 & g \\ 1 & c & f \\ f & d & 1 \end{vmatrix} + h \begin{vmatrix} 1 & b & g \\ 0 & f & 1 \end{vmatrix} + e \begin{vmatrix} 1 & b & 0 \\ 0 & f & d \end{vmatrix} = \\
 &= a(bc + g\bar{d} + \bar{f}cd + \bar{d}bf) + h + e(\bar{d} + \bar{f}c) = \\
 &= h + e\bar{d} + abc + e\bar{f}c + ag\bar{d} + ag\bar{f}c + abf\bar{d}.
 \end{aligned}$$

Піднесення структурної матриці B . Вихідна структурна матриця B є матрицею прямих зв'язків, тобто в ній записані всі шляхи першого рангу, які є в графі. При піднесенні матриці до квадрата кожен її елемент $b_{ij}^{(2)}$ буде являти собою множину шляхів до другого рангу включно, що існують між вершинами i та j . Піднесення матриці B до степеня $R_{\max} = n - 1$ призводить до отримання усіх шляхів, які існують між кожною парою (i, j) вершин графа, тобто $m_{ij} = b_{ij}^{(n-1)}$. Тут n – кількість вершин графа; R_{\max} – максимально можливий ранг шляхів у графі; $R_{\max} = n - 1$ у випадку, коли який-небудь шлях проходить через усі вершини графа.

При піднесенні матриці користуються відомими правилами лінійної алгебри

$$\begin{aligned}
 B^{(2)} &= B^{(1)} \cdot B^{(1)}, \\
 B^{(3)} &= B^{(2)} \cdot B^{(1)} \text{ тощо.}
 \end{aligned}$$

Елементи матриці-добутку $C = A \cdot D$ у лінійній алгебрі отримують відповідно до виразу

$$c_{ij} = \sum_{k=1}^n a_{ik} d_{kj}, \quad (9.115)$$

де n – загальний розмір матриць A і D (кількість стовпців матриці A та рядків матриці D ; у разі ступенювання квадратної матриці B n – кількість вершин).

Оскільки матриця B – бульова, ступенювання її здійснюється з використанням законів і тотожностей бульової алгебри. Крім того, у виразі (9.115) операції алгебричного множення та додавання замінюються операціями логічного множення та додавання, відповідно.

При піднесенні матриці B до степеня q кожен елемент $b_{ij}^{(q)}$ матриці $B^{(q)}$ обчислюється за такими правилами:

– елементи i -го рядка матриці $B^{(q-1)}$ логічно помножуються на відповідні елементи j -го стовпця матриці $B^{(1)}$;

– отримані логічні добутки логічно складаються, утворюючи шукане значення елемента $b_{ij}^{(q)}$.

переміщуватися в досить малому секторі напрямків α (MN – напрямок яру); рух уздовж координатних осей тут завжди приводить до зростання поверхні.

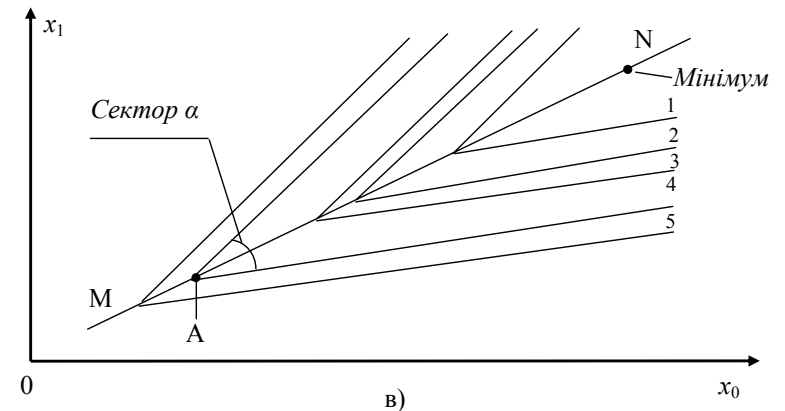
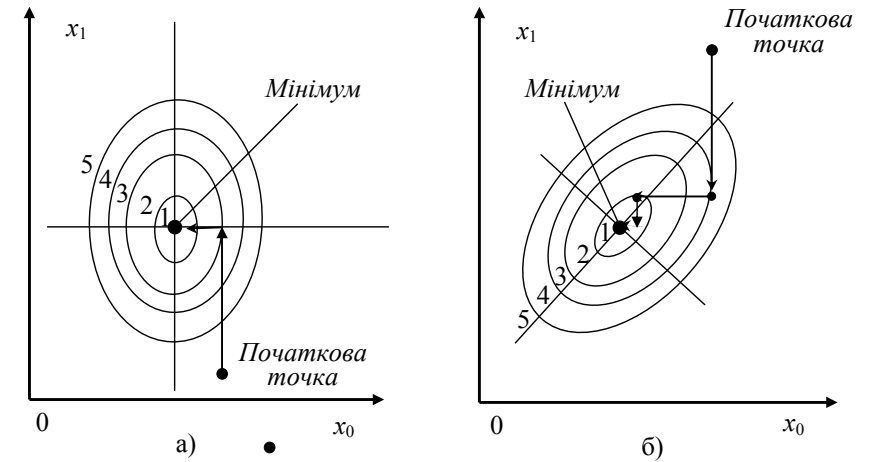


Рисунок 4.6 – Траєкторія пошуку: а) осі еліпса паралельні координатним осям; б) осі еліпса не паралельні координатним осям; в) поверхня зі зломом

Відзначимо, що в багатьох опуклих (опуклих униз) поверхнях, що зустрічаються на практиці, завжди можна зробити крок, нехай невеликий, уздовж однієї з координатних осей так, щоб поверхня зменшилася. Однак кінцева точність обчислень (або вимірів – у випадку апаратної оптимізації), що виключає кроки занадто малої довжини, може зупинити просування вниз по яру. У зв'язку з цим були запропоновані методи оптимізації, засновані на русі за напрямками, не паралельними координатним осям.

Зауваження 4.2. Уважний розгляд рис. 4.6 показує, що краї «сходинок» траєкторії пошуку знаходяться в точках торкання прямих ліній одномірного

пошуку і ліній рівня поверхні, що досліджується на екстремум. Пояснення цього явища таке: якби в точці торкання не відбулася зупинка і рух продовжувався далі уздовж тієї ж прямої у напрямку стрілки, то відбувалися б «перескоки» на лінії усе більш високого рівня, що суперечить описаній вище процедурі мінімізації.

Зауваження 4.3. Оскільки процедура перетинів базується на одномірній оптимізації, то природно виникає питання: яким з методів, описаних у розд. 2, варто користуватися? Відповідь така: тим, що адекватний можливостям дослідника (методом «золотого перетину», якщо недоступні відомості про похідні цільової функції; методом січних, якщо дані про похідні ϵ , і т. ін.). Це зауваження зберігає свою силу по відношенню і до інших процедур багатомірної оптимізації, наприклад, досліджуваному в наступному підрозділі методу найшвидшого спуска (підйому).

Зауваження 4.4. Ітерації в методах багатомірної оптимізації (зокрема, у методі перетинів) звичайно продовжують доти, поки збільшення функції при переході від однієї ітерації до наступної не стане менше якогось наперед заданого числа.

Розглянемо умови застосовності методів перетинів до задачі корекції лінійних спотворень за допомогою коректора, чий базисні функції $\Psi_k(\omega)$ задовольняють умові ортонормованості

$$\int_{-\Omega}^{\Omega} \Psi_k(\omega) \overline{\Psi_l(\omega)} d\omega = \begin{cases} 1, & k=l, \\ 0, & k \neq l. \end{cases} \quad (4.16)$$

Нагадаємо, що середньоквадратична похибка коректування у випадку коректора з двома регуляторами

$$\begin{aligned} D^2(g_0, g_1) &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \left| F(\omega) - \sum_{k=0}^n g_k \Psi_k(\omega) \right|^2 d\omega = \\ &= d - 2b_0g_0 - 2b_1g_1 + 2r_{01}g_0g_1 + r_{00}g_0^2 + r_{11}g_1^2, \end{aligned} \quad (4.17)$$

де

$$\begin{aligned} r_{kl} &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Phi_k(\omega) \Phi_l(\omega) d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Psi_k(\omega) H(\omega) \overline{\Psi_l(\omega)} \overline{H(\omega)} d\omega = \\ &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Psi_k(\omega) \overline{\Psi_l(\omega)} |H(\omega)|^2 d\omega; \quad l, k=0, 1, \dots, n, \\ b_k &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) \overline{\Psi_k(\omega)} \overline{H(\omega)} d\omega, \quad k=0, 1, \dots, n. \end{aligned}$$

Припустимо, що канал зв'язку має спотворену ФЧХ, але не вносить амплітудно-частотних спотворень

$$|H(\omega)| = 1, \quad -\Omega < \omega < \Omega \quad (4.18)$$

алгебричного множення та додавання замінюються на операції логічного множення та додавання, відповідно.

При перетворенні матриці та визначників слід використовувати основні правила й закони бульової алгебри:

- 1) $1 + a = 1; 1 \cdot a = a;$
- 2) $0 + a = a; 0 \cdot a = 0;$
- 3) $a + \bar{a} = 1; a \cdot \bar{a} = 0;$
- 4) $a + a = a; a \cdot a = a$ – закон повторення;
- 5) $a + ab = a$ – закон поглинання;
- 6) $(a + x) \cdot (a + y) = a + xy$ – розподільний закон.

При розкладанні бульових визначників необхідно також користуватися такими правилами:

- визначник, що має одиницю в кожному рядку та стовпці, тотожно дорівнює одиниці;
- якщо який-небудь рядок чи стовпець складається з нулів, то визначник тотожно дорівнює нулю;
- якщо перед визначником записано множник "a", тоді усі елементи "a" у визначнику можна замінити на "1", а усі елементи "a" – на "0";
- якщо у визначнику поміняти місцями два рядки або два стовпці, або провести його транспонування, то значення не зміниться.

Розглянемо отримання множини шляхів m_{14} з матриці B (9.113):

$$m_{14} = \det B_{14} = \begin{vmatrix} a & 0 & h & e \\ 1 & b & 0 & g \\ \bar{b} & 1 & c & f \\ 0 & \bar{f} & \bar{d} & 1 \end{vmatrix}.$$

Розкладання визначника проведемо за стовпцем з номером 4:

$$\begin{aligned} m_{14} &= h \begin{vmatrix} 1 & b & y \\ 0 & \bar{f} & 1 \end{vmatrix} + c \begin{vmatrix} a & 0 & e \\ 1 & b & g \\ \bar{b} & 1 & f \end{vmatrix} + \bar{d} \begin{vmatrix} a & 0 & e \\ 1 & b & g \\ 0 & \bar{f} & 1 \end{vmatrix} = \\ &= h + c(ab + ag\bar{f}) + \bar{d}(abf + e + ag) = \\ &= h + cab + ca\bar{f} + cag\bar{f} + \bar{d}abf + \bar{d}e + \bar{d}ag. \end{aligned}$$

Упорядкуємо отримані шляхи за рангом, врахуємо послідовність ребер у шляхах. Отримаємо:

$$m_{14} = h + e\bar{d} + abc + e\bar{f}c + ag\bar{d} + ag\bar{f}c + abf\bar{d}.$$

Такий самий результат отримаємо, розклавши визначник, наприклад, за першим рядком:

може розглядатися як бульова матриця, оскільки її елементами є бульові змінні. Дійсно, будь-який елемент матриці B – це змінна, її інверсія, константи 0 або 1.

Таким чином, шлях як послідовність ребер, може бути поданий кон'юнкцією (логічним множенням) ребер, утворюючих його. Розглянуті шляхи μ_{14} запишемо у формі

$$m_{14} = agfc; m_{14} = h; m_{14} = abc.$$

Множину шляхів можна записати в диз'юнктивній нормальній формі, застосувавши операцію диз'юнкції (логічного додавання) до шляхів, що складають множину. Так у розглянутому прикладі $m_{14} = agfc + h + abc$.

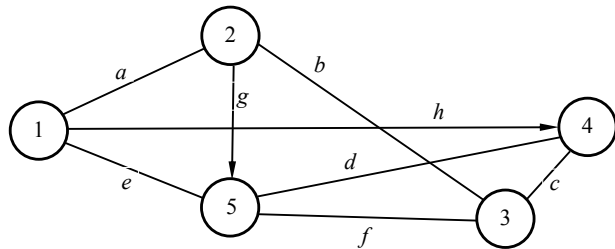


Рисунок 9.26 – Граф мережі

Тут для позначення логічного множення та додавання застосовуються символи алгебричного множення та додавання. Крім наведених символів використовуються також символи:

$\&$, \wedge – для позначення кон'юнкції;

\vee – для позначення диз'юнкції.

Множину шляхів m_{ij} можна одержати розкладанням визначника матриці B , з якій попередньо викреслюється стовпець з номером i та рядок з номером j , тобто:

$$m_{ij} = \det B_{ij} = b_{k1}|B_{k-1}| + b_{k2}|B_{k2}| + \dots + b_{k1}|B_{k1}| + \dots + b_{kn}|B_{kn}|. \quad (9.114)$$

Розкладання визначника матриці B_{ij} проведено за рядком з номером k .

Розкладання (9.114) можна проводити за будь-яким рядком (стовпцем). Тут B_{ij} – матриця, отримана з вихідної структурної матриці після викреслення i -го стовпця та j -го рядка.

У стовпцях кожної вершини графа записані ребра, що входять до цієї вершини; в рядках кожної вершини графа записані ребра, що виходять із вершини. Вилучивши стовпець i та рядок j , тим самим вилучимо ті ребра, котрі не можуть утворювати шляхів множини $m_{ij} \cdot |B_{kl}|$ – визначник, отриманий з $\det B_{ji}$ після вилучення k -того рядка i -го стовпця. Розкладання визначника (9.114) слід проводити за відомими правилами лінійної алгебри, однак операції

Підставивши (4.18) у вираз для коефіцієнта r_{kl} і з огляду на умову ортонормованості (4.16), виявляємо, що усі r_{kl} при $k \neq l$ дорівнюють нулю, а при $k = l - 1/(2\pi)$. Отже, середньоквадратична похибка перевтілюється до виду

$$D^2(g_0, g_1) = d - 2b_0g_0 - 2b_1g_1 + g_0^2 + g_1^2 \quad (4.19)$$

Неважко перевірити, що рівняння лінії рівня $D^2(g_0, g_1) = c$, у цьому випадку є рівняння кола [6] з центром у точці, координати якої рівні

$$g_0^* = b_0, \quad g_1^* = b_1 \quad (4.20)$$

Числа (4.20), як виходить з формули (4.13), є координатами мінімуму функції (4.19). Отже, лінії рівня поверхні (4.19) можна описати графіком на рис. 4.7. Тоді дійдемо до висновку, що канал зв'язку з великими фазочастотними спотвореннями і постійною АЧХ може бути відкоректований за допомогою, наприклад, гармонійного коректора, що допускає налаштування лише однією ітерацією методу перетинів (див. рис. 4.6,а).

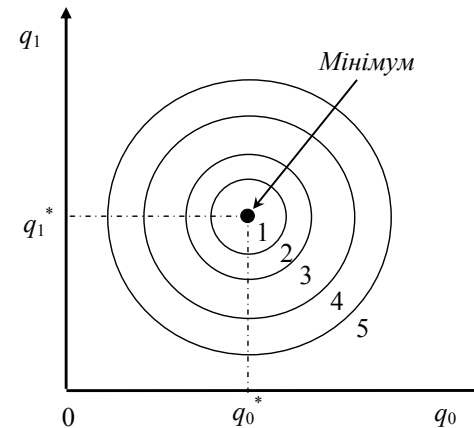


Рисунок 4.7 – Лінії рівня поверхні (4.17)

Проведені розрахунки й аналіз середньоквадратичної похибки коректування показали, що при наявності спотворень АЧХ лінії рівня (див. рис. 4.6,б) трансформуються в еліпси, чії осі не паралельні осям координат, причому, трохи спрощуючи опис ситуації, можна сформулювати наступне твердження: чим більша нерівномірність АЧХ, тим більш витягнутими є еліпси і тим менш ефективно буде застосування методу перетинів для налаштування ортогонального коректора.

Саме такі положення враховують при корекції гармонійним коректором телефонних каналів, нерівномірність залишкового загасання яких досягає 6 – 8 дБ.

Підводячи підсумки, можна сказати, що головним достоїнством методу перетинів є його простота, у зв'язку з чим, цей метод звичайно застосовується тоді, коли процес налаштування коректора автоматизувати неможливо і відшукування мінімуму похибки коректування випадає робити «ручним» способом.

Контрольні питання та задачі до розділу 4

- 4.1. У чому суть ітераційних методів відшукування екстремуму?
 4.2. Яку поверхню називають опуклою (опуклою вниз)?
 4.3. У чому суть методу перетинів?
 4.4. Який вид мають поверхні (лінії) рівня середньоквадратичної похибки коректування?
 4.5. Поясніть, чому при проведенні ітерацій значення середньоквадратичної похибки монотонно убуває.
 4.6. Який вид мають поверхні (лінії) рівня середньоквадратичної похибки коректування гармонійним коректором при відсутності амплітудно-частотних спотворень (порівняйте з випадком нерівномірної АЧХ каналу)?
 4.7. У чому полягає схожість визначення необхідних та достатніх умов екстремуму функцій однієї та багатьох змінних?
 4.8. Якщо не виконуються умови позитивної чи від'ємної знаковизначенності матриці за критерієм Сильвестра, то що можна сказати про таку стаціонарну точку?
 4.9. Покажіть, що функція

$$f(x) = 3x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 + 2x_2x_3 - 6x_1 - 4x_2 - 2x_3$$
 є опуклою.
 4.10. Знайдіть і класифікуйте стаціонарні точки функції

$$f(x) = 2x_1^3 + 4x_1x_2^2 - 10x_1x_2 + x_2^2$$

 4.11. У результаті пошуку мінімуму функції

$$f(x) = [x_1^2 + (x_2 + 1)^2] \cdot [x_1^2 - (x_2 - 1)^2]$$
 знайдено наступні точки:
 а) $x^{(1)} = [0, 0]^T$;
 б) $x^{(2)} = [0, 1]^T$;
 в) $x^{(3)} = [0, -1]^T$;
 г) $x^{(4)} = [1, 1]^T$.
 Класифікуйте отримані точки.
 4.12. Визначте і класифікуйте стаціонарні точки функції

$$f(x) = x_1^3 - x_1x_2 + x_2^2 - 2x_1 + 3x_2 - 4$$

 4.13. Нехай потрібно переправити 500 дм³ поштового вантажу в інше місто. Для перевезення вантажу необхідно побудувати контейнер. Відомі наступні дані: вартість кожного рейсу в інше місто і назад 500 грн.; вартість матеріалів для виготовлення дна контейнера складає 5,00 грн./дм²; бічних стінок контейнера – 1,00 грн./дм², кришки контейнера – 1,00 грн./дм².

9.9 Методи пошуку шляхів у мережах

Основне призначення мережі зв'язку полягає у тому, щоб давати абонентам з'єднувальні шляхи для передачі повідомлень, відповідно до адрес, та необхідних якісних показників (якості обслуговування, швидкості передачі, надійності і т. ін.); при цьому вибір з'єднувальних шляхів має здійснюватися таким чином, щоб забезпечити найбільш ефективне використання обладнання мережі, забезпечити мінімально можливі довжину шляхів і число транзитних ділянок у шляхах, забезпечити необхідну кількість каналів у шляхах і т. ін. [19].

При рішенні різних задач аналізу та синтезу мереж зв'язку виникає необхідність у пошуку множини шляхів, які існують між заданою парою вузлів зв'язку (вершин графа). Усі методи пошуку множини шляхів в мережі можна поділити на два класи: матричні та мережні. *Матричні методи* ґрунтуються на перетворенні різних матриць – топологічних або матриць характеристик ребер графів, мережні – на присвоєнні вершинам графа деяких позначень, що називаються позначками (або індексами).

9.10 Алгоритми пошуку множини шляхів

Матричні алгоритми пошуку множини шляхів

Розкладання бульового визначника структурної матриці В. Якщо в мережі є можливість передачі інформації з вузла i у вузол j (по прямих каналах або через проміжні вузли), тоді говорять, що існує зв'язок від i до j . Для здійснення зв'язку має існувати відповідний шлях (або шляхи).

Введемо такі позначення:

μ_{ij} – шлях від вузла i до вузла j (з вершини i у вершину j графа);

m_{ij} – множина шляхів з i в j . Шлях μ_{ij} – це упорядкований набір ребер, утворюючих його.

Довжина шляху $\mu_{ij} - L_{ij}$ – це сума довжин I_{xy} ребер, утворюючих його.

$$L_{ij} = \sum_{x,y \in \mu_{ij}} I_{xy}$$

Ранг шляху R_{ij} – кількість ребер, що утворюють шлях.

Для графа, показаного на рис. 9.26, для шляху $\mu_{14} = a_{12}, a_{25}, a_{53}, a_{34}$ ранг $R_{14} = 4$, для шляху $\mu_{14} = a_{14}$ ранг $R_{14} = 1$, для шляху $\mu_{14} = a_{12}, a_{23}, a_{34}$ ранг $R_{14} = 3$.

Структурна матриця графа, поданого на рис. 9.26,

$$B = \begin{bmatrix} 1 & a & 0 & h & e \\ \bar{a} & 1 & b & 0 & g \\ 0 & \bar{b} & 1 & c & f \\ 0 & 0 & \bar{c} & 1 & d \\ \bar{e} & 0 & \bar{f} & \bar{d} & 1 \end{bmatrix} \quad (9.113)$$

$$2\Omega = \frac{2\pi}{\Delta t} \quad (9.112)$$

Таким чином, для створення коефіцієнта передачі $\Phi(\omega)$, що дорівнює нулю за частотою зрізу Ω , необхідне каскадне з'єднання гармонійного синтезатора і ідеального фільтра з частотою зрізу Ω (рис. 9.24).

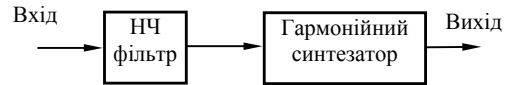


Рисунок 9.24 – Система НЧ фільтр-гармонійний синтезатор

У системі НЧ фільтр-гармонійний синтезатор рекомендується використати високоякісний НЧ фільтр, амплітудно-частотна характеристика якого показана на рис. 9.25.

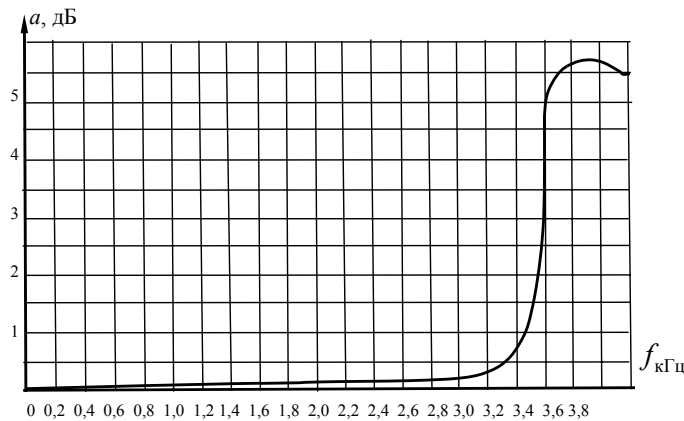


Рисунок 9.25 – Амплітудно-частотна характеристика НЧ фільтра, що включається послідовно з гармонійним синтезатором

Методика створення різних коефіцієнтів передачі за допомогою гармонійного синтезатора описана в [17].

Сконструйте контейнер так, щоб мінімізувати загальні витрати на перевезення вантажу.

Література до розділу 4

- 1 Захарченко Н.В. Оптимизация и моделирование систем связи: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, 1988. – Ч. 1.– 86 с.
- 2 Шуп Т. Решение инженерных задач на ЭВМ / Шуп Т. – М.: Мир, 1982.
- 3 Моисеев Н.Н. Методы оптимизации / Н.Н. Моисеев, Н.Н. Иванилов, Е.М. Столярова – М.: Наука, 1978.
- 4 Латхи Б.П. Системы передачи информации / Б.П. Латхи, пер. с англ., под общ. ред. Б. И. Кувшинова. – М.: Связь, 1971.
- 5 Обработка сигналов в линейных аналоговых и цифровых цепях: метод. руководство для научно-исследовательских работ студентов / Сост. П.Я. Нудельман. – Одесса: Одесск. электротехн. ин-т связи им. А.С. Попова. – 1987.– Ч. 1.
- 6 Привалов И.И. Аналитическая геометрия: учебник для ВТУЗов / И.И. Привалов [Изд. 30-е, стереотипн.] – М.: Наука, 1966.

5 ГРАДІЄНТНІ МЕТОДИ ПОШУКУ ЕКСТРЕМУМУ

5.1 Вектор-градієнт і його властивості

Координати вектора-градієнта збігаються з першими похідними функції за усіма незалежним змінним. Розглянемо більш докладно цей вектор.

Покажемо функцію $f(x_0, x_1)$ на рис. 5.1.

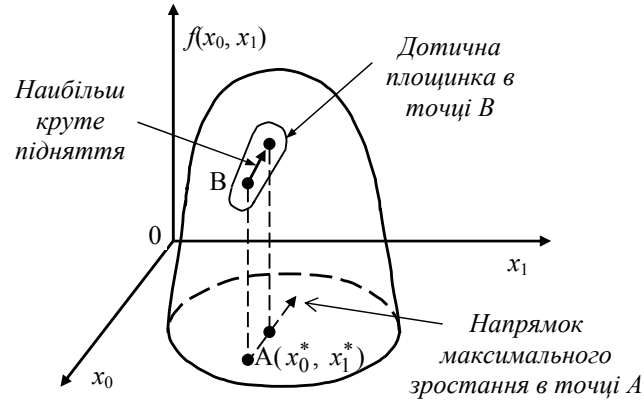


Рисунок 5.1 – Напрямок вектора максимального зростання на поверхні $f(x_0, x_1)$

Виберемо довільно на площині $x_0 \times x_1$ (області визначення) точку А з координатами (x_0^*, x_1^*) і знайдемо напрямок максимального зростання поверхні $f(x_0, x_1)$ у точці В, що лежить на цій поверхні над точкою А, тобто в точці перетинання перпендикуляра до площини $x_0 \times x_1$, відновленого у точці А, і поверхні $f(x_0, x_1)$. Це напрямок у площині $x_0 \times x_1$, рух уздовж якого з точки А приводить до найбільш крутого підйому поверхні $f(x_0, x_1)$, при відповідному переміщенні по цій поверхні з точки В, спробуємо охарактеризувати деяким вектором (див. рис. 5.1). У малому околі точки А довільно гладка поверхня може бути добре наближена дотичною площинкою, рівняння якої подамо за формулою [1]

$$y = m_0 x_0 + m_1 x_1 + c, \quad (5.1)$$

де

$$m_0 = \left. \frac{\partial f(x_0, x_1)}{\partial x_0} \right|_{\substack{x_0 = x_0^* \\ x_1 = x_1^*}}, \quad m_1 = \left. \frac{\partial f(x_0, x_1)}{\partial x_1} \right|_{\substack{x_0 = x_0^* \\ x_1 = x_1^*}}, \quad (5.2)$$

екстремум якого реалізується на рішеннях рівняння

$$g(t) = \mu \int_0^{\infty} \frac{1}{\pi} \frac{\sin \Omega(t - \tau)}{t - \tau} g(\tau) d\tau \quad (9.110)$$

Ядро $\frac{\sin \Omega(t - \tau)}{t - \tau}$ детально досліджене для кінцевого інтервалу (розділ 6) і для нескінченного інтервалу (рівняння Бейтмена).

Єдиним відомим нам рішенням рівняння (9.110) є бesselева функція $J_0\left(\frac{t}{\Omega}\right)$ при значенні $\mu = 2$.

Про одну можливість генерування сигналів із спектрами, близькими до оптимального $\bar{\Phi}(\omega)$. Генерування сигналів із спектрами, близькими до оптимального спектра $\bar{\Phi}(\omega)$ було здійснено нами за допомогою гармонійного синтезатора частотних характеристик.

Гармонійний синтезатор частотних характеристик є лінією затримки з відводами і суматором (рис.9.23).

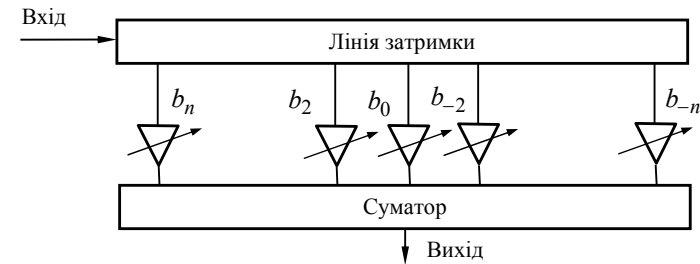


Рисунок 9.23 – Гармонійний синтезатор частотних характеристик

Принцип дії його аналогічний принципу дії синтезатора вокодера. Точніше кажучи, синтезатором гармонійного вокодера [25] є гармонійний синтезатор частотних характеристик з електрично керованими коефіцієнтами передачі по кожному відводу.

Гармонійний синтезатор дозволяє апроксимувати задану частотну характеристику за допомогою суми

$$\sum_{k=-N}^N b_k e^{j\omega k \Delta t}, \quad (9.111)$$

де Δt – часовий інтервал між сусідніми відводами ліній затримки.

Подібний пристрій реалізує, як це виходить з формули (9.111), частотну характеристику, що періодично повторюється по всій осі частот ω . Періодом розкладання (9.111) є частотний інтервал

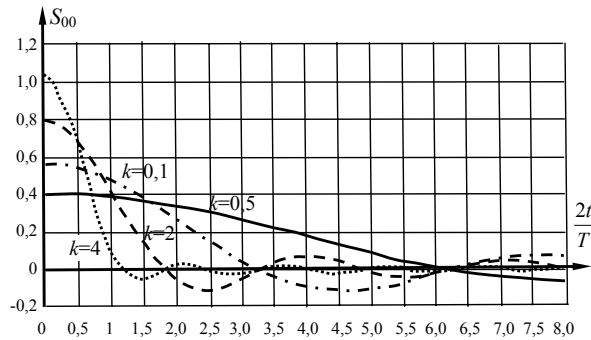


Рисунок 9.22 – Графіки рішення $S(t)$ (рівняння (9.101)) для значень t

Про одне екстремальне завдання для фізично реалізуемого фільтра. Нехай маємо імпульсну реакцію $g(t)$ деякої лінійної системи, про яку відомо, що

$$g(t) = 0 \quad \text{при} \quad t < 0 \quad (9.105)$$

та

$$\int_0^{\infty} g^2(t) dt = A^2. \quad (9.106)$$

Коефіцієнт передачі лінійного фільтра із імпульсною реакцією $g(t)$

$$G(\omega) = \int_0^{\infty} g(t) e^{-j\omega t} dt. \quad (9.107)$$

Будемо називати смугою прозорості фільтра $G(\omega)$ частотний інтервал $(0, \Omega)$, а смугою затримання (Ω, ∞) .

Очевидно, що лінійна система з коефіцієнтом передачі (9.107) фільтрує тим краще, чим менша різниця

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega - \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |G(\omega)|^2 d\omega. \quad (9.108)$$

“Ідеальні” системи з нескінченним загасанням в смузі затримання обертають різницю за формулою (9.108) в нуль, проте імпульсні реакції таких систем не задовольняють умові (9.105) (вони фізично не реалізуються).

Проведемо пошук лінійної системи, що якнайкраще фільтрує (у розумінні мінімуму (9.108)) та яка задовольняє умовам (9.105) і (9.106).

Підставимо (9.107) в (9.108) і з урахуванням (9.106) одержимо функціонал

$$\int_0^{\infty} g^2(t) dt - \mu \int_0^{\infty} \int_0^{\infty} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \varphi(t) \varphi(\tau) dt d\tau, \quad (9.109)$$

а константа c може бути визначена з рівняння

$$f(x_0^*, x_1^*) = m_0 x_0^* + m_1 x_1^* + c,$$

яке отримано прирівнюванням значень поверхні $f(x_0, x_1)$ і дотичної площинки (5.1) у розглянутій точці $A(x_0^*, x_1^*)$.

Геометричний зміст рівностей (5.2) полягає в тому, що похідні поверхні і дотичної площинки до цієї поверхні (похідні обчислюються в точці торкання) збігаються.

Побудуємо тепер коло малого радіуса r з центром у точці $A(x_0^*, x_1^*)$, як це показано на рис. 5.1, де зображені лінії рівня поверхні $f(x_0, x_1)$. Довільна точка на цьому колі може бути задана (при відомому значенні r) за допомогою кута повороту φ її радіуса-вектора щодо деякого фіксованого напрямку, наприклад, осі x_0 . Порушимо питання про відшукування максимального значення функції (5.1) на цьому колі. Точка, у якій буде досягнутий максимум, і покаже напрямок максимального зростання дотичної площинки (рис. 5.2), а разом з нею і функції $f(x_0, x_1)$ у малому околі точки A .

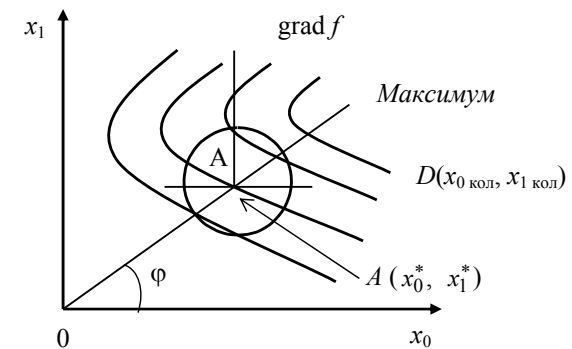


Рисунок 5.2 – Побудова кола малого радіуса у окіллі точки A

Для координат довільно обраної точки та кола $D(x_{0 \text{ кол}}, x_{1 \text{ кол}})$ справедливо співвідношення

$$x_{0 \text{ кол}} = x_0^* + r \cos \varphi, \quad x_{1 \text{ кол}} = x_1^* + r \sin \varphi,$$

$$x_{0 \text{ кол}} = x_0^* + r \cos \varphi, \quad x_{1 \text{ кол}} = x_1^* + r \sin \varphi.$$

У точці D розглянутого кола площинка (5.1) приймає значення

$$y(x_{0 \text{ кол}}, x_{1 \text{ кол}}) = m_0(x_0 + r \cos \varphi),$$

$$\begin{aligned} y(x_{0 \text{ кол}}, x_{1 \text{ кол}}) &= m_0(x_0^* + r \cos \varphi) + m_1(x_1^* + r \sin \varphi) + c = \\ &= m_0 x_0^* + m_1 x_1^* + r(m_0 \cos \varphi + m_1 \sin \varphi) + c. \end{aligned} \quad (5.3)$$

Для визначення напрямку максимального зростання дотичної площинки (5.1) порівняємо відповідно з теоремою 3.1 першу похідну виразу (5.3) по φ

нулю:

$$\frac{\partial y(x_{0 \text{ кол}}, x_{1 \text{ кол}})}{\partial \varphi} = -m_0 r \sin \varphi + m_1 r \cos \varphi = 0,$$

відкля

$$\operatorname{tg} \varphi = \frac{m_1}{m_0}, \quad \varphi = \operatorname{arc} \operatorname{tg} \frac{m_1}{m_0}. \quad (5.4)$$

Функція арктангенс, як відомо, багатозначна; однак, становлять інтерес лише ті її значення, що задовольняють умові $0 \leq \varphi < 360^\circ$. Таких значень два, вони відрізняються один від одного на 180° . Легко перевірити, що значення кута φ , визначеного рівностями

$$\cos \varphi = \frac{m_0}{\sqrt{m_0^2 + m_1^2}}, \quad \sin \varphi = \frac{m_1}{\sqrt{m_0^2 + m_1^2}}, \quad (5.5)$$

по-перше, задовольняє умові (5.4) і, по-друге, забезпечує виконання достатньої умови максимуму, заснованої на від'ємності другої похідної (теорема 3.2). Дійсно,

$$\frac{\partial^2 y(x_{0 \text{ кол}}, x_{1 \text{ кол}})}{\partial \varphi^2} = -m_0 r \cos \varphi + m_1 r \sin \varphi, \quad (5.6)$$

Підставивши (5.5) у (5.6), одержимо значення другої похідної в стаціонарній точці:

$$\frac{\partial^2 y}{\partial \varphi^2} = -\frac{m_0^2 r}{\sqrt{m_0^2 + m_1^2}} - \frac{m_1^2 r}{\sqrt{m_0^2 + m_1^2}} < 0.$$

Отже, вектор, чий проєкції на координатні осі рівні

$$\frac{m_0}{\sqrt{m_0^2 + m_1^2}}, \quad \frac{m_1}{\sqrt{m_0^2 + m_1^2}}, \quad (5.7)$$

або однаково спрямований з ним вектор, проєкції якого на координатні осі збігаються з числами (5.2), а саме, відрізняються від (5.7) множителем $1/\sqrt{m_0^2 + m_1^2}$, вказує напрямок максимального зростання (підйому) дотичної площини $y(x_0, x_1)$, а, виходить, і функції $f(x_0, x_1)$. При додаванні 180° до кута φ , який задається рівностями (5.5), одержимо, вектор, що вказує напрямок максимального спадання (спуску) згаданих поверхонь.

Можна показати, що й у загальному випадку $(n + 1)$ -мірного простору напрямок максимального зростання поверхні вказує вектор з координатами

$$\frac{\partial f}{\partial x_0}, \quad \frac{\partial f}{\partial x_1}, \quad \dots, \quad \frac{\partial f}{\partial x_n},$$

що називається градієнтом, який позначають $\operatorname{grad} f$ або ∇f .

Зауваження 1. Доведено [2, 3, 4], що вектор-градієнт є ортогональним лініям (поверхням) рівня, на яких поверхня приймає постійні значення.

перед відомими функціями з теореми Котельникова типу $\frac{\sin t}{t}$. Саме має місце

теорема: нехай функція $f(t)$, повна енергія, яка дорівнює 1, має найвищу частоту $F_{\bar{n}}$ і нехай

$$\int_{-T/2}^{T/2} |f(t)|^2 dt = 1 - \varepsilon_T^2, \quad (9.103)$$

тоді

$$\inf_{\{a_i\}_{-\infty}^{\infty}} \int_{-\infty}^{\infty} \left| f(t) - \sum_{i=0}^{[2TF_c]} a_i \varphi(t) \right|^2 dt < c \varepsilon_T^2, \quad (9.104)$$

а) справедливо для довільної функції $f(t)$ при $N=0$, $c=12$, якщо φ_n – кутові витягнуті сфероїдальні функції;

б) несправедливо для довільної функції $f(t)$ при будь-якій кінцевій величині N і c , якщо φ_n є відлікові функції з теореми Котельникова [16].

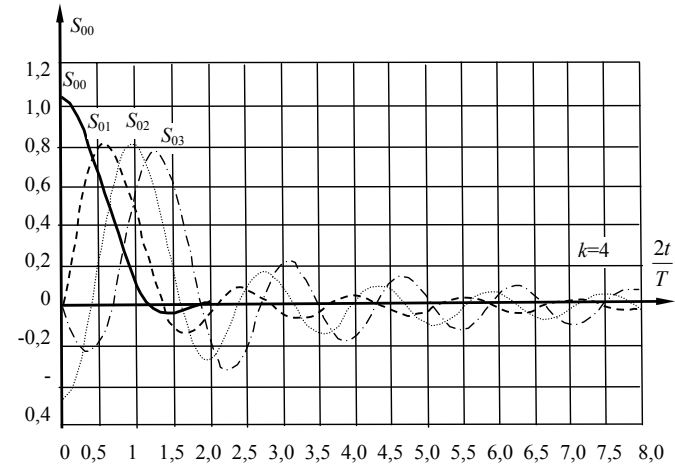


Рисунок 9.21 – Графіки рішення $S(t)$ (рівняння (9.101)) для значень t

У висновку вкажемо ще на один факт, який стосується екстремальних властивостей хвильової сфероїдальної функції. Саме функцією кінцевої тривалості, що найкраще апроксимується в середньоквадратичному результаті її проходження через ідеальний фільтр, є $S_{00}(k, t)$. Фізичне розуміння цього ствердження вельми просте: імпульс $\varphi(t)$ є найменш спотворюваним зі всіх сигналів при роботі у вказаному каналі. Ця властивість є безпосереднім наслідком того факту, що $\varphi(t)$ у заданій смузі несе максимальну частку своєї повної енергії.

Інша послідовність рішення рівняння (9.96) $R_{0n}^{(1)}$, $n = 0, 1, 2, \dots, \infty$, званих радіальними витягнутими сфероїдальними функціями, відрізняється від кутових постійними множниками

$$R_{0n}^{(1)}(k, t) = a_n(k) S_{0n}(k, t) \quad (9.97)$$

Радіальні функції нормовані так, що

$$R_{0n}^{(1)}(k, t) \rightarrow \frac{1}{kt} \cos \left[kt - \frac{1}{2}(n+1)\pi \right], \quad (9.98)$$

коли $t \rightarrow \infty$.

Кутові функції $S_{0n}(k, t)$ задовольняють наступним співвідношенням:

$$\frac{2k}{\pi} [R_{0n}^{(1)}(k, 1)]^2 S_{0n}(k, t) = \int_{-1}^1 \frac{\sin k(t-s)}{t-s} S_{0n}(k, s) ds \quad (9.99)$$

$$2i^n R_{0n}^{(1)}(k, 1) S_{0n}(k, t) = \int_{-1}^1 e^{ikts} S_{0n}(k, s) ds \quad (2.100)$$

$n = 0, 1, 2, \dots, \infty$.

Рівність (9.99) показує, що функції $S_{0n}(k, t)$ є власними функціями рівняння типу (9.78):

$$S(t) = \mu \int_{-1}^1 \frac{1}{\pi} \frac{\sin k(t-s)}{t-s} S(s) ds \quad (9.101)$$

відповідні характеристичним числам

$$\mu_n = \frac{\pi}{2k [R_{0n}^{(1)}(k, 1)]^2}, \quad n = 0, 1, 2, \dots, \infty. \quad (9.102)$$

Через позитивну визначеність ядра $\frac{1}{\pi} \frac{\sin k(t-s)}{t-s}$ всі характеристичні числа μ_n позитивні, а з фізичних міркувань, викладених в попередньому підрозділі, ясно, що $\mu_n > 1$, $n = 0, 1, 2, \dots, \infty$. Як показано, всі характеристичні числа μ_n – прості.

Перша з власних функцій $S_{00}(k, t)$, відповідна μ_0 , дає форму оптимального імпульсу $\bar{\varphi}(t)$.

Рівності (9.99) і (9.100) дозволили зробити висновок аналогічний рівності (9.93).

Графіки рішення рівняння (9.101) для значень t , що лежать поза інтервалом $(-T/2, T/2)$, показані на рис. 9.21 і 9.22. Вони запозичені з роботи [24].

У роботі [25] була виявлена важливіша властивість функцій $S_{0n}(k, t)$, $n = 0, 1, 2, \dots, \infty$, як оптимально скінченомірного базису для представлення функцій з обмеженим спектром, і довели перевагу цього базису

Довжина градієнта, що називається також нормою, визначається рівністю

$$\|\text{grad } f\| = \sqrt{\sum_{k=0}^n (\partial f / \partial x_k)^2}, \quad (5.7)$$

дорівнює швидкості зміни функції $f(x_0, x_1, \dots, x_n)$ у напрямку максимального зростання. В екстремальних точках, де функція стаціонарна, норма градієнта дорівнює нулю, тому що в цих точках дорівнюють нулю всі частинні похідні (4.5).

5.2 Метод найшвидшого підйому (спуску)

Дотримуючись образного викладу монографії [4], розглянемо окрему задачу пошуку вершини пагорба, що сильно заріс лісом і розташованого на низовині. Хоча через лісову хашу подорожанин не тільки не бачить вершини, але навіть не може довідатися про форму пагорба, він зрештою досягає вершини (у випадку унімодалності і, тим більше, опуклості поверхні) просто за рахунок того, що він безупинно піднімається нагору. До вершини його приведе будь-яка дорога, але якщо він квапиться, то, імовірно, він буде рухатися за тими напрямками, де нахил пагорба найбільший, за умови, звичайно, що йому під силу такий підйом.

Ця інтуїтивно приваблива ідея сходження по найкоротшій дорозі є основою методу пошуку, відомого за назвою методу найшвидшого підйому, що належить до класу градієнтних методів. В описаному вище географічному прикладі напрямок найбільш крутого підйому змінюється від точки до точки, але в кожній точці визначається однаково. Цей напрямок збігається, як неважко догадатися, з напрямком перпендикуляра до лінії рівня, або, по-іншому, вектора-градієнта, чії проекції мають вид (4.5).

Зауважимо, що довжина і напрямок градієнта є локальними характеристиками поверхні на противагу такій глобальній властивості (що характеризує всі точки поверхні), як, наприклад, унімодалність.

Існує кілька різновидів градієнтних методів оптимізації, ефективно «працюючих» на опуклих (опуклих униз) цільових функціях, однак загальним для них усіх є циклічна повторюваність наступних операцій. Спочатку виконується група спробних експериментів (або обчислень) для визначення напрямку градієнта. Ці експерименти полягають, власне кажучи, у наближеному (або точному) диференціюванні досліджуваної функції. Далі впливає просування в напрямку градієнта (або в зворотному градієнту напрямку – у випадку мінімізації) на деяким чином вибрану відстань. Потім усе повторюється – впливає новий цикл (ітерація) процесу.

Якщо просування в напрямку градієнта здійснюється аж до точки, де починається спадання функції, то такий алгоритм зветься методом найшвидшого підйому (у задачах мінімізації шукають мінімум у напрямку, зворотному градієнту; тоді ітераційну процедуру називають методом найшвидшого спуску).

Математичний опис переходу від k -ї точки з координатами $(x_0^{(k)}, x_1^{(k)}, \dots, x_n^{(k)})$ до $(k+1)$ -ї точки з координатами $(x_0^{(k+1)}, x_1^{(k+1)}, \dots, x_n^{(k+1)})$ при послідовному наближенні до максимуму методом найшвидшого підйому полягає в наступному:

$$x_l^{(k+1)} = x_l^{(k)} + \alpha^{(k)} \frac{\partial f(x_0^{(k)}, x_1^{(k)}, \dots, x_n^{(k)})}{\partial x_l}, \quad l = 0, 1, \dots, n, \quad (5.8)$$

де величина $\alpha^{(k)}$ вибирається внаслідок одномірної оптимізації – відшукування максимуму поверхні $f(x_0, x_1, \dots, x_n)$ при русі з точки $x^{(k)}$ з координатами $x_0^{(k)}, x_1^{(k)}, \dots, x_n^{(k)}$ у напрямку вектора-градієнта $\nabla f = (\partial f / \partial x_0, \partial f / \partial x_1, \dots, \partial f / \partial x_n)$, розрахованого в точці $x^{(k)}$.

Траєкторію пошуку максимуму двовимірної поверхні методом найшвидшого підйому ілюструє рис. 5.3, де послідовність точок, одержуваних ітераціями, позначена буквами $x^{(k)}$. У цих точках лінія градієнта, обумовлена, наприклад, точкою $x^{(k)}$, стосується поверхні (у даному випадку, лінії) рівня в точці $x^{(1)}$. Таке поведіння траєкторії пошуку породжується його стратегією: рух уздовж лінії градієнта в методі найшвидшого підйому супроводжується переходами з лінії деякого рівня на лінію більш високого рівня доти, поки це можливо, тобто до тієї точки $x^{(k)}$, де подальше просування уздовж лінії градієнта приведе до переходу на поверхню (точніше, лінію) більш низького рівня. Як і в методі перетинів, траєкторія пошуку тут має вигляд сходів; тепер, однак, окремі ланки траєкторії вже не обов'язково паралельні координатним осям. Така вільність у виборі напрямку руху істотно підвищує ефективність пошуку.

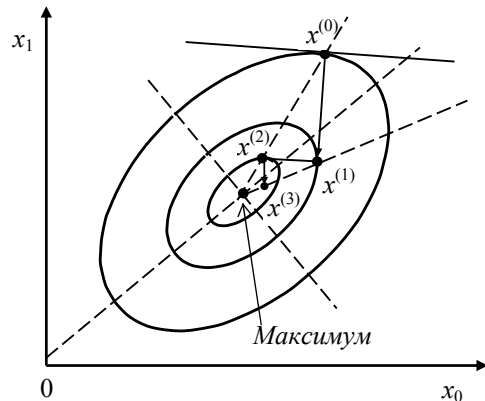


Рисунок 5.3 – Траєкторія пошуку максимуму двовимірної поверхні методом найшвидшого підйому

У випадку задачі відшукування мінімуму усі подані міркування зберігають силу, варто лише замінити рух у напрямку градієнта рухом у зворотному напрямку.

$$|Z(\omega)e^{-j\omega\tau(\varepsilon)} - L(\omega)| < \varepsilon, \quad (9.94)$$

де ε – наперед задана мала величина, а $\tau(\varepsilon)$ – час запізнювання, який вносить лінійна система $L(\omega)$, що забезпечує необхідну близькість до ідеального фільтра.

Позначимо $f(t)$, як і раніше, відгук ідеального фільтра на імпульс $\varphi(t)$, а $f^*(t)$ – відгук фільтра з коефіцієнтом передачі $L(\omega)$ на той же імпульс $\varphi(t)$.

Оцінімо середньоквадратичне значення різниці $f(t) - f^*(t)$. Отримуємо:

$$\int_{-\infty}^{\infty} [f(t) - f^*(t)]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi(\omega)Z(\omega)e^{-i\omega\tau(\varepsilon)} - \Phi(\omega)L(\omega)|^2 d\omega \leq \varepsilon^2 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi(\omega)|^2 d\omega = \frac{\varepsilon^2}{2\pi} \int_{-T/2}^{T/2} \varphi^2(t) dt. \quad (9.95)$$

Звідси ясно, що середньоквадратична помилка, що виникла при заміні реального перехідного процесу ідеалізованим, пропорційна квадрату ε .

Іншими словами, отримані рекомендації з зменшення перехідного процесу можуть бути перенесені на реальні системи з помилкою, що не перевищує (9.95), якщо виконується умова (9.94).

Про деякі властивості вирішень рівняння (9.78). Приведемо короткі відомості про властивості власних функцій і характеристичних чисел рівняння (9.78), які можуть виявитися корисними не лише при розрахунку перехідних процесів, що викликаються імпульсом $\bar{\varphi}(t)$, але і у ряді інших задач теорії зв'язку.

Доказ того факту, що кутові витягнуті сфероїдальні функції є рішеннями рівнянь (9.78), показано в роботі [24], спеціально присвяченій хвильовим сфероїдальним функціям.

Для дійсних k диференціальне рівняння

$$(1-t^2) \frac{d^2 u}{dt^2} - 2t \frac{du}{dt} + (X - k^2 t^2) \cdot u = 0 \quad (9.96)$$

має безперервні рішення в замкнутому інтервалі $(-1, 1)$ тільки для певних дискретних дійсних позитивних значень $0 < X_0(k) < X_1(k) < X_2(k) \dots$ параметра X . Кожному власному значенню $X_n(k)$, $n = 0, 1, 2, \dots, \infty$ відповідає власна функція $S_{0n}(k, t)$, така, що $S_{0n}(k, 0) = P_n(0)$, де $P_n(t)$ – поліном Лежандра. Функції $S_{0n}(k, t)$ називаються кутовими витягнутими сфероїдальними функціями. Це дійсні функції дійсного змінного t , безперервно залежні від параметра k для $k \geq 0$. Вони ортогональні в інтервалі $(-1, 1)$ і утворюють в ньому замкнуту систему ортогональних функцій. У інтервалі $(-1, 1)$ функції $S_{0n}(t)$ мають в точності n нулів. Функції з парними індексами парні, а з непарними – непарні.

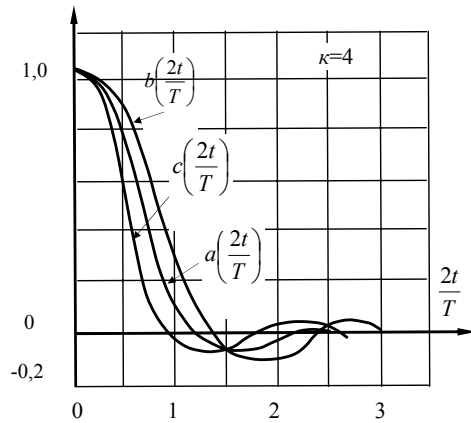


Рисунок 9.20 – Відгуки ідеального фільтра на імпульси $\bar{\varphi}(t)$ і прямокутний імпульс при $k = 4$

Як видно з рис. 9.18 косинусквдратний імпульс дуже близький до $\bar{\varphi}(t)$ для великих значень k . Тому застосування косинусквдратного імпульсу для “завади” при малих k не є виправданим. З рис. 9.19 видно, що при $k = 2$ прямокутний імпульс дає менший перехідний процес, ніж косинусквдратний.

Основний висновок з виконаних розрахунків полягає в тому, що лінійна корекція у вузькосмугових системах стає доцільною при $k > 3 \div 4$. При k , менших вказаної величини, перехідний процес, викликаний проходженням прямокутного імпульсу через ідеальний фільтр, лінійними засобами не може бути значно зменшений.

Теоретичне дослідження цього підрозділу було проведене для “ідеальних” систем, тобто систем, які за частотою зрізу вносять нескінченно велике загасання.

Ідеальний фільтр, як відомо, фізично реалізований бути не може: його імпульсна реакція розтягується в часі від $-\infty$ до ∞ .

Проте, ідеальний фільтр виявляється зручною моделлю для багатьох теоретичних досліджень. Використання ідеальних систем дозволило легко уникнути тривіального рішення задачі: перехідний процес може бути зменшений розширенням смуги частот каналу зв'язку.

Ідеалізована постановка задачі примушує обережно підходити до отриманого рішення. Тут виникає питання: з якою точністю результат, справедливий для ідеальних систем, може бути перенесений на реальні фізичні системи?

Відомо, що ідеальний фільтр (9.70) може бути фізично реалізований за будь-якої міри точності. При цьому, чим ближче амплітудно-частотна характеристика (9.70) апроксимується фізично реалізованою лінійною системою $L(\omega)$, тим більше запізнювання вносить ця система. Це значить, що можна виконати умову

Застосування даного алгоритму до мінімізації середньоквадратичної похибки коректування – опуклої униз функції – виявляється досить ефективним і часто знаходить застосування при дослідженні і проектуванні пристроїв і систем зв'язку.

Зауваження 2. Для збіжності процедури найшвидшого підйому (спуску) властивість опуклості (опуклості вниз) практично не є необхідною; у цій ситуації часто достатнє виконання більш слабкої умови унімодалності. Однак фізична реалізація процесу налаштування коректора можлива лише при наявності відповідного автоматичного пристрою, тому що проведення «ручного» налаштування в цьому випадку вкрай складно.

5.3 Спрощені градієнтні процедури

Одномірний пошук, який варто проводити в методі найшвидшого спуску (підйому), є досить громіздкою процедурою, що вимагає часом великих часових витрат, а також деякого ускладнення програмної реалізації в порівнянні зі спрощеними градієнтними процедурами.

Однією з важливих процедур такого типу є *градієнтний пошук з фіксованим (постійним) кроком*. Координати точок, що відповідають ітераціям, як і раніше опишемо рівністю:

$$x_l^{(k+1)} = x_l^{(k)} + \alpha^{(k)} \frac{\partial f(x_0^{(k)}, x_l^{(k)}, \dots, x_n^{(k)})}{\partial x_l}, \quad l = 0, 1, \dots, n \quad (5.9)$$

так що збільшення кожної з координат при переході з k -ї спробої точки в $(k + 1)$ -у дорівнює

$$x_l^{(k+1)} - x_l^{(k)} = \alpha^{(k)} \frac{\partial f(x_0^{(k)}, x_l^{(k)}, \dots, x_n^{(k)})}{\partial x_l}, \quad l = 0, 1, \dots, n \quad (5.10)$$

Для руху з постійним кроком варто заздалегідь задати, щоб довжина кроку

$$\sqrt{\sum_{l=0}^n (x_l^{(k+1)} - x_l^{(k)})^2} = \alpha^{(k)} \sqrt{\sum_{l=0}^n \left(\frac{\partial f}{\partial x_l} \right)^2} \quad (5.11)$$

залишалася постійною (не залежною від номера ітерації k) величиною. Якщо довжина кроку задана (позначимо її буквою Δ), то величину $\alpha^{(k)}$ на кожному кроці варто вибирати з умови

$$\alpha^{(k)} \sqrt{\sum_{l=0}^n \left(\frac{\partial f}{\partial x_l} \right)^2} = \Delta, \quad (5.12)$$

$$\alpha^{(k)} = \frac{\Delta}{\sqrt{\sum_{l=0}^n \left(\frac{\partial f}{\partial x_l} \right)^2}}. \quad (5.13)$$

Вид траєкторії пошуку екстремуму в даному випадку характеризує рис. 5.4. Як видно, розглянута процедура дозволяє «швидким кроком» наблизитися до точки екстремуму, однак надалі спостерігається блукання в досить «великому» околі екстремальної точки.

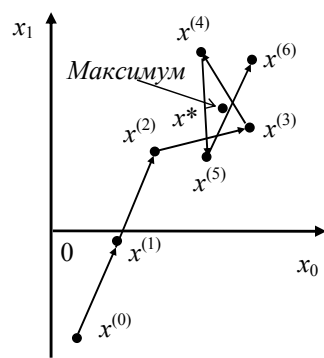


Рисунок 5.4 – Траєкторія пошуку екстремуму при русі з постійним шагом

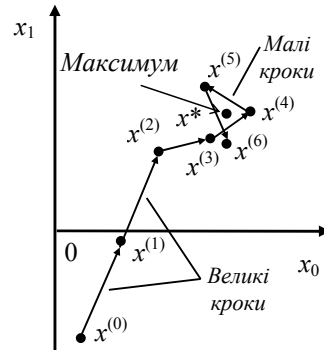


Рисунок 5.5 – Траєкторія пошуку екстремуму при русі зі змінним шагом

Практично зручно використовувати кілька значень величини Δ , наприклад, два. У віддалених від екстремуму точках параметр Δ вибирають досить великим, при підході до екстремальної точки (щоб її не «проскочити») переходять до меншого значення Δ (рис. 5.5).

Описаний метод позбавлений, як сказано, від необхідності проводити на кожнім кроці одномірну оптимізацію. Але він істотно спирається на обчислення точних значень похідних, завдяки чому виявляється можливим рух уздовж лінії градієнта.

Іншою важливою спрощеною градієнтною процедурою є *знаковий алгоритм налаштування коректорів*. Експериментальні дослідження, проведені розроблювачами коректорів [5], показали, що заміна похідних

$$\frac{\partial f}{\partial x_l}, \quad l=0, 1, \dots, n,$$

величинами

$$\text{sign}(\partial f / \partial x_l), \quad l=0, 1, \dots, n,$$

де

$$\text{sign } z = \begin{cases} 1, & z > 0, \\ -1, & z < 0. \end{cases}$$

(при цьому зберігається інформація тільки про знак похідної), дозволяє успішно налаштувати коректори навіть за допомогою такої спрощеної градієнтної процедури, що використовує, до того ж, кроки фіксованої довжини (рис. 5.6). Даний алгоритм налаштування коректорів, що одержав назву *знакового* [5], знайшов широке поширення в апаратурі високошвидкісної передачі дискретної інформації (при цьому реалізують автоматичний режим налаштування).

Виконані розрахунки показують, що при $k \leq 2$ практично немає відмінності у відгуках фільтра (9.71) на імпульси $\bar{\varphi}(t)$ і прямокутний. Перевага $\bar{\varphi}(t)$ перед прямокутним імпульсом позначається у міру збільшення параметра k ; при $k = 4$ воно вельми помітне: відношення амплітуди першої осциляції до максимального значення сигналу при відгуку фільтра (9.70) на прямокутний імпульс удвічі більше, ніж при відгуку на $\bar{\varphi}(t)$. Крім того, амплітудне значення першої осциляції у разі прямокутного імпульсу знаходиться в 1,3 рази далі від максимального значення, ніж у разі $\varphi(t)$.

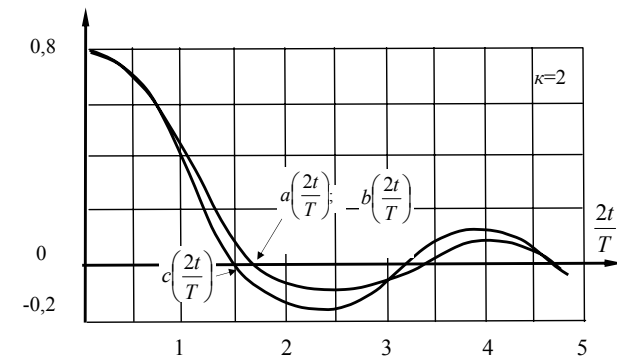


Рисунок 9.19 – Відгуки ідеального фільтра на імпульси $\bar{\varphi}(t)$ і прямокутний імпульс при $k = 2$

Рис. 9.19 і 9.20 ілюструють сказане. Тут $a(t)$ і $b(t)$ позначають відповідно відгуки ідеального фільтра на імпульси $\bar{\varphi}(t)$ і прямокутний імпульс ($a(0)$ вибрано з умови $\int_{-\infty}^{\infty} a^2(t) dt = 1$). На рис. 9.19 функції $a(t)$ і $b(t)$ зображені однією кривою, оскільки відмінність між ними менше точності виконання рисунка.

Крива $c(t)$ відповідає проходженню через ідеальний фільтр косинусквадратного імпульсу.

$$\bar{\Phi}(\omega) = \mu \int_{-\Omega}^{\Omega} \frac{1}{\pi} \frac{\sin T/2(\omega - y)}{\omega - y} \bar{\Phi}(y) dy. \quad (9.92)$$

Відповідно до сказаного позначимо через $\bar{\Phi}(\omega)$ ту із допустимих функцій $F(\omega)$, на якій функціонал (9.90) приймає стаціонарне значення. Порівнюючи рівняння (9.78) і (9.92) приходимо до висновку, що функція $\bar{\Phi}(\omega)$, $-\infty < \omega < \infty$ є цілою функцією, перетворення Фур'є якої є функція, обмежена за часом або точніше, імпульс тривалості від $-T/2$ до $T/2$. Очевидно, цей імпульс і є $\bar{\varphi}(t)$.

Порівняння інтегральних рівнянь (9.78) і (9.92) дозволяє зробити ще один важливий висновок:

$$\bar{\varphi}\left(\frac{r}{\Omega}\right) = \alpha \bar{\Phi}\left(\frac{2r}{T}\right), \quad -\frac{T\Omega}{2} \leq r \leq \frac{T\Omega}{2}, \quad (9.93)$$

який справедливий, при простому характеристичному числі μ_{\min} ; де α можна визначити за допомогою теореми Релея.

Підсумовуючи викладене, можемо сказати, що імпульс $\bar{\varphi}(t)$ є максимально ущільненим за частотою. При проходженні через ідеальний фільтр він перетворюється на функцію, максимально ущільнену в часі, повторюючи форму $\bar{\varphi}(t)$ у інтервалі первинної тривалості. Спектр імпульсу $\varphi(t)$ пов'язаний з ним співвідношенням (9.92), тобто подвійний йому у розумінні перетворення Фур'є. Система з частотою зрізу Ω дає на виході максимально сконцентровану в часі функцію, якщо добуток спектрів імпульсу і системи в смузі пропускання збігаються з $\bar{\Phi}(\omega)$.

Звідси ясно, що корекція повинна зводитися до підбору такого імпульсу при заданій лінійній системі (або фільтра – коректора при заданому кодовому сигналі), щоб спектр сигналу на виході збігався з $\bar{\Phi}(\omega)$ у смузі пропускання системи.

Фізичний сенс найменшого характеристичного числа μ_{\min} полягає в наступному. Імпульс $\bar{\varphi}(t)$ у смузі частот $(-\Omega, \Omega)$ містить $\frac{1}{\mu_{\min}}$ від своєї повної енергії, а після проходження ідеального фільтра в інтервал первинної тривалості потрапляє $\frac{1}{2\mu_{\min}}$ від його повної енергії.

Останнє ствердження виходить з доведеного нами факту, що, проходячи через фільтр (9.70), сигнал $\bar{\varphi}(t)$ не змінює форми, але зменшується за амплітудою в μ_{\min} раз.

Представляє інтерес порівняти (при різних значеннях k) перехідний процес на виході ідеального фільтра при подачі на його вхід $\bar{\varphi}(t)$ і прямокутного імпульсу тієї ж тривалості. Природно чекати, що оскільки $\bar{\varphi}(t)$ дає мінімальний перехідний процес в сенсі енергії, то і амплітуди осциляції відгуку фільтра (9.70) на $\bar{\varphi}(t)$ поза інтервалом первинної тривалості $(-T/2, T/2)$ будуть малі.

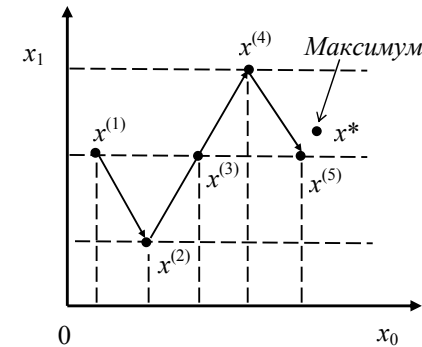


Рисунок 5.6 – Траєкторія пошуку екстремуму знаковим методом

5.4 Метод сполучених градієнтів

Швидкість збіжності методу найшвидшого підйому (спуску) при розв'язанні ряду обчислювальних задач виявляється недостатньою. Це цілком з'ясовано, тому що змінення змінних x_0, x_1, \dots, x_n при переході від однієї ітерації до наступної стають все меншими (див. рис. 5.3) і в міру наближення до екстремуму спостерігається істотне уповільнення процесу. Метод найшвидшого підйому (спуску), досить ефективний при русі з точок, далеких від екстремуму (і тому часто використовується на початковому етапі оптимізації), може бути істотно поліпшений на завершальному етапі пошуку екстремуму. (Ця обставина показує, що назва методу найшвидшого підйому чи спуску є у відомому розумінні невдалою).

Сформулюємо одну з можливих ідей поліпшення методу найшвидшого підйому. У загальному випадку траєкторія пошуку максимуму має вид ламаної лінії $x^{(0)}, x^{(1)}, x^{(2)}$, (див. рис. 5.3). При цьому зигзагоподібна траєкторія пошуку $x^{(1)}, x^{(2)}, x^{(3)}, \dots$ цілком лежить в області, обмеженій двома штрих пунктирними лініями, що перетинаються в точці екстремуму. Це наводить на думку про те, що пошук із точки $x^{(1)}$ варто вести не в напрямку лінії градієнта до точки $x^{(2)}$, а уздовж прямої $x^{(1)}, x^{(3)}$. У цьому випадку точка максимуму може бути знайдена після трьох одномірних переміщень від точки $x^{(1)}$ до $x^{(2)}$ у напрямку градієнта, від точки $x^{(2)}$ до $x^{(3)}$ також у напрямку градієнта і з точки $x^{(1)}$ у напрямку прямої $x^{(1)}, x^{(3)}$.

Викладений метод прискорення найшвидшого підйому являє собою двовимірний варіант прискореного методу паралельних дотичних (ПМПД) [4].

Однак найбільшу популярність серед фахівців, що використовують при дослідженнях, проектуванні і розрахунках методи оптимізації, здобув метод сполучених градієнтів, суть якого зводиться до наступного.

Початкові операції тут такі ж, як і в методі найшвидшого підйому: визначається напрямок градієнта в початковій точці $x^{(0)}$ (рис. 5.7) і робиться крок у напрямку градієнта (або в зворотному напрямку – при відшуванні мінімуму) до точки $x^{(1)}$, що максимізує функцію f на прямій лінії, що визначається вектором-градієнтом.

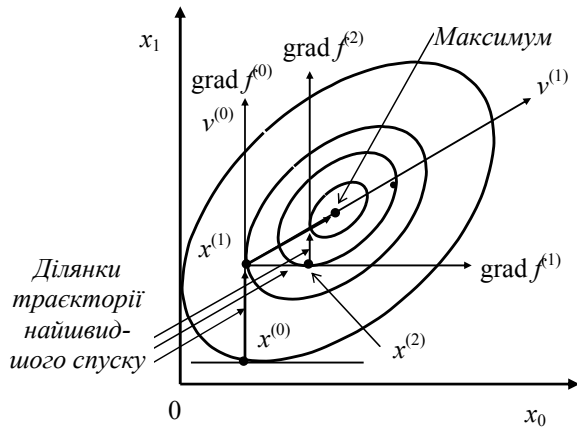


Рисунок 5.7 – Траєкторія пошуку екстремуму за методом сполучених градієнтів

Знайдений на цьому етапі градієнт нормується

$$v = \frac{\text{grad } f^{(0)}}{\|\text{grad } f^{(0)}\|}, \quad (5.14)$$

При цьому проєкції вектора $v^{(0)}$ мають вид

$$v_l^{(0)} = \frac{\partial f^{(0)} / \partial x_l}{\sqrt{\sum_{k=0}^n [\partial f^{(0)} / \partial x_k]^2}}, \quad l=0, 1, \dots, n, \quad (5.15)$$

У точці $x^{(1)}$ аналогічним чином обчислюється градієнт $\text{grad } f^{(1)}$, але крок робиться у напрямку вектора $v^{(1)}$, який є лінійною комбінацією (тобто сумою з деякими коефіцієнтами) вектора $v^{(0)}$ і знову знайденого вектора-градієнта (рис. 5.7):

$$v_l^{(1)} = \frac{\frac{\partial f^{(1)}}{\partial x_l} + \beta^{(1)} v_l^{(0)}}{\sqrt{\sum_{k=0}^n \left[\frac{\partial f^{(1)}}{\partial x_k} + \beta^{(1)} v_k^{(0)} \right]^2}}, \quad l=0, 1, \dots, n, \quad (5.16)$$

де

$$\beta^{(1)} = \frac{\sum_{k=0}^n \left[\frac{\partial f^{(1)}}{\partial x_k} \right]^2}{\sum_{k=0}^n \left[\frac{\partial f^{(0)}}{\partial x_k} \right]^2}. \quad (5.17)$$

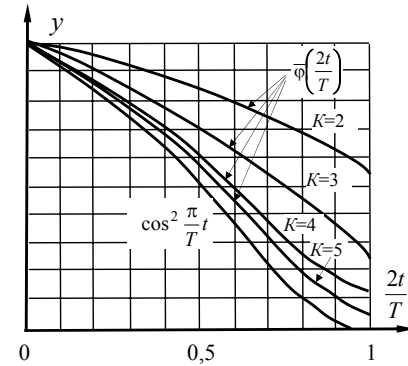


Рисунок 9.18 – Залежність $\bar{\varphi}(t)$ при різних значеннях параметра $k = \Omega T/2$

Інакше кажучи, імпульс необхідно погоджувати з системою так, щоб на виході системи був спектр, відповідний функції $f(t)$, щонайкраще енергетично сконцентрованої у часі.

Сказане можна покласти в основу іншого підходу до задачі. Потрібно знайти $F(t)$ – спектр на виході системи з скінченною шириною смуги пропускання, рівним нулю при $|\omega| > \Omega$, такий щоб енергія “завади”

$$W_n = \int_{-\infty}^{\infty} f^2(t) dt - \int_{-T/2}^{T/2} f^2(t) dt \quad (9.88)$$

була мінімальною, тобто ми приходимо до задачі М.С. Гуревича, з тією лише різницею, що вісь часу і частот помінялися місцями.

Тут

$$f(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{j\omega t} d\omega. \quad (9.89)$$

Замість визначення $\bar{\varphi}(t)$ можна розшукати його перетворення Фур'є $\bar{\Phi}(\omega)$ у інтервалі $(-\Omega, \Omega)$, що однозначно визначить $\Phi(\omega)$ на всій вісі частот, а отже, і $\bar{\varphi}(t)$. Покажемо це.

Дослідження на стаціонарність функціоналу

$$\int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{j\omega t} d\omega \right]^2 dt - \mu \int_{-T/2}^{T/2} \left[\frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{j\omega t} d\omega \right]^2 dt, \quad (9.90)$$

або

$$\frac{1}{2\pi} \int_{-\Omega}^{\Omega} |F(\omega)|^2 d\omega - \frac{\mu}{2\pi^2} \int_{-\Omega}^{\Omega} \int_{-\Omega}^{\Omega} \frac{\sin T/2(\omega - y)}{\omega - y} F(\omega) F^*(y) d\omega dy \quad (9.91)$$

приводить до рівняння

З доказу виходить, що імпульсом оптимальної форми $\bar{\varphi}(t)$ є перша власна функція рівняння (9.78), відповідна $\mu_{\min} - \mu_0$.

У роботах Ландау і Полака показано, що власні функції рівняння (9.78) збігаються з добре відомими кутовими витягнутими сфероїдальними функціями, які зустрічаються в різних задачах математичної фізики.

Відзначимо, що рівняння (9.78) можна за рахунок заміни змінних привести до виду, де межі інтегрування будуть $-\frac{\Omega T}{2}$ і $\frac{\Omega T}{2}$, а в підінтегральному виразі – Ω і T , а від їх добутку, що позначається далі скорочено

$$2k = \Omega T. \quad (9.84)$$

Рис. 9.18. дає деяке уявлення про форму $\bar{\varphi}(t)$ при різних значеннях параметра $k = \Omega T/2$. Для порівняння надамо графік косинусквадратного імпульсу

$$\bar{\varphi}(t) = \mu \int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t - \tau)}{t - \tau} \varphi(\tau) d\tau.$$

Рівняння (9.78) отримане Чоком і Гуревичем для імпульсу, що несе максимальну частину повної енергії в заданій смузі частот.

Таким чином, імпульс з максимально сконцентрованим спектром (у сенсі енергії) при проходженні через ідеальний фільтр дає мінімальний наслідок (також в сенсі енергії).

Відзначимо одну властивість $\bar{\varphi}(t)$, що слідує з фізичного сенсу рівняння (9.78). При проходженні через систему (9.70) імпульс $\bar{\varphi}(t)$ не змінює своєї форми в інтервалі $(-T/2, T/2)$, а тільки зменшує амплітуду в μ_{\min} раз і аналітично продовжується з інтервалу $(-T/2, T/2)$ на всю вісь часу $(-\infty, \infty)$.

Якщо замість ідеального фільтра розглянути іншу вузькосмугову систему $K(\omega)$ з тією ж частотою зрізу Ω , то, очевидно, імпульс $\bar{\varphi}(t)$ потрібно замінити імпульсом $\psi(t)$ так, щоб

$$\Psi(\omega) \cdot K(\omega) = \Phi(\omega), \quad (9.85)$$

де

$$\Phi(\omega) = \int_{-T/2}^{T/2} \varphi(t) e^{-j\omega t} dt, \quad (9.86)$$

$$\Psi(\omega) = \int_{-T/2}^{T/2} \psi(t) e^{j\omega t} dt. \quad (9.87)$$

Новий напрямок називається «сполученим». За цим напрямком знову проводиться одномірний пошук і знаходиться екстремальна точка $x^{(2)}$. Якщо необхідний ступінь близькості значень функції $f(x)$ у сусідніх спробних точках у даному випадку $x^{(1)}$ і $x^{(2)}$ не досягнутий, знову визначається градієнт $\text{grad } f^{(2)}$ і дається переміщення в напрямку вектора $v^{(1)}$, що має бути рівний лінійній комбінації вектора $v^{(1)}$ і градієнта $\text{grad } f^{(2)}$, і т. ін.

Якщо функція є квадратичною, як, наприклад, $D^2(g_0, g_1, \dots, g_n)$, то знаходження екстремуму гарантується, у точності за $n + 1$ кроків, як це показано на рис. 5.7 (нагадаємо, що $n + 1$ – число незалежних змінних). У загальному ж випадку приходиться переходити до нового циклу перебору всіх змінних g_0, g_1, \dots, g_n .

Даний метод, очевидно, є найбільш ефективним серед відомих методів, що використовують дані лише про перші похідні.

Контрольні питання та задачі до розділу 5

5.1 Дайте визначення вектора-градієнта.

5.2 Що характеризує довжина (норма) градієнта? Що характеризує напрямок градієнта?

5.3 Чому дорівнює норма градієнта в точці мінімуму (максимуму)?

5.4 У чому суть методу найшвидшого спуску (підйому)?

5.5 У чому полягає ефективність градієнтних методів пошуку екстремуму?

5.6 У чому полягає відмінність траєкторії пошуку точки екстремуму за методом найшвидшого спуску від аналогічної траєкторії за методом перетинів?

5.7 У чому суть градієнтної процедури з постійним кроком?

5.8 Чому при наближенні до екстремуму довжину кроку бажано зменшити?

5.9 У чому суть знакового алгоритму мінімізації середньоквадратичної похибки коректування?

5.10 Які об'ємні обчислювальні процедури дозволяє "обійти" знаковий алгоритм?

5.11 Які напрямки руху виявляються дозволеними в знаковому алгоритмі?

5.12 Задана функція

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

і дві перші точки, отримані в процесі пошуку точки мінімуму функції f : $x^{(0)} = [-1, 2, 1]^T$, $x^{(1)} = [1, 3, 1, 07]^T$.

Визначте напрямок пошуку з точки $x^{(1)}$ користуючись методом найшвидшого спуску.

5.13 Задана функція

$$q(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2,$$

початкова точка $x^{(0)} = [10, 10]^T$ і два лінійно незалежних напрямки $d^{(0)} = \text{grad } q(x^{(0)}) = [200, 140]^T$, $d^{(1)} = [7, -10]^T$.

Визначте новий напрямок пошуку ($S^{(1)} = d^{(1)} + \beta_{11} \Delta g^{(1)}$), сполучений з $d^{(0)}$. Використовуйте ці напрямки при пошуку точки x^* : спочатку проведіть пошук у напрямку $S^{(0)} = d^{(0)}$, потім з отриманої точки мінімуму проведіть пошук у напрямку $S^{(1)}$.

5.14 Знайдіть напрямок, що є ортогональним вектору

$$s_1 = \left[\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right]^T, \quad x = [0, 0, 0]^T.$$

Знайдіть також напрямок s_2 , сполучений з s_1 , у тій же точці за умови, що цільова функція дорівнює

$$f(x) = x_1 + 2x_2^2 - x_1x_2 + 3x_1^2 - 2x_1x_3 + x_3^2.$$

Література до розділу 5

1 Захарченко Н.В. Оптимизация и моделирование систем связи: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, 1988. – Ч. 1. – 86 с.

2 Реклейтис Г. Оптимизация в технике: В 2-х кн. – Кн.1 / Реклейтис Г., Рейвиндран А., Рэгсдел К. – М.: Мир, 1986. – 352 с.

3 Моисеев Н.Н. Методы оптимизации / Н.Н. Моисеев, Н. Н. Иванилов, Е.М. Столярова – М.: Наука, 1978.

4 Уайлд Д. Дж. Методы поиска экстремума / Уайлд Д. Дж. – М.: Наука, 1967.

5 Тамм Ю.А. Адаптивная коррекция сигнала ПД / Ю.А. Тамм – М.: Связь, 1978.

При цьому перетворенні використовується основна лема варіаційного числення.

Неважко показати, що рівнянню (9.77) задовольняють тільки власні функції одного інтегрального рівняння типу Фредгольма з симетричним позитивно визначеним ядром:

$$\tilde{\varphi}(t) = \mu \int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \varphi(\tau) d\tau \quad (9.78)$$

при значеннях μ , що співпадають з характеристичними числами рівняння (9.78).

Останнє твердження витікає з властивостей ітерированих ядер, у яких власні функції не відрізняються від власних функцій початкового ядра, а характеристичні числа виходять і характеристичних чисел початкового ядра піднесенням до степені.

Безпосередньою підстановкою переконаємося в тому, що власні функції (9.70) задовольняють рівнянню (9.77).

Переконаємося тепер, що функції, що не задовольняють рівнянню (9.78), не задовольняють також рівнянню (9.77). Передбачимо протилежне. Позначимо рішення (9.77), що не задовольняє (9.78) через $\tilde{\varphi}(\tau)$.

Рівняння (9.77) можна розглядати як два рівняння Фредгольма першого роду, по черзі позначаючи ліву і праву частину однією і тією ж функцією $g(t)$, тобто

$$\begin{cases} \int_{-T/2}^{T/2} \tilde{\varphi}(\tau) \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} d\tau = g(t), \end{cases} \quad (9.79)$$

$$\begin{cases} \mu \int_{-T/2}^{T/2} \tilde{\varphi}(\tau) \int_{-T/2}^{T/2} \frac{1}{\pi^2} \frac{\sin \Omega(t-s)}{t-s} \cdot \frac{\sin \Omega(s-\tau)}{s-\tau} ds d\tau = g(t). \end{cases} \quad (9.80)$$

З теорії рівнянь Фредгольма першого роду відомо, що вирішення рівнянь (9.58) і (9.59) мають, відповідно, вид:

$$\tilde{\varphi}(\tau) = \lim_{m \rightarrow \infty} \sum_{n=0}^m a_n \mu_n \varphi_n(x), \quad (9.81)$$

$$\varphi(\tau) = \lim_{m \rightarrow \infty} \frac{1}{\mu} \sum_{n=0}^m a_n \mu_n^2 \varphi_n(\tau), \quad (9.82)$$

де μ_n і φ_n , $n = 0, 1, 2, \dots, \infty$ – характеристичні числа і власні функції рівняння (9.78), а

$$a_n = \int_{-T/2}^{T/2} g(t) \varphi_n(t) dt. \quad (9.83)$$

Оскільки характеристичні числа (9.78) μ_n прості і більші одиниці, то одночасне виконання (9.81) і (9.82) неможливе. Це є доказом другої частини твердження.

З'ясуємо, яка форма імпульсу $\varphi(t)$ тривалості T буде з мінімальною енергією «завади»

$$W_n = W_{-\infty, \infty} - W_{-T/2, T/2} = \int_{-\infty}^{\infty} f^2(t) dt - \int_{-T/2}^{T/2} f^2(t) dt \quad (9.72)$$

при проходженні через ідеальний фільтр.

Повна енергія і енергія імпульсу $\varphi(t)$ в інтервалі $(-T/2, T/2)$ на виході фільтра відповідно рівні

$$\begin{aligned} W_{-\infty, \infty} &= \int_{-\infty}^{\infty} \left[\int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \varphi(\tau) d\tau \right]^2 dt = \\ &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \int_{-\infty}^{\infty} \frac{1}{\pi^2} \frac{\sin \Omega(t-\tau)}{t-\tau} \cdot \frac{\sin \Omega(t-s)}{t-s} dt d\tau ds = \\ &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \frac{1}{\pi} \frac{\sin \Omega(\tau-s)}{\tau-s} d\tau ds, \end{aligned} \quad (9.73)$$

$$W_{-T/2, T/2} = \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \int_{-T/2}^{T/2} \frac{1}{\pi^2} \frac{\sin \Omega(t-\tau)}{t-\tau} \cdot \frac{\sin \Omega(t-s)}{t-s} dt d\tau ds. \quad (9.74)$$

Для відшукування імпульсу, що дає мінімум (9.72), проведемо дослідження функціоналу

$$\begin{aligned} \mathcal{J}[\varphi] &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \frac{1}{\pi} \frac{\sin \Omega(\tau-s)}{\tau-s} d\tau ds - \\ &- \mu \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \varphi(\tau) \varphi(s) \int_{-T/2}^{T/2} \frac{1}{\pi^2} \frac{\sin \Omega(t-\tau)}{t-\tau} \cdot \frac{\sin \Omega(t-s)}{t-s} dt d\tau ds. \end{aligned} \quad (9.75)$$

Необхідно знайти функцію (що надалі позначається $\bar{\varphi}(t)$), для якої $\delta \mathcal{J}[\varphi] = 0$ при найменшому значенні параметра $\mu = \mu_{\min}$.

Вводячи довільну функцію $\zeta(t)$ і параметр ε , отримуємо наступне рівняння для визначення μ_{\min} і $\bar{\varphi}(t)$:

$$\frac{\partial}{\partial \varepsilon} \mathcal{J}[\bar{\varphi}(t) + \varepsilon \zeta(t)]_{\varepsilon=0} = 0 \quad (9.76)$$

Рівняння (9.76) перетворюється до виду:

$$\begin{aligned} &\int_{-T/2}^{T/2} \bar{\varphi}(\tau) \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} d\tau = \\ &= \mu \int_{-T/2}^{T/2} \bar{\varphi}(\tau) \int_{-T/2}^{T/2} \frac{1}{\pi^2} \frac{\sin \Omega(t-s)}{t-s} \cdot \frac{\sin \Omega(s-\tau)}{s-\tau} ds d\tau. \end{aligned} \quad (9.77)$$

6. ОСНОВИ ВАРІАЦІЙНОГО ЧИСЛЕННЯ

6.1 Класична задача про брахістохрону і поняття функціоналу

На дитячих майданчиках в районах новобудов майже завжди є «устаткування», за допомогою якого діти із задоволенням використовують рішення однієї класичної задачі, що привертає увагу кращих фахівців з математики кінця сімнадцятого століття. Мова йде про дитячу «гірку» (рис. 6.1).

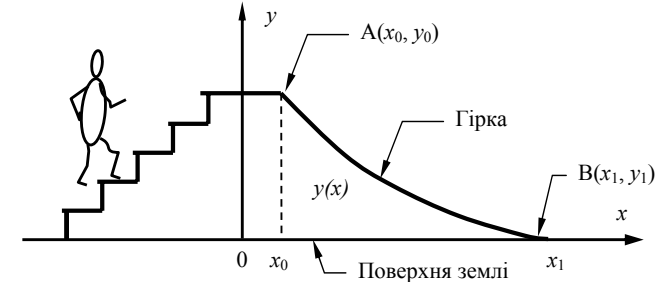


Рисунок 6.1 – Дитяча гірка

Дитина сідає на «гірку» в точці А і з'їжджає в точку В. Який має бути профіль «гірки» $y(x)$, щоб з'їхати з точки А у точку В за мінімальний час $T_{AB}[y(x)]$?

Іоганн Бернуллі в 1696 році опублікував лист, в якому пропонував математикам знайти криву найшвидшого спуску – брахістохрону. У цій задачі потрібно визначити лінію, що з'єднує дві задані точки А і В, які не лежать на одній вертикальній прямій та мають таку властивість, що матеріальна точка (кулька з малим радіусом, яким можна нехтувати) скотиться по цій лінії з точки А в точку В у найкоротший час.

Неважко показати, що лінією якнайшвидшого спуску не буде пряма, що з'єднує точки А і В, хоча вона і є найкоротшою відстанню між ними, оскільки при русі по прямій швидкість руху наростатиме порівняно повільно; якщо ж ми візьмемо криву, що крутіше спускається біля точки А вниз, то хоча шлях і подовжиться, але значна частина його буде пройдена з великою швидкістю. Рішення задачі про брахістохрону було дане І.Бернуллі, Я.Бернуллі, Ньютоном, Лопіталем. Виявилось, що лінією якнайшвидшого скочування є циклоїда.

Відзначимо, що в задачі І.Бернуллі мова йде про мінімум T_{AB} на деякому спеціальному класі функцій $y(x)$, а саме, усі порівнювані функції $y(x)$ повинні задовольняти умовам:

- 1) $y(x_0) = y_0$,
- 2) $y(x_1) = y_1$.

Варіаційне числення почало розвиватися із задачі про брахістохрону, але оформилося в самостійну математичну дисципліну після фундаментальних робіт петербурзького академіка Л. Ейлера.

Чим же характеризується задача про брахістохрону, що породила новий розділ математичної науки?

Поряд із задачами, в яких необхідно визначити максимальні і мінімальні значення деякої функції, в інженерній практиці нерідко виникає необхідність знайти максимальні і мінімальні значення величин особливого роду, званих функціоналами. Функціоналами називаються змінні величини, значення яких визначаються вибором однієї або декількох функцій [1].

Наприклад, в задачі про брахістохрону [2] функціоналом є час руху з точки А в точку В, оскільки ця величина визначається вибором профілю «гірки», функції $y(x)$, тобто $T_{AB} = T_{AB}[y(x)]$.

Інший приклад. При передачі імпульсних сигналів за смугово-обмеженим каналом зв'язку енергія позасмугових випромінювань не повинна перевищувати деякої нормативної величини. Позначимо сигнал через $y(t)$, де $0 \leq t \leq T$, а його спектр – через $Y(\omega)$. Як відомо, сигнал пов'язаний зі своїм спектром парою перетворень Фур'є:

$$Y(\omega) = \int_0^T y(t) e^{-i\omega t} dt, \quad (6.1)$$

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) e^{i\omega t} d\omega. \quad (6.2)$$

Інтенсивність випромінювання енергії поза діапазоном частот (ω_1, ω_2) характеризуватимемо відношенням енергії сигналу поза цим діапазоном до повної енергії сигналу:

$$\mu = \frac{\frac{1}{\pi} \int_0^{\infty} |Y(\omega)|^2 d\omega - \frac{1}{\pi} \int_{\omega_1}^{\omega_2} |Y(\omega)|^2 d\omega}{\frac{1}{\pi} \int_0^{\infty} |Y(\omega)|^2 d\omega}, \quad (6.3)$$

де $\frac{1}{\pi} \int_0^{\infty} |Y(\omega)|^2 d\omega$ – повна енергія сигналу, а $\frac{1}{\pi} \int_{\omega_1}^{\omega_2} |Y(\omega)|^2 d\omega$ – його енергія у діапазоні частот (ω_1, ω_2) .

Підставивши (6.1) в (6.3) ми виявляємо, що $\mu = \mu[y(t)]$, тобто відносна частка енергії сигналу в заданому діапазоні частот (ω_1, ω_2) є функціоналом від форми цього сигналу. У багатьох практичних завданнях потрібні сигнали, що мінімізують функціонал $\mu = \mu[y(t)]$.

зберігав би в інтервалі початкової тривалості максимальну долю повної енергії, тобто давав би в енергетичному трактуванні мінімальний перехідний процес.

У роботах Полака і Ландау [24] приведено більш точніше рішення цієї задачі, тому використаємо розрахунки за оцінкою інтенсивності перехідних процесів цих авторів. Ландау і Полак довели енергетичний “варіант” принципу невизначеності.

Якщо

$$\frac{\int_{t_0-T/2}^{t_0+T/2} |\varphi(t)|^2 dt}{\int_{-\infty}^{\infty} |\varphi(t)|^2 dt} = \alpha^2 \quad (9.68)$$

і

$$\frac{\int_{-\Omega}^{\Omega} |F(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |F(\omega)|^2 d\omega} = \beta^2, \quad (9.69)$$

тоді $\Omega T \geq \Phi(\alpha, \beta)$, де $\Phi(\alpha, \beta)$ і функція $\varphi(t)$, для якої має місце рівність, вказані.

Відзначимо, що зменшення перехідного процесу на виході НЧ фільтра грає суттєву роль у вокодерній техніці.

Фільтри, через які передаються коефіцієнти розкладання обвідної миттєвого спектра мови в ряд за системою ортогональних функцій, повинні мати малий перехідний процес; інакше виявляються характерні дефекти звучання в синтезованій мові.

Про коректування перехідного процесу на виході фільтра низьких частот. Нехай маємо ідеальний фільтр

$$Z(\omega) = \begin{cases} 1; & |\omega| \leq \Omega \\ 0; & |\omega| > \Omega. \end{cases} \quad (9.70)$$

Фазові зсуви передбачаємо рівними нулю, що, як стане ясним з подальшого, не зменшить спільності розгляду.

Якщо на вхід такого фільтра подати імпульс $\varphi(t)$, то на виході з урахуванням формули Дюамеля [3], отримаємо:

$$f(t) = \int_{-T/2}^{T/2} \frac{1}{\pi} \frac{\sin \Omega(t-\tau)}{t-\tau} \varphi(\tau) d\tau. \quad (9.71)$$

Тут $(-T/2, T/2)$ – інтервал тривалості імпульсу $\varphi(t)$. Всі значення $f(t)$, відповідні моментам часу поза інтервалом $(-T/2, T/2)$, називатимемо “завадою”.

Далі розглядатимемо спеціальний випадок лінійної корекції перехідних процесів, де критерієм інтенсивності перехідного процесу буде прийнята енергія.

Для тільки лінійних методів коректування задача полягає у виборі такого коефіцієнта передачі каналу зв'язку при заданій формі кодового сигналу (або виборі форми кодового сигналу при заданому коефіцієнті передачі каналу), щоб перехідний процес на виході каналу був мінімальним. Грубо кажучи, задача полягає у відшукуванні сигналу, який, займаючи дану смугу частот, мав би найменшу протяжність в часі. Неточність останнього формулювання полягає в тому, що сигнали з обмеженим спектром не обертаються тотожно в нуль на всій часовій осі. Мова йде, очевидно, про якийсь умовний спосіб виміру тривалості сигналу.

Відзначимо, що в приведеній, хоча ще і не чітко сформульованій постановці, задача виглядає добре знайомою: тут явно видно зв'язок з принципом невизначеності Гейзенберга, математична суть якого зводиться до наступного. Якщо зміряти протяжність функції $\varphi(t)$ у часі виразом

$$T^2 = \frac{\int_{-\infty}^{\infty} (t - t_0)^2 |\varphi(t)|^2 dt}{\int_{-\infty}^{\infty} |\varphi(t)|^2 dt}, \quad (9.66)$$

а протяжність її перетворення Фур'є в частотній області виразом

$$\Omega^2 = \frac{\int_{-\infty}^{\infty} (\omega - \omega_0) |F(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |F(\omega)|^2 d\omega}, \quad (9.67)$$

тоді, при будь-якому виборі t_0 , ω_0 і $\varphi(t)$, $\Omega T \geq 1/2$, рівність має місце лише для гаусової дзвіноподібної кривої [15].

Чок і Гуревич з'ясували [24], яка форма імпульсу заданої тривалості, що має в заданій смузі частот максимальну долю повної енергії. Якщо вважати шириною спектра імпульсу смугу частот, в якій міститься 99 % повної енергії, то форма імпульсу відповідає найбільш короткому з фізичної трактовки енергії спектра. Тим самим реалізується мінімальне ΩT , де T – тривалість імпульсу в загальноприйнятому сенсі, а Ω – умовна ширина спектра, що оцінюється процентним вмістом в ній енергії імпульсу.

Пошук форми імпульсу тривалості T , що займає найменшу смугу частот, є дуже важливим при частотному ущільненні ліній зв'язку, оскільки цій формі імпульсу відповідають найменші позасмугові випромінювання. Питанню про зменшення позасмугових випромінювань присвячена монографія М.С. Гуревича.

У багатьох роботах авторів [16] рішалася задача про відшукування форми імпульсу, який при проходженні через ідеальний фільтр низьких частот

Ще одним прикладом функціоналу може служити залежність імовірності помилки одиночного елемента P_0 в системі передачі даних (ПД) від форми амплітудно-частотної характеристики $|K(\omega)|$ каналу зв'язку

$$P_0 = P_0[|K(\omega)|, \varphi(\omega)]. \quad (6.4.)$$

При цьому мається на увазі, що передаточну функцію каналу, як завжди, можна представити у виді:

$$K(\omega) = |K(\omega)| e^{-i\varphi(\omega)}. \quad (6.5)$$

Функціонал P_0 , як видно з (6.4), залежить від двох змінних, і його мінімізацію здійснюють апаратно за допомогою корегуючих пристроїв системи ПД.

Відзначимо, що в багатьох практичних ситуаціях оптимальна АЧХ $|K(\omega)|$ відрізняється від постійної.

Приведемо ще один вельми повчальний приклад, який не має відношення до техніки зв'язку. Ще древні греки знали рішення наступної задачі: знайти форму замкнутої кривої $y(x)$ із заданим периметром l і максимально можливою площею $S[y(x)]$.

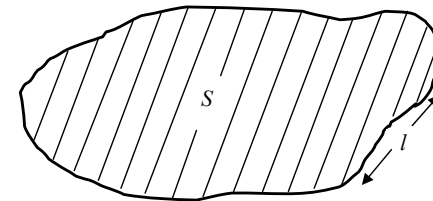


Рисунок 6.2 – Замкнута крива $y(x)$ із заданим периметром l

Ця задача має суттєву відмінність від двох інших, наведених вище. Дослідження функціоналу $S[y(x)]$ на екстремум має бути проведене за додаткової умови – периметр фігури функціонал $l[y(x)]$ зберігає фіксоване значення на всіх порівнюваних кривих $y(x)$.

Покажемо за допомогою простого прикладу, що дана задача не позбавлена сенсу, і, дійсно, різні фігури з однаковим периметром можуть мати різні площі. Порівняємо для цього дві фігури: круг радіуса r і квадрат із стороною a .

Периметри круга $l_{кр}$ і квадрата $l_{кв}$ рівні відповідно

$$l_{кр} = 2\pi r, \quad l_{кв} = 4a.$$

За умовою задачі дані периметри дорівнюють один одному. Тому, прирівнюючи величини $l_{кр}$ і $l_{кв}$, знаходимо вирази для сторони квадрата через радіус круга

$$a = \frac{\pi r}{2}.$$

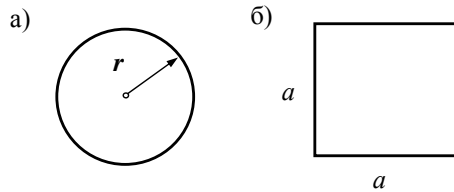


Рисунок 6.3 – Круг радіуса r і квадрат із стороною a

Тепер можна порівняти площі цих фігур, що мають однаковий периметр

$$S_{\text{кр}} = \pi r^2; \quad S_{\text{кв}} = a^2 = \frac{\pi^2 r^2}{4}.$$

Як бачимо, площа круга більше площі квадрата:

$$\frac{S_{\text{кр}}}{S_{\text{кв}}} = \frac{\pi r^2}{\frac{\pi^2 r^2}{4}} = \frac{4}{\pi} > 1.$$

Не обмежуючись цим поодиноким прикладом, можна знайти за допомогою варіаційного числення спільне рішення задачі.

Для кращого засвоєння поняття функціоналу приведемо ще одне еквівалентне визначення: функціонал (його позначатимемо за допомогою букви J) $J = J[y(x)]$ – це закон, який ставить у відповідність функціям дійсні числа. Його можна задати нескінченною послідовністю:

$y(x)$	$J[y(x)]$
$y_\alpha(x)$	a_α
$y_\beta(x)$	a_β
$y_\gamma(x)$	a_γ
.....
$y_\mu(x)$	a_μ
.....

де $\alpha, \beta, \gamma \dots$ – індекси, за допомогою яких позначаються криві з області визначення функціоналу, а також числа з області значень функціоналу, в якій функціонал відображає функції.

Наприклад,

$$a_\alpha = J[y_\alpha(x)].$$

Відзначимо, що не виключена ситуація, коли двом різним функціям з області визначення ставиться у відповідність одне і те ж число (порівняйте другий і третій рядки таблиці). Зворотне, проте, вважається за неможливе: довільно вибраній з області визначення функції функціонал ставить у відповідність лише одне цілком певне (для даного функціоналу) число.

На рис. 6.4 функціонал умовно представлений в звичній для спеціалістів зв'язку формі у виді пристрою з входом і виходом.

Визначення константи c_2 , як вже говорилося, не викликає утруднень, вона рівна $\sqrt{2/T}$. Графік імпульсу з мінімальною ефективною шириною спектра показаний на рис. 9.17.

Відзначимо складність дослідження функціоналу (9.48), в якому $y(t)$ аналізується на всьому відрізку $[0, T]$. Зрозуміло, що рівняння Ейлера і Якобі, а також їх розв'язування мали б той же самий вид, який описаний вище. Але досягти успіху за допомогою пункту б) достатніх умов, очевидно, виявилось б неможливим. Дійсно, розв'язання рівняння Якобі (9.64) в точці $t = T$ порушується у випадку $k = 0$: $u_0(T) = 0$ при $k = 0$. Значить, не існує жодного цілого числа k , при якому пункт б) був би виконаний. І хоча при цьому не порушується необхідна умова Якобі, питання про те, чи реалізується мінімум функціоналу (9.56) на якій-небудь з кривих (9.63) залишається відкритим.

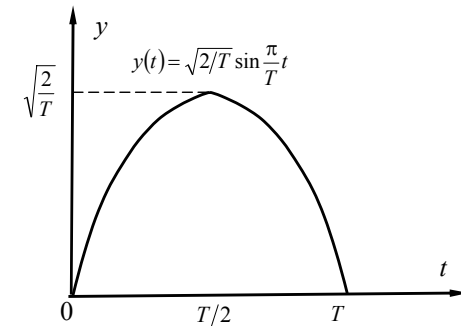


Рисунок 9.17 – Графік імпульсу з мінімальною ефективною шириною спектра

9.8 Лінійна корекція перехідних процесів

Історія питання. Нехай є канал зв'язку із заданою смугою частот Ω . Для спрощення питання розглянемо канал, який є еквівалентом фільтра нижніх частот. Сигнали передаються по каналу у вигляді імпульсної послідовності, так що інформація закладена, наприклад, в амплітудні значення імпульсів.

Відомо, що при цьому імпульси не можна розташовувати занадто близько один до одного: перехідний процес від попереднього накладається на наступний імпульс і стає неможливим провести точний звіт їх амплітудних значень [15, 16].

Очевидний спосіб зменшення перехідного процесу розширенням смуги частот вилучається сформульованою вище умовою: смуга частот каналу Ω наперед задана. Ця умова диктується добре відомою практичною задачею – вторинним ущільненням телефонних каналів інформацією, що передається в цифровій формі (при цьому смуга пропускання каналу складає 3,1 кГц).

Зменшенню (корекції) перехідних процесів присвячена ціла низка робіт.

$$y(0) = c_1 \cos \sqrt{-\lambda} \cdot 0 + c_2 \sin \sqrt{-\lambda} \cdot 0 = c_1 = 0.$$

Очевидно $c_1 = 0$, $c_2 \neq 0$. Для визначення λ скористаємося умовою (9.61) (з урахуванням того, що $c_1 = 0$):

$$y'(T/2) = c_2 \sqrt{-\lambda} \cdot \sin \sqrt{-\lambda} T/2 = 0,$$

звідки

$$\sqrt{-\lambda} = \left(\frac{\pi}{2} + k\pi\right) \frac{2}{T}, \quad k = 0, \pm 1, \pm 2, \dots \quad (9.62)$$

Таким чином,

$$y(t) = c_2 \sin(\pi + 2\pi k) \frac{t}{T}, \quad (9.63)$$

$$k = 0, \pm 1, \pm 2, \dots,$$

де c_2 і ціле число k поки не визначені. Відшукування значень c_2 не представляє труда і може бути легко здійснене за допомогою підстановки (9.63) в умову нормування енергії імпульсу $y(t)$ (9.58).

Дещо складніше знайти число k . Для того, щоб виділити з сімейства екстремалей (9.63) криву, яка дійсно відповідає мінімуму функціоналу (9.57), звернемося до достатніх умов сильного мінімуму, приведених в підрозділі 6.5.

Пункт а) виконаний, бо всі криві сімейства (9.63) – екстремалі.

Для перевірки пункту б) складемо диференційне рівняння Якобі (6.13), яке в даному випадку набирає виду:

$$u'' - \lambda u = 0,$$

тобто збігається за формою з рівнянням Ейлера даної задачі.

Його загальне рішення

$$U(t) = c_1 \cos \sqrt{-\lambda} t + c_2 \sin \sqrt{-\lambda} t,$$

а рішення, що обертається в нуль на лівому кінці

$$U_0(t) = c_2 \sin \sqrt{-\lambda} t, \quad (9.64)$$

Для виконання пункту б) необхідно, щоб функція не оберталася в нуль ні в одній точці відрізка $(0, T/2)$, окрім точки $t = 0$. Легко перевірити, що серед всіх значень $\sqrt{-\lambda}$, що задовольняють (9.62), тільки випадок $k = 0$ задовольняє цій умові. Більш «високочастотні» ($k = \pm 1, \pm 2, \dots$) синусоїди (9.64) мають додаткові нулі на відрізку $(0, T/2)$.

Підставляючи $k = 0$ у (9.63), отримуємо єдину криву, на якій може бути реалізований мінімум (9.57):

$$y(t) = c_2 \sin \frac{\pi}{T} t, \quad 0 \leq t \leq T \quad (9.65)$$

– півхвиллю синуса.

Аби довести, що (9.65) дійсно є рішенням даної задачі, покажемо що виконується і останній (третій) пункт в) достатніх умов.

Дійсно

$$[(y')^2 + \lambda y^2]_{y'y'} = 2 > 0.$$

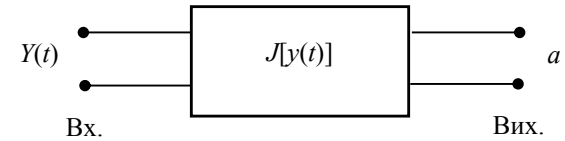


Рисунок 6.4 – Пристрій з входом і виходом

На «вхід» функціоналу надходить сигнал $y(t)$, а на його «виході» з'являється напруга a . Даний рисунок підказує наступне питання: чи є в апаратурі зв'язку пристрої, які з математичної точки зору можна розглядати як функціонали?

Відповідь, як легко виявити, позитивна. Велике число вузлів, блоків апаратури зв'язку є по суті електричними моделями функціоналів. Як приклади досить вказати приймач передання даних (який ставить у відповідність дуже складній кривій, рівній сумі сигналу і завади, число, співпадаюче з одним з можливих значень амплітуди або фази), піковий вольтметр (який ставить у відповідність сигналу довільної форми його максимальне значення), вагметр і т. ін.

6.2 Варіація функціоналу і його властивості

Методи рішення варіаційних задач [1, 3], тобто задач з дослідження функціоналів на максимум або мінімум, дуже схожі з методом дослідження на максимум або мінімум функцій. Тому доцільно нагадати коротко теорію максимуму і мінімуму функцій і паралельно ввести аналогічні поняття та довести схожі теореми для функціоналів.

1. Змінна величина z називається функцією змінної величини x , що позначається так: $z = f(x)$, якщо кожному значенню x з деякої області зміни x відповідає значення z , тобто має місце відповідність: числу x відповідає число z

1. Змінна величина J називається функціоналом, залежним від функції $y(x)$, що позначається так: $J = J[y(x)]$, якщо кожній функції $y(x)$ з деякого класу функцій $y(x)$ відповідає значення J , тобто має місце відповідність: функції $y(x)$ відповідає число J .

2. Приростом Δx аргументу x функції $f(x)$ називається різниця між двома значеннями цієї змінної $\Delta x = x - x_1$. Якщо x – незалежне змінне, то диференціал збігається з приростом $dx = \Delta x$.

2. Приростом або варіацією δy аргументу $y(x)$ функціоналу $J[y(x)]$ називається різниця між двома функціями $\delta y = y(x) - y_1(x)$. При цьому передбачається, що $y(x)$ міняється довільно в деякому класі функцій.

3. Функція $f(x)$ називається безперервною, якщо малій зміні x відповідає мала зміна функції $f(x)$.

3. Функціонал $J[y(x)]$ називається безперервним, якщо малій зміні $y(x)$ відповідає мала зміна функціоналу $J[y(x)]$.

Останнє визначення потребує уточнення і роз'яснення, оскільки зараз же виникає питання: які зміни функції $y(x)$, що є аргументом функціоналу, називаються малими або, що те ж саме, які криві $y = y(x)$ і $y = y_1(x)$ мало між собою відрізняються або є близькими.

Можна вважати, що функції $y(x)$ і $y_1(x)$ близькі в тому випадку, якщо модуль їх різниці $|y(x) - y_1(x)|$ малий для всіх значень x , для яких задаються функції $y(x)$ і $y_1(x)$, тобто вважати за близькі ті криві, що близькі за ординатами. Проте в багатьох випадках природніше вважати близькими тільки ті криві, які близькі за ординатами і за напрямками дотичних у відповідних точках, тобто вимагати, щоб для близьких кривих не лише модуль різниці $|y(x) - y_1(x)|$ був би малий, але крім того був би малий і модуль різниці похідних $|y'(x) - y_1'(x)|$.

На рис. 6.5 показані криві, що близькі в сенсі близькості нульового порядку (це значить, що модуль їх різниці є малим, але відхилення похідних може бути дуже великим), а на рис. 6.6 показані криві, що близькі в сенсі близькості першого порядку. Це значить, що модулі різниці цих функцій і різниці їх похідних малі.

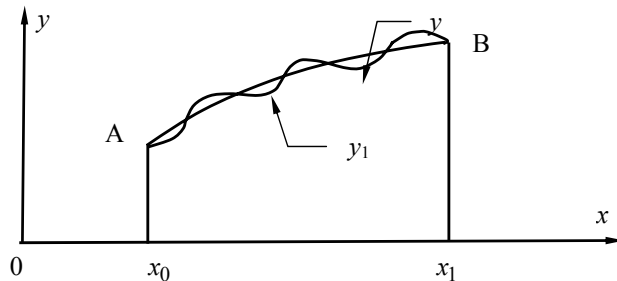


Рисунок 6.5 – Криві, що близькі в розумінні близькості нульового порядку

$$\int_0^{T/2} y^2(t) dt = \frac{1}{2}. \quad (9.58)$$

Правий кінець відрізка $[0, T/2]$ вважатимемо вільним, тобто передбачаємо, що $y(T/2)$ може набувати будь-яких значень. Що стосується лівого краю інтервалу – точки $t = 0$ (так само як і симетричної відносно центра точки $t = T$), то тут точно можна сказати, що

$$y(0) = 0, \quad (9.59)$$

хоча в самій постановці завдання немає ніяких вказівок щодо поведінки $y(t)$ на кінцях. Проте одна важлива обставина приводить до умови (9.59). Для збіжності інтеграла (9.59)

$$\int_{-\infty}^{\infty} \omega^2 |y(\omega)|^2 d\omega \quad (9.60)$$

і існування кінцевої величини σ_y^2 (див. (9.51)) потрібно, щоб функція $|y(\omega)|$ убувала досить швидко при $|\omega| \rightarrow \infty$. Нагадаємо відомий результат: якщо сигнал $y(t)$, $0 \leq t \leq T$ має розриви, його спектр убуває на нескінченність, як $1/\omega$; якщо цей сигнал безперервний, його спектр убуває на нескінченність, як $1/\omega^2$; якщо сигнал безперервний і має безперервну першу похідну, то характер убавання спектра при $|\omega| \rightarrow \infty$ визначається функцією $1/\omega^3$ і т. ін.

У випадку для збіжності інтеграла (9.60) досить зажадати, щоб квадрат модуля спектра $|Y(\omega)|^2$ убував як $1/\omega$ при $|\omega| \rightarrow \infty$ (або $|Y(\omega)|$ убував як $1/\omega^2$). Це означає, що імпульс $y(t)$ має бути безперервним. Але з безперервності функції виходить рівність меж зліва і справа в будь-якій точці області її визначення. Наприклад, на лівому краю області визначення для безперервного сигналу $y(t)$ справедлива рівність

$$y(t-0) = y(t+0), \quad t = 0.$$

Оскільки поза відрізком $0 \leq t \leq T$ функція $y(t)$ вважається рівною нулю, справедливість умови (9.59) очевидна.

Що стосується вільного кінця $t = T/2$, то через теорему про рухливі кінці стосовно функціоналу (9.57), можемо записати відповідні обмеження

$$[(y')^2 + \lambda y^2]' y' |_{t=T/2} = 2y' |_{t=T/2} = 0 \quad (9.61)$$

або

$$y'(T/2) = 0.$$

Рівняння Ейлера для функціоналу (9.49) має вид:

$$y'' - \lambda y = 0,$$

а його рішення, що містить дві довільні постійні

$$y(t) = c_1 \cos \sqrt{-\lambda t} + c_2 \sin \sqrt{-\lambda t}.$$

Скориставшись (9.49), запишемо

(він дорівнює одиниці). Надалі, проте, зручніше використовувати еквівалентну (9.47) рівність (9.50).

Нагадаємо, що дисперсія характеризує ступінь зосередженості щільності $p(\omega)$ навколо її середнього значення. Чим менше дисперсія, тим більш «вузким» є графік функції $p(\omega)$. В принципі, ця функція в межі при $\sigma_y \rightarrow 0$ переходить в σ -функцію (для сигналів $y(t)$ кінцевої тривалості останнє неможливо). Ця обставина і обґрунтовує застосування теоретико-імовірнісного критерію – дисперсії до оцінки ширини смуги частот, займаної сигналом $y(t)$.

Вираз перетворимо так, щоб представити його як функціонал $y(t)$. Для цього проведемо наступні допоміжні міркування, що відносяться до виразу зворотного перетворення Фур'є

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) e^{i\omega t} d\omega. \quad (9.52)$$

Продиференціюємо обидві частини рівності (9.52) по t

$$y'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} i\omega Y(\omega) e^{i\omega t} d\omega. \quad (9.53)$$

Застосуємо тепер теорему Парсеваля-Релея (9.49) до сигналу $y(t)$, $0 \leq t \leq T$ та з урахуванням (9.53) отримаємо:

$$\int_0^T [y'(t)]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\omega Y(\omega)|^2 d\omega. \quad (9.54)$$

Порівнюючи рівняння, можемо записати

$$\delta_y^2 = \delta_y^2[y(t)] = \int_0^T [y'(t)]^2 dt. \quad (9.55)$$

Для мінімізації функціоналу (9.55) при обмеженні (9.50) складемо допоміжний функціонал

$$I[y(t)] = \int_0^T [(y')^2 + \lambda y^2] dt \quad (9.56)$$

Зробимо просте припущення (воно полегшить перевірку достатніх умов мінімізації), яке є природним з фізичної точки зору: імпульс $y(t)$ повинен мати парну симетрію відносно середини відрізка $[0, T]$ – точки $t = T/2$.

Тоді задачу мінімізації функціоналу (9.56) можна замінити задачею мінімізації функціоналу

$$I_1[y(t)] = \int_0^{T/2} [(y')^2 + \lambda y^2] dt \quad (9.57)$$

за умови

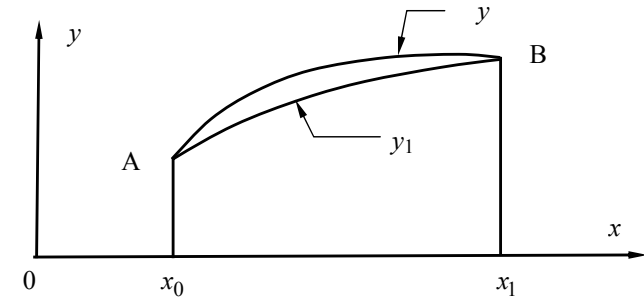


Рисунок 6.6 – Криві, що близькі в розумінні близькості першого порядку

Уточнимо поняття безперервності функціоналу, сформулювавши паралельно точніше визначення безперервності функцій.

4. Функція $y(x)$ безперервна при $x = x_0$, якщо для будь-якого позитивного ε можна підібрати $\delta > 0$ таке, що $|f(x) - f(x_0)| < \varepsilon$ при $|x - x_0| < \delta$. При цьому мається на увазі, що x набуває значень, в яких функція $f(x)$ визначена.

5. Лінійною функцією називається функція $l(x)$, яка задовольняє наступним умовам: $l(cx) = cl(x)$, де c – довільна постійна і $l(x_1 + x_2) = l(x_1) + l(x_2)$.

Лінійна функція однієї змінної має вид $l(x) = kx$, де k – постійна.

6. Якщо приріст функції $\Delta f = f(x + \Delta x) - f(x)$ може бути представлено у виді $\Delta f = A(x)\Delta x + \beta(x, \Delta x)\Delta x$, де $A(x)$ не залежить від Δx , а $\beta(x, \Delta x) \rightarrow 0$ при $\Delta x \rightarrow 0$, то функція називається

4. Функціонал $J[y(x)]$ безперервний при $y = y_0(x)$ у сенсі близькості:
а) 0-го порядку, якщо для будь-якого позитивного ε можна підібрати $\delta > 0$ таке, що $|J[y(x)] - J[y_0(x)]| < \varepsilon$ при $|y(x) - y_0(x)| < \delta$;
б) 1-го порядку, якщо для будь-якого позитивного ε можна підібрати $\delta > 0$ таке, що $|J[y(x)] - J[y_0(x)]| < \varepsilon$ при $|y(x) - y_0(x)| < \delta$, $|y'(x) - y_0'(x)| < \delta$.

5. Лінійним функціоналом називають функціонал $L[y(x)]$, який задовольняє наступним умовам: $L[cy(x)] = cL[y(x)]$, де c – постійна (довільна) і $L[y_1(x) + y_2(x)] = L[y_1(x)] + L[y_2(x)]$.

Прикладом лінійного функціоналу є $L[y(x)] = \int_{x_0}^{x_1} [p(x)y(x) + q(x)y'(x)] dx$.

6. Якщо приріст функціоналу $\Delta J = J[y(x) + \delta y] - J[y(x)]$ можна представити у виді $\Delta J = L[y(x), \delta y] + \beta[y(x), \delta y] \max|\delta y|$, де $L[y(x), \delta y]$ – лінійний по відношенню

диференційованою, а лінійна по відношенню до Δx частина приросту $A(x)\Delta x$ називається диференціалом функції і позначається df . Розділивши на Δx і переходячи до межі при $\Delta x \rightarrow 0$ отримаємо, що $A(x) = f'(x)$ і, отже, $df = f'(x)\Delta x$.

ню до δy функціонал, $\max|\delta y|$ – максимальне значення $|\delta y|$ і $\beta[y(x), \delta y] \rightarrow 0$ при $\max|\delta y| \rightarrow 0$, то лінійна по відношенню до δy частина приросту функціоналу, тобто $L[y(x), \delta y]$ називається варіацією функціоналу і позначається δJ .

Отже, варіація функціоналу – це головна лінійна по відношенню до δy частина приросту функціоналу.

При дослідженні функціоналів варіація грає таку ж роль, яку грає диференціал при дослідженні функцій.

Можна дати і інше, майже еквівалентне, визначення диференціала функції і варіації функціоналу. Розглянемо значення функції $f(x + \alpha\Delta x)$ при фіксованому α . При $\alpha = 1$ отримаємо прирощене значення функції $f(x + \alpha\Delta x)$, а при $\alpha = 0$ – початкове значення функції $f(x)$. Неважко перевірити, що похідна від $f(x + \alpha\Delta x)$ по α при $\alpha = 0$ дорівнює диференціалу функції $f(x)$ в точці x . Дійсно, за правилом диференціювання складної функції

$$\frac{\partial}{\partial \alpha} f(x + \alpha\Delta x) = f'(x + \alpha\Delta x) /_{\alpha=0} = f'(x)\Delta x = \alpha f'(x).$$

Доведено, що для функціоналів можна визначити варіацію, як похідну функціоналу $J[y(x) + \alpha\delta y]$ по α при $\alpha = 0$.

7. Диференціал функції $f(x)$ дорівнює $\frac{\partial}{\partial x} f(x + \alpha\Delta x) /_{\alpha=0}$	7. Варіація функціоналу $J[y(x)]$ дорівнює $\frac{\partial}{\partial \alpha} J[y(x) + \alpha\delta y]_{\alpha=0}$
---	---

Визначення. Функціонал $J[y(x)]$ досягає на кривій $y = y_0(x)$ максимуму, якщо його значення на будь-якій близькій до $y = y_0(x)$ кривої не більше, ніж $J[y_0(x)]$, тобто $\Delta J = J[y(x)] - J[y_0(x)] \leq 0$.

Якщо $\Delta J \leq 0$, причому $\Delta J = 0$ тільки при $y(x) = y_0(x)$ досягається строгий максимум. Аналогічно визначається крива $y = y_0(x)$, на якій реалізується мінімум. В цьому випадку $\Delta J \geq 0$ для всіх кривих, близьких до кривої $y = y_0(x)$.

8. Теорема. Якщо функція $f(x)$, що диференціюється, досягає максимуму або мінімуму у внутрішній області визначення $x = x_0$, то в цій точці $df = 0$.	8. Якщо функціонал $J[y(x)]$, що має варіацію, досягає максимуму або мінімуму при $y = y_0(x)$, то при $y = y_0(x)$ $\delta J = 0$.
--	--

$$\int_{-\infty}^{\infty} y^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega. \quad (9.49)$$

У нашому випадку це рівняння може бути переписане так:

$$\int_0^T y^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega.$$

Тому умова (9.47) еквівалентна наступному:

$$\int_0^T y^2(t) dt = 1. \quad (9.50)$$

Варіаційну задачу максимізації (9.48) за умови (9.50) зводять до вирішення інтегрального рівняння відносно невідомої функції $y(t)$. Викладення досягнутих тут цікавих і важливих результатів вимагає, проте, використання досить складного математичного апарата. Тому використовуємо інший підхід до мінімізації позасмугових випромінювань, для чого введемо поняття про ефективну ширину спектра.

Щоб ввести це поняття, яке аналогічне дисперсії розподілу, спробуємо перенести характеристики закону розподілу імовірності випадкових величин на спектр сигналу. Припускаємо, що виконується умова (9.47), і розглядатимемо невід'ємну функцію

$$\frac{1}{2\pi} |Y(\omega)|^2, \quad -\infty < \omega < \infty,$$

як щільність розподілу імовірності $p(\omega)$ деякої випадкової величини.

Оскільки модуль спектра довільного речового сигналу є парною функцією частоти, тобто

$$|Y(-\omega)| = |Y(\omega)|,$$

то середнє значення цієї випадкової величини дорівнює нулю:

$$\begin{aligned} \int_{-\infty}^{\infty} \omega p(\omega) d\omega &= \int_{-\infty}^{\infty} \omega \frac{1}{2\pi} |Y(\omega)|^2 d\omega = \frac{1}{2\pi} \int_0^{\infty} \omega |Y(\omega)|^2 d\omega + \int_{-\infty}^0 \omega |Y(\omega)|^2 d\omega = \\ &= \frac{1}{2\pi} \left[\int_0^{\infty} \omega |Y(\omega)|^2 d\omega - \int_0^{\infty} \omega |Y(\omega)|^2 d\omega \right] = 0, \end{aligned}$$

а її дисперсія

$$\sigma_y^2 = \int_{-\infty}^{\infty} \omega^2 p(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |Y(\omega)|^2 d\omega. \quad (9.51)$$

Позитивну величину σ_y назвемо ефективною шириною спектра сигналу $y(t)$, $0 \leq t \leq T$ і поставимо питання про мінімізацію σ_y або, що еквівалентно, мінімізацію σ_y^2 . При цьому за додаткову умову приймемо рівність (9.47), яка відображає відому властивість інтеграла від щільності розподілу імовірності

протяжності і, отже, принципово неусувними позасмуговими випромінюваннями. Спектр прямокутного імпульсу $y(t)=1, 0 \leq t \leq T$ є достатньою мірою типовим (рис. 9.16).

Іншими словами, не існує частотного діапазону $(-\Omega, \Omega), \Omega > 0$, всередині якого помістився б цілком спектр прямокутного (та і будь-якого іншого) імпульсу. Разом з тим, ясно, що позасмугові випромінювання можуть, залежно від форми імпульсу, мати більшу або меншу інтенсивність.

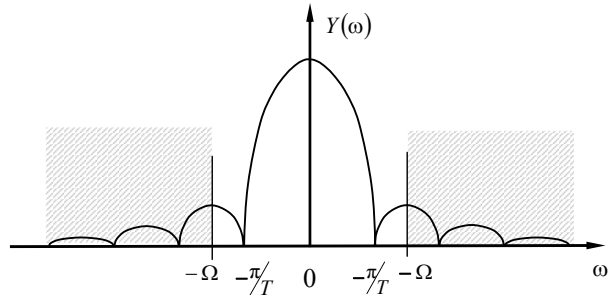


Рисунок 9.16 – Спектр прямокутного імпульсу

Існують різні способи оцінки позасмугових випромінювань. Мабуть, найбільш поширений з них – енергетичний, при якому інтенсивність позасмугових випромінювань характеризують величиною μ .

У випадку низькочастотного робочого діапазону частот критерій ефективності запишемо у виді:

$$\mu = \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega - \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |Y(\omega)|^2 d\omega}{\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega}. \quad (9.45)$$

Задачі мінімізації величини μ присвячена безліч літератури. Відзначимо, що для мінімізації відношення (9.45) переходять зазвичай до іншої, еквівалентної задачі. Вважаючи

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega = 1, \quad (9.47)$$

можна вирішити питання про максимізацію енергії імпульсу $y(t)$ у робочій смузі частот $(-\Omega, \Omega)$ –

$$\frac{1}{2\pi} \int_{-\Omega}^{\Omega} |Y(\omega)|^2 d\omega. \quad (9.48)$$

Нагадаємо, що за теоремою Парсеваля-Релея [16] справедливе наступне рівняння для енергії сигналу

Приведемо доказ теореми для функціоналів.

При фіксованих $y_0(x)$ і δy , $J[y(x) + \alpha \delta y] = \varphi(\alpha)$ є функцією від α , яка при $\alpha = 0$ за припущенням, досягає максимуму або мінімуму; отже похідна $\varphi'(0) = 0$ або $\frac{\partial}{\partial \alpha} J[y(x) + \alpha \delta y]_{\alpha=0}$ тобто $\delta J = 0$. Отже, на кривих, на яких досягається екстремум функціоналу, його варіація дорівнює нулю.

Уточнимо поняття екстремуму функціоналу. Якщо функціонал $J[y(x)]$ досягає на кривій $y = y_0(x)$ максимуму або мінімуму по відношенню до всіх кривих, для яких модуль різниці $|y(x) - y_0(x)|$ малий, тобто по відношенню до всіх кривих, близьких до $y = y_0(x)$ у розумінні близькості нульового порядку, то максимум або мінімум називається сильним.

Якщо ж функціонал $J[y(x)]$ досягає на кривій $y = y_0(x)$ максимуму або мінімуму лише по відношенню до кривих $y = y_0(x)$, близьких до $y = y_0(x)$ у розумінні близькості першого порядку, тобто по відношенню до кривих близьких до $y = y_0(x)$ не лише за ординатами, але і за напрямками дотичних, то максимум або мінімум називається слабким.

Очевидно, якщо кривою $y_0(x)$ не досягається сильний максимум (або мінімум), то напевно досягається і слабкий. Зворотне твердження в загальному випадку не справедливе.

Всі визначення даного параграфу і основна теорема майже без жодної зміни переносяться на функціонали, залежні від декількох функцій

$$J[y_1(x), y_2(x), \dots, y_n(x)],$$

кожна з яких також може залежати від декількох змінних:

$$J[y_1(x_1, x_2, \dots, x_m), y_2(x_1, x_2, \dots, x_m), \dots, y_n(x_1, x_2, \dots, x_m)].$$

6.3 Рівняння Ейлера

Досліджуємо на екстремум простіший [4, 5, 7], і разом з тим досить поширений в додатках, функціонал

$$J[y(x)] = \int_{x_0}^{x_1} F(x, y(x), y'(x)) dx, \quad (6.6)$$

причому граничні точки допустимих кривих закріплені: $y(x_0) = y_0, y(x_1) = y_1$ (рис. 6.7), функцію $F(x, y, y')$ вважатимемо за точку, що тричі диференціюється.

Як було визначено, необхідною умовою екстремуму є перетворення на нуль варіації функціоналу. Розглянемо тепер, як «працює» ця основна теорема у випадку даного функціоналу. Припустимо, що екстремум досягається на кривій, що двічі диференціюється $y = y(x)$ (потрібно лише існування похідних першого порядку у допустимих кривих, тоді можна іншим методом довести, що у кривій, на якій реалізується екстремум, існує і друга похідна).

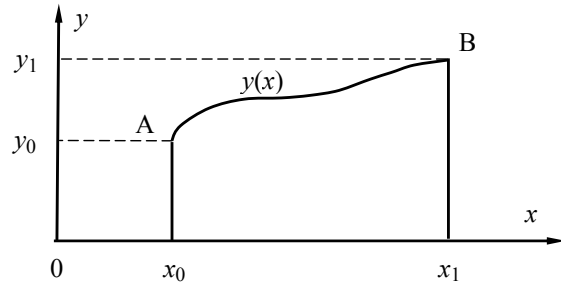


Рисунок 6.7 – Крива $y = y(x)$

Візьмемо яку-небудь близьку до $y = y(x)$ допустиму криву $y = y^*(x)$ та включимо криві $y = y(x)$ і $y = y^*(x)$ в одне параметричне сімейство кривих

$$y(x, \alpha) = y(x) + \alpha(y^*(x) - y(x)).$$

При $\alpha = 0$ отримаємо криву $y = y(x)$, при $\alpha = 1$ маємо $y = y^*(x)$ (рис.6.8).

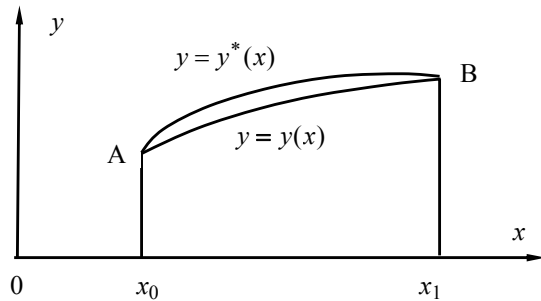


Рисунок 6.8 – Крива $y = y(x)$

Як ми вже знаємо, різниця $y^*(x) - y(x)$ називається варіацією функції $y(x)$ і позначається δy .

Варіація δy у варіаційних задачах грає роль, аналогічну ролі приросту незалежного змінного Δx в задачах на дослідження екстремумів функції $f(x)$.

Варіація функції $\delta y = y^*(x) - y(x)$ є функцією x . Цю функцію можна диференціювати один або кілька разів, при чому

$$(\delta y)' = y^{*'}(x) - y'(x) = \delta y',$$

тобто похідна варіації дорівнює варіації похідної і аналогічно

$$(\delta y)' = y^{*''}(x) - y''(x) = \delta y''.$$

$$\int_{-\infty}^{\infty} \log_2 \frac{1}{p(x)} dx \quad (9.43)$$

максимальна.

Допоміжний функціонал (з урахуванням обмежень (9.41) і (9.42) має вид:

$$\hat{J}[p(x)] = \int_{-\infty}^{\infty} [p \log_2 \frac{1}{p} + \lambda_1 p + \lambda_2 x^2 p] dx,$$

а його рівняння Ейлера

$$-\frac{1}{\ln 2} \cdot p \cdot \frac{1}{p} - \log_2 p + \lambda_1 + \lambda_2 x^2 = 0,$$

звідки

$$p(x) = 2^{\lambda_1 + \lambda_2 x^2 - \frac{1}{\ln 2}} = e^{\ln 2 [\lambda_1 + \lambda_2 x^2 - \frac{1}{\ln 2}]}$$

Опустимо громіздкі (але елементарні) розрахунки, за допомогою яких з урахуванням (9.33) і (9.34), отримуємо константи λ_1 і λ_2 , а остаточний вираз для шуканої щільності

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{x^2}{2\sigma_0^2}}. \quad (9.44)$$

Отже розподіл імовірностей за Гаусом при заданій дисперсії має максимальну ентропію.

9.7 Задача пошуку імпульсу з мінімально ефективною шириною спектра

Як правило, передача інформації по каналах зв'язку здійснюється в строго обмеженому частотному діапазоні, поза яким «позасмугові випромінювання» не повинні перевищувати деякої заданої існуючими нормами величини. При передачі даних займана смуга частот переважно визначається формою сигналу-переносника. Тому представляє суттєвий інтерес відшукування форми сигналів кінцевої тривалості з мінімально можливою смугою частот.

Сказане, проте, потребує деякого роз'яснення. Позначимо сигнал-переносник тривалості T через $y(t)$, $0 \leq t \leq T$.

Тоді його спектр кінцевої тривалості [16]

$$Y(\omega) = \int_0^T y(t) e^{-i\omega t} dt, \quad -\infty < \omega < \infty \quad (9.45)$$

визначається перетворенням Фур'є (9.45), де $\omega = \alpha + i\beta$ – комплексна змінна.

Відомо, що цілі функції можуть обернутися в нуль лише в ізолюваних точках, тобто множини точок – «міра більше нуля». Прикладом такої множини може бути відрізок дійсної або уявної осі комплексної площини, коло або сукупність фігур на цій площині, дійсна піввісь і т. ін. Практично це означає, що спектри сигналів кінцевої тривалості являються спектрами нескінченної

за умовою нормування щільності імовірностей

$$\int_{-x_0}^{x_0} p(x) dx = 1. \quad (9.41)$$

Метою є знаходження максимального значення функціоналу (9.40) за обставини (9.41). Приступимо до рішення цієї задачі, залишивши без уваги той факт, що щільність розподілу імовірності є невід'ємна функція. Якщо знайдене рішення рівняння Ейлера задовольнятиме умові $p(x) \geq 0$, то його можна буде розглядати як рішення задачі. Якщо цього не трапиться, мають бути прийняті спеціальні заходи, які розглядалися у 6-му розділі.

Допоміжний функціонал має вид:

$$\hat{J}[p(x)] = \int_{-x_0}^{x_0} [p \log_2 \frac{1}{p} + \lambda p] dx,$$

а його рівняння Ейлера –

$$\frac{1}{\ln 2} p \cdot \frac{1}{p} - \log_2 p + \lambda = 0,$$

звідки

$$\log_2 p = -\frac{1}{\ln 2} + \lambda, \\ p(x) = 2^{\lambda - 1/\ln 2} = \text{const}.$$

З (9.41) знаходимо

$$\lambda = -\log_2 x_0 + \frac{1}{\ln 2} - 1,$$

і значить, шукана щільність розподілу має вид:

$$p(x) = \frac{1}{2x_0}, \quad -x_0 \leq x \leq x_0.$$

Отже, в даній задачі ентропія, у разі рівномірного розподілу, є максимальною. Чисельна величина ентропії

$$H_{\max} = \int_{-x_0}^{x_0} p \log_2 \frac{1}{p} dx = \int_{-x_0}^{x_0} \frac{1}{2x_0} \log_2 2x_0 dx = \log_2 2x_0.$$

Задача 2. У даній задачі є обмеження на дисперсію розподілу імовірності інформаційного параметра (пікове значення його не передбачається обмеженням):

$$\int_{-x_0}^{x_0} x^2 p(x) dx = \sigma^2, \quad (9.42)$$

де σ_0^2 – задана (позитивна) константа.

Знайдемо щільність $p(x)$, при якій ентропія

Отже, розглянемо сімейство $y = y(x, \alpha)$ або $y = y(x) + \alpha \delta y$, що містить при $\alpha = 0$ криву, на якій досягається екстремум, а при $\alpha = 1$ – деяку близьку допустиму криву – так звану криву порівняння, і покажемо як «працює» теорема п.7 попереднього підрозділу у випадку функціоналу (6.1).

Якщо розглядати значення функціоналу

$$J[y(x)] = \int_{x_0}^{x_1} F(x, y, y') dx$$

лише на кривих $y = y(x, \alpha)$, то функціонал перетворюється на функцію

$$J[y(x, \alpha)] = \varphi(\alpha),$$

оскільки значення параметра α визначає криву $y = y(x, \alpha)$ і тим самим визначає і значення функціоналу $J[y(x)]$. Ця функція досягає свого екстремуму при $\alpha = 0$ тому, що при $\alpha = 0$ одержуємо $y = y(x)$ і функціонал, за припущенням, досягає екстремуму в порівнянні з будь-якою близькою допустимою кривою і, зокрема, по відношенню до близьких кривих $y = y(x, \alpha)$. Необхідною умовою екстремуму функції $\varphi(\alpha)$ при $\alpha = 0$, як відомо з курсу аналізу, є перетворення на нуль її похідної при $\alpha = 0$

$$\varphi'(0) = 0.$$

Оскільки

$$\varphi(\alpha) = \int_{x_0}^{x_1} F(x, y(x_1, \alpha), y'(x_1, \alpha)) dx,$$

тоді

$$\varphi'(\alpha) = \int_{x_0}^{x_1} [F_y \frac{\partial}{\partial \alpha} y(x_1, \alpha) + F_{y'} \frac{\partial}{\partial \alpha} y'(x_1, \alpha)] dx,$$

де

$$F_y = \frac{\partial}{\partial y} F(x_1, y(x_1, \alpha), y'(x_1, \alpha)),$$

$$F_{y'} = \frac{\partial}{\partial y'} F(x, y(x_1, \alpha), y'(x_1, \alpha)).$$

Внаслідок того, що

$$\frac{\partial}{\partial \alpha} y(x_1, \alpha) = \frac{\partial}{\partial \alpha} [y(x) + \alpha \delta y] = \delta y$$

і

$$\frac{\partial}{\partial \alpha} y'(x_1, \alpha) = \frac{\partial}{\partial \alpha} [y'(x) + \alpha \delta y'] = \delta y'$$

мають місце рівності

$$\varphi'(\alpha) = \int_{x_0}^{x_1} [F_y(x_1, y(x_1, \alpha), y'(x_1, \alpha)) \delta y + F_{y'}(x_1, y(x_1, \alpha), y'(x_1, \alpha)) \delta y'] dx;$$

$$\phi'(\alpha) = \int_{x_0}^{x_1} [F_y(x_1, y(x), y'(x))\delta y + F_{y'}(x_1, y(x), y'(x))\delta y'] dx. \quad (9.38)$$

Як ми вже знаємо, $\phi'(0)$ називається варіацією функціоналу і позначається δJ . Необхідна умова екстремуму функціоналу J полягає в перетворенні на нуль його варіації $\delta J = 0$.

Для функціоналу

$$J[y(x)] = \int_{x_0}^{x_1} F(x, y, y') dx$$

ця умова має вид

$$\int_{x_0}^{x_1} [F_y \delta y + F_{y'} \delta y'] dx = 0.$$

Інтегруємо другий доданок по частинам, і, з урахування того, що $\delta y' = (\delta y)'$, отримаємо

$$\int_{x_0}^{x_1} F_{y'} \delta y' dx = \int_{x_0}^{x_1} F_{y'} d(\delta y) = F_{y'} \delta y \Big|_{x_0}^{x_1} - \int_{x_0}^{x_1} \delta y \frac{d}{dx} F_{y'} dx;$$

$$\delta J = F_{y'} \delta y \Big|_{x_0}^{x_1} + \int_{x_0}^{x_1} (F_y - \frac{d}{dx} F_{y'}) \delta y dx.$$

Але

$$\frac{\delta y}{x=x_0} = y(x_0) - y(x_0) = 0 \quad \text{і} \quad \frac{\delta y}{x=x_1} = y^*(x_1) - y(x_1) = 0,$$

тому, що всі допустимі криві в даній простій задачі проходять через фіксовані граничні точки, і, отже

$$\delta J = \int_{x_0}^{x_1} \left(F_y - \frac{d}{dx} F_{y'} \right) \delta y dx.$$

Тоді необхідна умова екстремуму має вид

$$\int_{x_0}^{x_1} \left(F_y - \frac{d}{dx} F_{y'} \right) \delta y dx = 0, \quad (6.7)$$

при чому перший множник $\left(F_y - \frac{d}{dx} F_{y'} \right)$ на кривій $y = y(x)$, що реалізує екстремум, є заданою безперервною функцією, а другий множник δy , зважаючи на вільність у виборі кривої порівняння $y = y^*(x)$, є довільною функцією, що задовольняє лише деяким, вельми спільним умовам, а саме: функція δy у граничних точках $x = x_0$ і $x = x_1$ обертається в нуль, безперервна і та, що диференціюється один або декілька разів, δy або $\delta y'$ малі за абсолютною величиною.

$$S(f) \geq 0. \quad (9.38)$$

На рішеннях подібних задач зупинятися не будемо. Вкажемо також на один простий захід, який дозволяє звести дане завдання до завдання без обмежень виду (9.38). Для цього досить ввести замість шуканої функції $S(f)$ нову функцію $r(f)$, зв'язану з нею співвідношенням

$$S(f) = r^2(f),$$

і мінімізувати функціонал варіюючи $r(f)$. Але така заміна може інколи привести до істотного ускладнення рівняння Ейлера.

9.6 Екстремальні задачі, що пов'язані з поняттям ентропії

Для характеристики ансамблю випадкових сигналів (які породжені, наприклад, деяким джерелом повідомлень) прийнято використовувати математичне очікування кількості інформації, що називається ентропією [15, 16]

$$H = M \left\{ \log_2 \frac{1}{p(x)} \right\},$$

де $p(x)$ – щільність розподілу імовірності інформаційного параметра x , при цьому $\log_2 \frac{1}{p(x)} dx$ рівний, за визначенням, кількості інформації в сигналі з інформаційним параметром, що набуває значення в інтервалі $(x, x + dx)$.

У наведеному виразі для ентропії математичне очікування означає усереднювання по всьому ансамблю сигналів, тобто

$$H = \int p(x) \log_2 \frac{1}{p(x)} dx. \quad (9.39)$$

Нагадаємо, що чим більше ентропія джерела, тим більше ступінь несподіванки переданих ним повідомлень в середньому, тобто тим більш невизначеним є очікуване повідомлення. Тому ентропію часто називають мірою невизначеності повідомлень. Таким чином, до передавання повідомлення існує деяка його невизначеність.

Після прийому повідомлення (якщо воно свідомо приймається вірно) будь-яка невизначеність усувається. Це дозволяє трактувати кількість інформації як міру усунення невизначеності. Можна характеризувати ентропію як міру різноманітності повідомлень від джерела.

Поставимо питання про знаходження щільності розподілу імовірностей $p(x)$, що мають за заданими умовами максимальну ентропію.

Задача 1. Розглянемо випадок, коли в ансамблі обмежені пікові значення інформаційного параметра

$$H = H[p(x)] = \int_{-x_0}^{x_0} p(x) \log_2 \frac{1}{p(x)} dx, \quad (9.40)$$

$$\frac{\partial}{\partial S} \{ \log_2 \left[\frac{S+n}{n} \right] + \lambda S \} = \frac{1}{\ln 2} \cdot \frac{n}{S+n} \cdot \frac{1}{n} + \lambda = 0,$$

звідки

$$S = -\frac{1}{\lambda \ln 2} - n. \quad (9.35)$$

Підставляючи (9.35) в (9.36) і враховуючи позначення (9.31), знаходимо значення λ :

$$\int_{f_1}^{f_2} S(f) df = \int_{f_1}^{f_2} \left[-\frac{1}{\lambda \ln 2} - n(f) \right] df = -\frac{f_2 - f_1}{\lambda \ln 2} - P_{\Phi} = \tilde{P}_{\Phi};$$

$$\lambda = -\frac{f_2 - f_1}{\ln 2 (P_c + P_{\text{ш}})}.$$

Остаточний вираз для оптимальної форми спектра щільності потужності сигналу має вид:

$$S(f) = -\frac{\tilde{P}_c + P_{\text{ш}}}{f_2 - f_1} - n(f). \quad (9.36)$$

Як видно, оптимальний спектр щільності потужності сигналу доповнює спектр щільності потужності шуму до константи. Іншими словами, енергію передавача доцільно розподіляти в робочому діапазоні частот нерівномірно, направляючи її, в основному, в ті ділянки, де потужність шуму мала.

Цей висновок має безперечний практичний інтерес, проте він, мабуть, зроблений поспішно, адже не доведено, що на екстремалі (9.36) дійсно досягається мінімум. Втім, із зауваження про функціонали, що не містять похідної невідомої функції витікає обґрунтування того факту, що на функції (9.35) дійсно реалізується екстремум функціонала (9.34), а разом з ним і функціоналу (9.32) за умови (9.33). Цей екстремум може бути тільки максимумом, бо, наближаючи $S(f)$ в довільно малому, але кінцевому підінтервалі інтервалу (f_1, f_2) до функції $n(f)$, узятій із зворотним знаком $S(f) \approx n(f)$, можна зробити значення функціоналу (9.31) меншим будь-якого наперед заданого числа.

У зв'язку із записом наближеної рівності $S(f) \approx n(f)$ доцільно нагадати, що за фізичним трактуванням функції $S(f)$ і $n(f)$ невід'ємні. Вирішуючи поставлену задачу формально, ніде не була урахована умова $S(f) \geq 0$. Тому формула (9.36) дійсно дає рішення поставленої задачі з урахуванням фізичних обмежень, якщо в усіх точках інтервалу (f_1, f_2) виконується нерівність

$$\frac{\tilde{P}_c + P_{\text{ш}}}{f_2 - f_1} - n(f) \geq 0. \quad (9.37)$$

Проте нерівність (9.37) може виявитися порушеною і ця обставина сигналізує про те, що математична задача максимізації пропускну здатності каналу $R[S(f)]$ була поставлена некоректно, і щоб виправити положення, треба до умови (9.33) додати умову

Для управління отриманої вище необхідної умови екстремуму скористаємося наступною лемою.

Основна лема варіаційного числення. Якщо функція $\Phi(x)$ безперервна на відрізку $[x_0, x_1]$ і $\int_{x_0}^{x_1} \Phi(x) \eta(x) dx = 0$ при довільному виборі функції $\eta(x)$, яка

задовольняє лише деяким спільним умовам, наприклад, диференціюється один або кілька разів, перетворюється на нуль на кінцях відрізка $[x_0, x_1]$, або $|\eta(x)| < \varepsilon$ і $|\eta'(x)| < \varepsilon$, то $\Phi(x) \equiv 0$ на відрізку $x_0 \leq x \leq x_1$.

Застосуємо тепер основну лему для спрощення отриманої вище необхідної умови екстремуму (6.7) функціоналу (6.6). Всі умови леми виконані: на кривій, що реалізує екстремум, множник $(F_y - \frac{d}{dx} F_{y'})$ є безперервною функцією, а варіація δy є довільною функцією, на яку накладені лише передбачені в основній лемі обмеження спільного характеру, отже $F_y - \frac{d}{dx} F_{y'} \equiv 0$ на кривій $y = y(x)$, що реалізує екстремум даного функціоналу, тобто є рішенням диференційного рівняння другого порядку

$$F_y - \frac{d}{dx} F_{y'} = 0, \quad (6.8)$$

або в розгорненому виді

$$F_y - F_{xy'} - F_{yy'} y' - F_{y'y'} y'' = 0. \quad (6.9)$$

Це рівняння носить ім'я Ейлера, який вперше отримав його в 1744 році. Інтегральні криві рівняння Ейлера $y = (x, C_1, C_2)$ називаються екстремаліями.

Тільки на екстремаліях може досягатися екстремум функціоналу (6.6), притому на таких, які задовольняють згаданим вище умовам на кінцях інтервалу (ці умови дозволяють визначити константи C_1 і C_2).

Однак для того, щоб встановити, чи реалізується насправді на екстремаліях екстремум, і при тому максимум або мінімум, треба скористатися достатніми умовами екстремуму, про які буде сказано далі (підрозділ 6.4).

У ряді випадків, проте, вже одне рівняння Ейлера дає вичерпне рішення задачі. Дійсно, саме існування екстремуму буває очевидним з фізичного або геометричного сенсу задачі (наприклад, задачі про брахістохрону – найкоротшу відстань між двома точками і т. ін.). Якщо при цьому існує лише єдина екстремаль, що задовольняє граничним умовам задачі, то саме вона і буде неодмінно тією кривою, яка реалізує шуканий екстремум.

Як бачимо з формули (6.9), рівняння Ейлера є диференціальним рівнянням другого порядку, взагалі кажучи, нелінійним.

Розглянемо деякі приклади.

Приклад 1. На яких кривих може досягати екстремуму функціонал

$$I[y(x)] = \int_0^{\pi/2} [(y')^2 - y^2] dx; \quad y(0) = 0; \quad y(\pi/2) = 1?$$

Легко перевіряється, що

$$F_y = -2y; F_{y'} = 2y'; \frac{d}{dx} F_{y'} = 2y''.$$

Рівняння Ейлера має вид

$$y'' + y = 0. \quad (6.10)$$

Це лінійне диференціальне рівняння другого порядку. Воно зустрічалося в курсі теоретичної електротехніки, цьому рівнянню задовольняє напруга вільних коливань в коливальному контурі другого порядку без втрат (рис. 6.9).

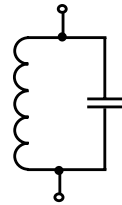


Рисунок 6.9 – Коливальний контур другого порядку без втрат

Загальний інтеграл диференційного рівняння (6.10) має вид

$$y(x, C_1, C_2) = C_1 \cos x + C_2 \sin x. \quad (6.11)$$

Використовуючи умови на краях $x=0$ і $x=\pi/2$ та (6.11), знаходимо

систему рівнянь для знаходження C_1 і C_2

$$\left. \begin{aligned} y(0) &= C_1 \cos 0 + C_2 \sin 0 = 0, \\ y(\pi/2) &= C_1 \cos \pi/2 + C_2 \sin \pi/2 = 1, \end{aligned} \right\}$$

звідки $C_1 = 0$, $C_2 = 1$. Отже, екстремум може досягатися на кривій $y = \sin x$.

Приклад 2. На яких кривих може досягати екстремуму функціонал

$$J[y(x)] = \int_0^1 [(y')^2 + 12xy] dx; \quad y(0) = 0; \quad y(1) = 1 \quad ?$$

Знаходимо:

$$F_y = 12x; \quad F_{y'} = 2y'; \quad \frac{d}{dx} F_{y'} = 2y''.$$

Рівняння Ейлера має вид

$$y'' - 6x = 0.$$

Знайдемо загальний інтеграл цього рівняння, яке перепишемо у виді

$$\frac{d^2 y}{dx^2} = 6x,$$

звідки

$$\frac{dy}{dx} = \int 6x dx + C_1 = 3x^2 + C_1,$$

і, отже

9.5 Задача максимізації пропускної спроможності каналу зв'язку

Максимальна кількість інформації, яка може бути передана по каналу зв'язку із смугою частот $f_1 < f < f_2$ при будь-якій малій імовірності помилки, визначається (за К. Шенноном) за формулою [3, 15]:

$$R = \int_{f_1}^{f_2} \log_2 \left[\frac{S(f) + n(f)}{n(f)} \right] df, \quad (9.30)$$

де $S(f)$ і $n(f)$ – функції спектральної щільності потужності корисного сигналу і шуму, відповідно.

Якщо спектральні щільності потужності сигналу і шуму є частотно незалежними в смузі (f_1, f_2) , тобто

$$R = \Delta f \log_2 \left(\frac{P_c}{P_{ш}} + 1 \right),$$

де $\Delta f = f_2 - f_1$, $P_n = \int_{f_1}^{f_2} S(f) df$ – повна потужність сигналу,

$$P_o = \int_{f_1}^{f_2} n(f) df - \quad (9.31)$$

повна потужність шуму.

Поставимо задачу про відшукування спектра щільності потужності корисного сигналу $S(f)$ з метою забезпечення (при фіксованій потужності сигналу $P_c = \tilde{P}_c$ і заданій спектральній щільності потужності шуму $n(f)$) максимальної швидкості передачі інформації. Таким чином, потрібно знайти максимум функціоналу

$$R[S(f)] = \int_{f_1}^{f_2} \log_2 \left[\frac{S(f) + n(f)}{n(f)} \right] df \quad (9.32)$$

за додатковою умовою

$$P_c = \int_{f_1}^{f_2} S(f) df = \tilde{P}. \quad (9.33)$$

Використовуючи термінологію шостого розділу, можемо сказати, що поставлене завдання є ізопериметричним з вільними кінцями, причому підінтегральні вирази в (9.32) і (9.33) не містять функції $S'(f)$.

Склавши відповідно до методу множників Лагранжа допоміжний функціонал

$$\hat{R}[S(f)] = \int_{f_1}^{f_2} \left\{ \log_2 \left[\frac{S(f) + n(f)}{n(f)} \right] + \lambda S(f) \right\} df, \quad (9.34)$$

випишемо для нього рівняння Ейлера

Підставимо значення $g_0^{(1)}$ і $g_1^{(1)}$ у вираз для функції $D^2(g_0, g_1)$ і знайдемо $\alpha_1^{(0)}$, для чого дослідимо отриманий вираз (функцію $D^{2(1)}(g_0^{(1)}, g_1^{(1)})$) на екстремум по змінній α_1 аналітично

$$D^2(\alpha_1) = 3\alpha_1^2 - 12\alpha_1 + 12.$$

Відповідно до необхідних умов екстремуму функції – рівності нулю її першої похідної по змінній α_1 , маємо

$$\frac{d(D^2(\alpha_1))}{d\alpha_1} = 6\alpha_1 - 12 = 0; \quad \alpha_1^{* \text{ опт}} = 2.$$

Знаходимо коефіцієнти коректора на 2-й внутрішній ітерації

$$g_0^{(1)} = 5, \\ g_1^{(1)} = g_1^{(0)} + \alpha_1^{(0)} = 0 + 2 = 2.$$

Визначаємо значення цільової функції на 2-ій ітерації

$$D^{2(1)}(g_0^{(1)}, g_1^{(1)}, \alpha_1^{(0)}) = 3 \cdot 2^2 - 12 \cdot 2 + 12 = 0; \\ D^{2(1)}(g_0^{(1)}, g_1^{(1)}) < D^{2(0)}(g_0^{(0)}, g_1^{(0)}).$$

Мінімум досягнутий, $D^2 = 0$ при $g_0 = 5$ і $g_1 = 2$.

На рис.9.15 показана траєкторія пошуку мінімуму функції $D^2(g_0, g_1)$ методом покоординатного спуску, а саме функції $D^2(5, 2)$.

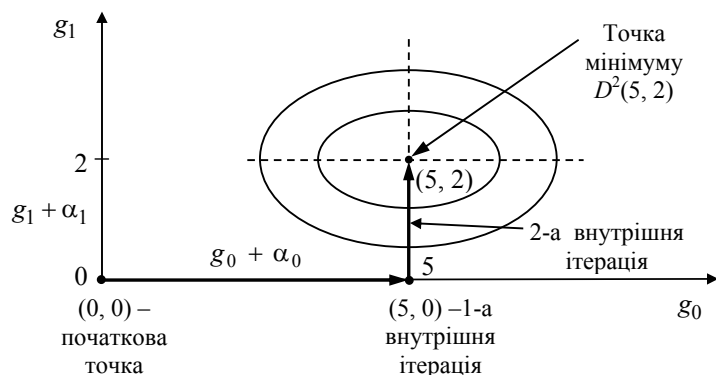


Рисунок 9.15 – Траєкторія пошуку методом покоординатного спуску мінімуму функції $D^2(g_0, g_1) = 2g_0^2 + 3g_1^2 - 20g_0 - 12g_1 + 62$

$$y(x) = \int [3x^2 + C_1] dx + C_2 = x^3 + C_1x + C_2.$$

Запишемо систему рівнянь для визначення C_1 і C_2

$$\left. \begin{aligned} y(0) &= 0^3 + C_1 \cdot 0 + C_2 = 0, \\ y(1) &= 1^3 + C_1 \cdot 1 + C_2 = 1, \end{aligned} \right\}$$

Отже, $C_1 = 0$, $C_2 = 0$ і $y(x) = x^3$.

На жаль, у загальному випадку питання про інтегрування рівняння Ейлера вирішується далеко не так просто, як в розглянутих прикладах. Вкажемо деякі окремі випадки, в яких це рівняння може бути зведене до рівняння першого порядку або навіть повністю проінтегровано у квадратурах:

1. Підінтегральна функція $P(x, y, y')$ не залежить від y . Припустимо, що даний функціонал має вид

$$\int_{x_0}^{x_1} F(x, y') dx,$$

тобто F явно не містить y . В цьому випадку рівняння Ейлера приймає вид:

$$\frac{d}{dx} Fy' = 0,$$

і має, очевидно, перший інтеграл

$$Fy' = C.$$

Це рівняння першого порядку, що не містить y . Вирішивши його відносно y' , отримуємо співвідношення виду:

$$y' = f(x, C),$$

звідки y знаходиться квадратурою.

2. Підінтегральна функція не залежить від x тобто

$$J[y(x)] = \int_{x_0}^{x_1} F(x, y') dx.$$

В цьому випадку $F_{xy'} = 0$ і рівняння Ейлера має вид:

$$F_y - F_{yy'} - F_{y'y'} y'' = 0.$$

Помноживши обидві частини цього рівняння на y' , отримаємо нове рівняння, яке можна переписати у виді:

$$\frac{d}{dx} [F - y'Fy'] = 0.$$

Дійсно

$$\begin{aligned} \frac{d}{dx} [F - y'Fy'] &= F_y y' + F_{y'y'} y'' - y'' F_{y'} - y' \frac{d}{dx} F_{y'} = \\ &= F_y y' + F_{y'y'} y'' - y'' F_{y'} - y' [F_{yy'} y' + F_{y'y'} y''] = \\ &= F_y y' - y' [F_{yy'} + F_{y'y'} y''] = y' [F_y - F_{yy'} y' - F_{y'y'} y''], \end{aligned}$$

звідки отримуємо, що в даному випадку рівняння Ейлера має перший інтеграл $F - y'F_{y'} = C$.

3. F не залежить від y' : $F = F(x, y)$. Рівняння Ейлера в даній ситуації має вид:

$$F_y(x, y) = 0, \text{ оскільки } F_{y'} = 0.$$

Рішення отриманого кінцевого рівняння $F_y(x, y) = 0$ не містить елементів вільності і тому не задовольняє граничним умовам $y(x_0 = x_0)$ і $y(x_1 = x_1)$, отже, рішення даної варіаційної задачі не існує. Лише у виняткових випадках, коли крива

$$F(x, y) = 0$$

проходить через граничні точки (x_0, y_0) і (x_1, y_1) існує функція, на якій може досягатися екстремум.

4. F залежить лише від y' . Рівняння Ейлера має вид $F_{y'y'}y''$, оскільки $F_y = F_{xy'} = F_{yy''} = 0$ звідси $y'' = 0$ або $F_{y'y'} = 0$. Якщо $y'' = 0$, то $y = c_1x + c_2$ – двопараметричне сімейство прямих ліній.

Якщо ж рівняння $F_{y'y'} = 0$ має один або декілька дійсних коренів $y' = k_i$, то $y = k_i x + c$ і одержуємо однопараметричне сімейство прямих, що міститься в отриманому вище двопараметричному сімействі $c_1x + c_2$.

Таким чином, у випадку $F = F(y')$ екстремалами є всілякі прямі лінії $y = c_1x + c_2$.

Знову звернемося до прикладів.

Приклад 3. (Відноситься одночасно до випадків 2 і 3).

Скористаємося результатами випадку 3 для відшукування екстремалей функціоналу

$$I = \int_{x_0}^{x_2} y^2 dx; \quad y(x_0) = y_0; \quad y(x_1) = y_1.$$

Маємо

$$F_y = 2y = 0.$$

Екстремаль $y = 0$ проходить через граничні точки тільки при $y_0 = 0$, $y_1 = 0$. Якщо $y_0 = 0$ і $y_1 = 0$, то, очевидно, функція $y = 0$ реалізує мінімум функціоналу I , оскільки $I \geq 0$, причому $y = 0$. Якщо ж хоча б одне з чисел y_0 і y_1 не дорівнює нулю, то мінімум функціоналу на безперервних функціях не досягається. Це пояснюється тим, що можна вибрати послідовність безперервних функцій $y_n(x)$, графіки яких більш крутіше спускаються із точки до вісі абсцис дуги кривої, потім з відрізка вісі абсцис, майже співпадаючого з відрізком $[x_0, x_1]$ і, нарешті, біля точки x_1 , що круто піднімається до точки (x_1, y_1) дуги кривої (рис. 6.10).

Очевидно, що на кривих такої послідовності значення функціоналу скільки завгодно мало відрізняються від нуля і, отже, нижня грань не може досягатися

9.4.5 Приклад розрахунку оптимальних значень коефіцієнтів гармонійного коректора методом покоординатного спуску

Знайти оптимальні значення коефіцієнтів гармонійного коректора g_0 і g_1 , що доставляють мінімум функції (середньоквадратичної похибки)

$$D^2(g_0, g_1) = 2g_0^2 + 3g_1^2 - 20g_0 - 12g_1 + 62$$

методом покоординатного спуску.

1-ий крок. Задання початкових значень:

$$g_0^{(0)} = 0; \quad g_1^{(0)} = 0.$$

Тоді

$$D^{2(0)}(g_0^{(0)}, g_1^{(0)}) = 62.$$

2-ий крок (1-а та 2-а внутрішні ітерації).

1 внутрішня ітерація:

$$g_0^{(1)} = g_0^{(0)} + \alpha_0^{(0)} = 0 + \alpha_0^{(0)};$$

$$g_1^{(1)} = g_1^{(0)} = 0.$$

Підставимо значення $g_0^{(1)}$ і $g_1^{(1)}$ у вираз для функції $D^2(g_0, g_1)$.

Знайдемо $\alpha_0^{(0)}$ опт, для чого виконаємо дослідження отриманого виразу (функції $D^{2(1)}(g_0^{(1)}, g_1^{(1)})$) на екстремум по змінній α_0 аналітично

$$D^2(\alpha_0) = 2\alpha_0^2 - 20\alpha_0 + 62.$$

Відповідно до необхідних умов екстремуму функції – рівності нулю її першої похідної маємо

$$\frac{d(D^2(\alpha_0))}{d\alpha_0} = 4\alpha_0 - 20 = 0; \quad \alpha_0^* \text{ опт} = 5.$$

Знаходимо коефіцієнти коректора на 1-ій ітерації

$$g_0^{(1)} = g_0^{(0)} + \alpha_0^{(0) \text{ опт}} = 0 + 5 = 5;$$

$$g_1^{(1)} = g_1^{(0)} = 0.$$

Визначаємо значення цільової функції на 1-ій ітерації

$$D^{2(1)}(g_0^{(1)}, g_1^{(1)}, \alpha_0^{(0) \text{ опт}}) = 2 \cdot 5^2 - 20 \cdot 5 + 62 = 12.$$

Перевіряємо виконання умови

$$D^{2(1)}(g_0^{(1)}, g_1^{(1)}) < D^{2(0)}(g_0^{(0)}, g_1^{(0)}).$$

2-а внутрішня ітерація

$$g_0^{(1)} = 5;$$

$$g_1^{(1)} = g_1^{(0)} + \alpha_1^{(0)} = 0 + \alpha_1^{(0)}.$$

середньоквадратичної похибки можна використовувати всі результати, що отримані в теорії корекції, вважаючи передатну функцію каналу рівною одиниці.

Зауваження 9.3. Якщо множина частот P збігається з проміжком $(-\pi/T, \pi/T)$ – інтервалом ортогональності функцій $e^{-i\omega k}$, $k = 0, 1, \dots, M-1$, то коефіцієнти, що мінімізують похибку (9.7) можуть бути знайдені як коефіцієнти Фур'є функції $G_{\text{ід}}(e^{i\omega T})$ [9, 10]

$$g_k = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} G_{\text{ід}}(e^{i\omega T}) e^{-i\omega k T} d\omega, \quad k = 0, 1, \dots, M-1$$

Якщо множина P не збігається з інтервалом ортогональності розглянутих експонентних функцій, то аналітичне рішення задачі важке і варто скористатися якою-небудь з чисельних процедур оптимізації. У силу ряду причин, однак, найбільш часто ЦФ проєктують, спираючись на чебишевський критерій точності синтезу; при цьому коефіцієнти g_k , $k = 0, 1, \dots, M-1$, знаходять з умови мінімуму наступного виразу:

$$\max W(\omega) \left| G_{\text{ід}}(e^{i\omega T}) - \int_{k=0}^{M-1} g_k e^{-i\omega k T} \right|. \quad (9.29)$$

Критерій (9.29) називають мінімаксним. Відповідно до цього критерію слід мінімізувати максимум (на множині точок P), зважений з невід'ємною ваговою функцією $W(\omega)$ модуля відхилення передатної функції ЦФ

$$\int_{k=0}^{M-1} g_k e^{-i\omega k T},$$

від заданої функції $G_{\text{ід}}(e^{i\omega T})$.

Згідно з теорією передачі сигналів, критерій (9.29) можна інтерпретувати у такий спосіб. Якщо аргумент ω умовно вважати часом і для простоти вагу $W(\omega)$ покласти рівною одиниці, то число (9.29) є піковим значенням узятого за абсолютною величиною різницевого сигналу, що одержується відніманням із «сигналу» $G_{\text{ід}}(e^{i\omega T})$ «сигналу»

$$\int_{k=0}^{M-1} g_k e^{-i\omega k T}.$$

Для відшукування мінімуму (9.29) методи, які викладені в розділі щодо корекції лінійних спотворень каналів зв'язку, неефективні, тому що функційна залежність величини (9.29) від незалежних змінних g_0, g_1, \dots, g_{M-1} виявляється поверхнею зі зламами. Добрі результати по розв'язанню даної задачі забезпечують методи лінійного програмування, однак найбільшою ефективністю серед відомих методів відшукування чебишевських наближень має спеціально розроблений для подібних ситуацій алгоритм, програмне забезпечення якого є в монографії [9].

на безперервній кривій, оскільки для будь-якої безперервної кривої $y = y(x)$, що відрізняється від тотожного нуля, інтеграл $\int_{x_0}^{x_1} y^2(x) dx > 0$. Ця нижня грань значень функціоналу досягається на розривній функції.

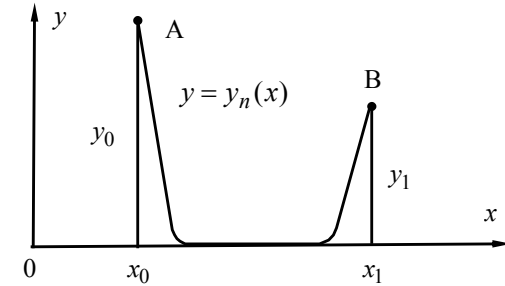


Рисунок 6.10 – Крива $y = y(x)$

Приклад 4. Відомо, що довжина дуги кривої, що з'єднує точки $A(x_0, y_0)$ і $B(x_1, y_1)$ є

$$l[y(x)] = \int_{x_0}^{x_1} \sqrt{1 + (y')^2} dx.$$

Поставимо питання про відшукування кривої $y(x)$, дуга якої з'єднує точки A і B (тобто $y(x_0) = y_0$, $y(x_1) = y_1$) має мінімальну довжину. Відповідь на це питання загальновідома: найменшою довжиною має відрізок прямої лінії, що з'єднує дані точки. Цей же висновок виходить формально з рівнянь Ейлера, оскільки згідно з результатами випадку 4 екстремаліями тут є всілякі прямі $y = c_1 x + c_2$.

На закінчення відзначимо, що приведені тут результати узагальнюються на функціонали більш загального виду, залежні як від більшого числа змінних, так і похідних вищого порядку, наприклад

$$J[x, y_1(x), y_2(x), y'_1(x), y'_2(x)] = \int_{x_0}^{x_1} F(x, y_1, y_2, y'_1, y'_2) dx.$$

6.4 Задачі з рухливими кінцями і задачі на умовний екстремум

У попередньому параграфі при дослідженні функціоналу $J = \int_{x_0}^{x_1} F(x, y, y') dx$ передбачалося, що граничні точки (x_0, y_0) і (x_1, y_1) задані.

Передбачимо тепер, що одна або обидві граничні точки можуть переміщатися (рис. 6.11).

Тоді клас допустимих кривих розширюється. Окрім кривих порівняння, що мають спільні граничні точки з досліджуваною кривою, можна вже брати і криві зі зміщеними граничними точками [5, 6, 7].

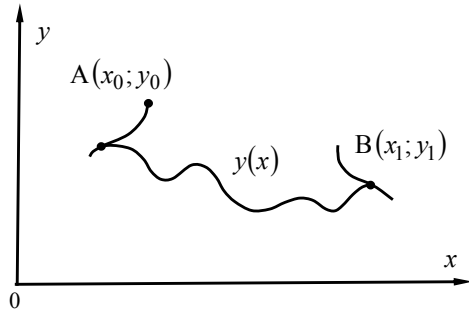


Рисунок 6.11 – Крива $y(x)$

Тому, якщо на кривих $y = y(x)$ досягається екстремум в задачі з рухливим граничними точками, то екстремум тим більше досягається по відношенню до вузького класу кривих, що мають спільні граничні точки з кривою $y = y(x)$. Отже, повинна бути виконана основна, необхідна для досягнення екстремуму в задачі з нерухомими границями умова – функція має бути вирішенням рівняння Ейлера, тобто криві $y(x)$, на яких реалізується екстремум в задачі з рухливими границями, повинні бути екстремаліями.

Загальне рішення рівняння Ейлера містить дві довільні постійні, для визначення яких необхідно мати дві умови. У задачі з нерухомими граничними точками такими умовами були:

$$y(x_0) = y_0 \text{ і } y(x_1) = y_1.$$

У задачі з рухливими границями одна або обидві ці умови відсутні, і відсутні умови для визначення довільних постійних загального вирішення рівняння Ейлера мають бути отримані із основної необхідної умови екстремуму – рівності нулю варіації δJ .

Приведемо без доказу ці умови для поодинокого і дуже поширеного випадку, званого задачею з вільними кінцями: серед всіх кривих, кінці яких лежать на двох заданих вертикалях $x = x_0$, $x = x_1$, (рис. 6.12), знайти ту, яка дає екстремум функціоналу

$$J = \int_{x_2}^{x_1} F(x, y, y') dx.$$

Теорема. У задачі з рухливими границями умови для визначення довільних постійних мають вид:

$$F_{y'}|_{x=x_0} = 0, \quad F_{y'}|_{x=x_1} = 0.$$

Приклад 5. Задача мінімізації функціоналу

$$G(e^{i\omega T}) = \int_{k=0}^{N-1} a_k e^{-i\omega k T}$$

Які достоїнства і недоліки ЦФ, побудованих за різними схемами рис. 9.13 і 9.14?

Головні особливості порівнюваних варіантів такі: БІХ-фільтр, що має відносно невеликий порядок $N - 1$ ($N = 7 - 9$) дозволяє реалізувати високоселективну фільтрацію з вгамівністю приблизно 70 – 85 дБ у смузі затримування при практично довільному співвідношенні смуг затримування і пропускання і смузі розфільтрування, рівній приблизно 30 % від смуги пропускання (смуга розфільтрування розташовується між смугою пропускання і смугою затримування). Застосування КІХ-фільтра для розв'язання цієї ж задачі може зажадати, щоб число відводів M у регістра зсуву на рис. 9.13 досягало багатьох десятків і навіть сотень, причому це число, як показали численні дослідження і розрахунки, пропорційне відношенню смуг затримування і пропускання [10].

Разом з тим БІХ-фільтри мають і істотний недолік: на протигагу ЦФ із кінцевою імпульсною характеристикою, вони мають, завдяки зворотним зв'язкам, схильність до самозбудження і виявляються досить чутливими до округлень результатів арифметичних операцій. У цьому зв'язку процесор, що реалізує БІХ-фільтр, повинний надавати більше двійкових розрядів для представлення чисел, ніж аналогічний процесор, призначений для КІХ-фільтрації.

Зазначені особливості КІХ- і БІХ-фільтрів обумовлюють їхню конкуренцію і співіснування в сучасній цифровій обробці сигналів.

Розглянемо оптимальний вибір коефіцієнтів передатних функцій ЦФ.

Центральним питанням проектування цифрових фільтрів є оптимальний вибір коефіцієнтів $g(kT)$, $k = 0, 1, \dots, M - 1$, якщо мова йде про КІХ-фільтр, або коефіцієнтів a_k , $k = 0, 1, \dots, N - 1$; $b_k = 0, 1, \dots, N - 1$, у випадку БІХ-фільтра. Зупинимося тільки на відносно простому випадку нерекурсивної фільтрації (див. рис. 9.13).

Звичайно в задачі проектування ЦФ зазначений деякий діапазон (множина) частот – позначимо її через P , – на якому задається «ідеальна» передатна функція $G_{ид}(e^{i\omega T})$. У випадку застосування, середньоквадратичного критерію точності синтезу ЦФ (або, середньоквадратичної похибки) коефіцієнти $g(kT)$, $k = 0, 1, \dots, M - 1$, у формулі (9.26) визначають з умови мінімуму виразу

$$D_{ид}^2 = \int_P \left| G_{ид}(e^{i\omega T}) - \int_{k=0}^{M-1} g_k e^{-i\omega k T} \right|^2 d\omega, \quad (9.28)$$

де для спрощення введено позначення $gk = g(kT)$, $k = 0, 1, \dots, M - 1$.

Вираз (9.28) нагадує середньоквадратичну похибку коректування; однак розглянута тут ситуація є більш простою: коректований канал відсутній. Іншими словами, при проектуванні цифрових фільтрів за критерієм мінімуму

Формула (9.26) записана стосовно до випадку обробки дискретного сигналу за допомогою ЦФ. Щоб підкреслити цю обставину, як аргумент передатної функції, пишуть експоненту $e^{i\omega T}$, а не просто частоту ω , як це має місце в передатній функції гармонійного синтезатора.

Перейдемо до побудови БІХ-фільтрів. Структурну схему БІХ-фільтра можна одержати, якщо в схему рис. 9.13 ввести зворотні зв'язки, як показано на рис. 9.14. Фільтр за схемою рис. 9.14 прийнято називати рекурсивним, а за схемою рис. 9.13 – нерекурсивним.

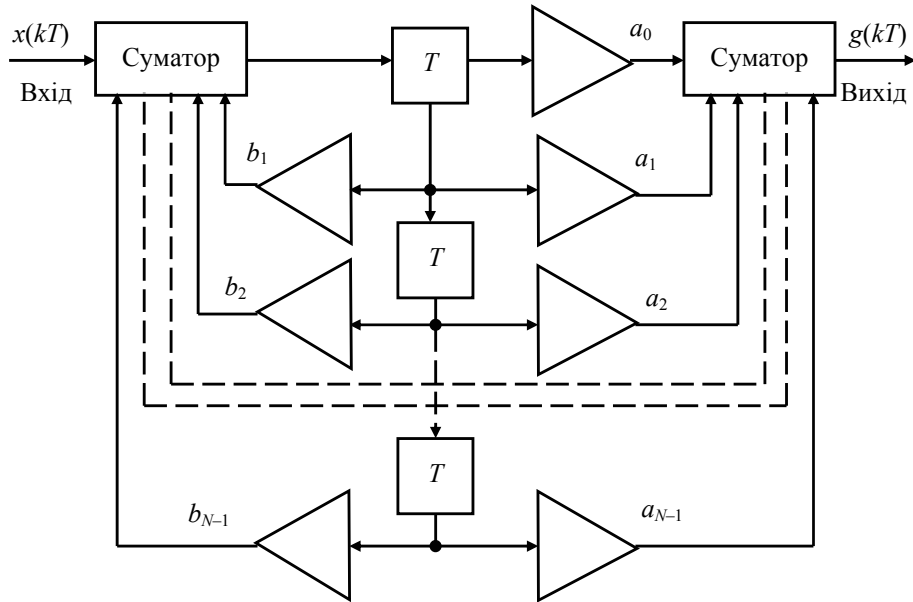


Рисунок 9.14 – Структурна схема алгоритму БІХ-фільтрації

Можна показати, що передатна функція БІХ-фільтра має вид

$$G(e^{i\omega T}) = \frac{\sum_{k=0}^{N-1} a_k e^{-i\omega k T}}{1 + \sum_{k=0}^{N-1} b_k e^{-i\omega k T}} \quad (9.27)$$

Якщо розірвати всі зворотні зв'язки рис. 9.14 (при цьому варто покласти $b_k = 0, k = 1, \dots, N - 1$), то з одного боку, схема рис. 9.14 перетвориться в схему рис. 9.13, а з іншого боку – дробово-раціональна передатна функція (9.27) перетвориться в багаточленний вираз типу (9.26):

$$J = \int_{x_2}^{x_1} y^2 dx, \quad y(x_0) = y_0, \quad y(x_1) = y_1$$

має рішення лише при $y_0 = y_1 = 0$.

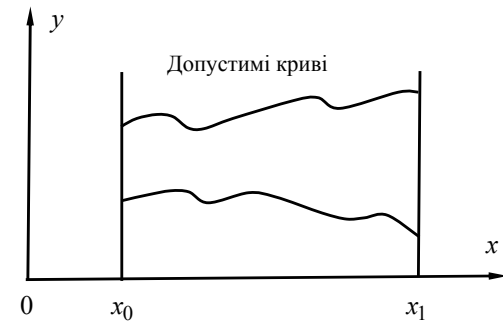


Рисунок 6.12 – Криві на заданих вертикалях $x = x_0, x = x_1$

Ця задача тепер формулюється без граничних умов, тобто з вільними кінцями. Оскільки підінтегральний вираз функціоналу J від y' не залежить, то приведені в теоремі умови

$$F_{y'}|_{x=x_0} = 0 \quad \text{і} \quad F_{y'}|_{x=x_1} = 0$$

визначаються на тотожності виду $0 = 0$ і, значить, виконуються автоматично при будь-яких $y(x_0)$ і $y(x_1)$, а за відсутності яких-небудь обмежень рішення рівняння Ейлера $y(x) \equiv 0$ є рішенням поставленої задачі на екстремум і задає функціоналу мінімально можливе значення, яке дорівнює нулю.

Даний приклад дозволяє, очевидно, встановити одну загальну для всіх функціоналів виду (6.6), що не містять y' , закономірність. У екстремальних задачах з вільними кінцями для функціоналів цього типу умови для визначення довільних постійних відсутні. Проте ця обставина допомагає рішенням задачі, оскільки в даній ситуації рівняння Ейлера є кінцевим і його рішення не містить довільних постійних.

Задачі на умовний екстремум. Низка додатків варіаційного числення приводить до задач, в яких на допустимі криві, окрім граничних умов, накладаються умови зовсім іншого типу. Розглянемо, як приклад, так звану ізопериметричну задачу.

Серед всіх кривих, що задовольняють умови $y(x_0) = y_0$ і $y(x_1) = y_1$, на яких функціонал

$$K[y(x)] = \int_{x_0}^{x_1} G(x, y, y') dx$$

набуває заданого значення L , знайти ту, для якої інший функціонал $J[y(x)]$ досягає екстремуму.

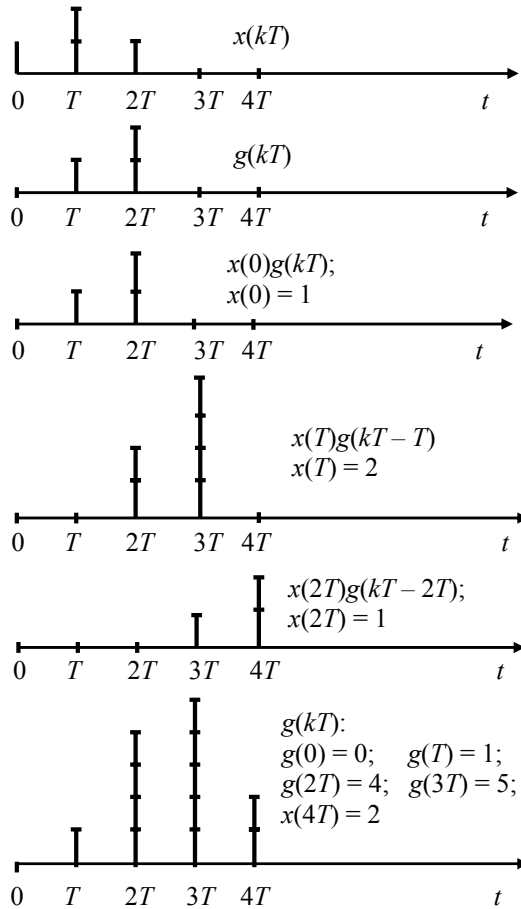


Рисунок 9.12 – Оброблення сигналів у цифровій формі

Істотно, що для переходу від одного типу фільтра до іншого, а саме від ФНЧ до ФВЧ, варто перемінити числа, що зберігаються в пам'яті процесора, $g(0), g(T), \dots$ на інші, відповідним чином, розраховані значення.

У значному числі ситуацій, що виникають на практиці, необхідна імпульсна реакція (характеристика) має кінцеве число не рівних нулю відліків $g(m), m = 0, 1, \dots, M - 1$. Фільтри такого типу називають КІХ-фільтрами. На противагу цьому, фільтри, чиї імпульсні характеристики мають нескінченне число ненульових відліків, називають БІХ-фільтрами.

Зрозуміло, значення нескінченного числа відліків $g(m), k = 0, \pm 1, \dots$, не можуть бути введені в цифрову пам'ять безпосередньо, у цьому випадку в пам'ять записують кінцеве число параметрів, що непрямим шляхом реалізують нескінченну імпульсну характеристику.

$$f = ax_1^2 + bx_2^2$$

за додаткової умови

$$\varphi = x_1^2 + x_2^2 - 1 = 0,$$

або, що те ж саме

$$x_1^2 + x_2^2 = 1. \quad (6.13)$$

Поверхня $f = ax_1^2 + bx_2^2$ (двовимірна парабола) при $b > a$ показана на рис. 6.13.

Жирною лінією на цьому рисунку показана крива, що лежить, з одного боку, на цій поверхні і, з іншого боку, задовольняюча умові (6.13). Ця крива виходить, як перетин циліндра, основою якого є коло (6.13) з поверхнею $f = ax_1^2 + bx_2^2$. З ситуацією, що виникла допомагають розібратися лінії рівнів поверхні $f(x_1, x_2) = C_k, k = 1, 2, \dots$, які показані на рис. 6.14.

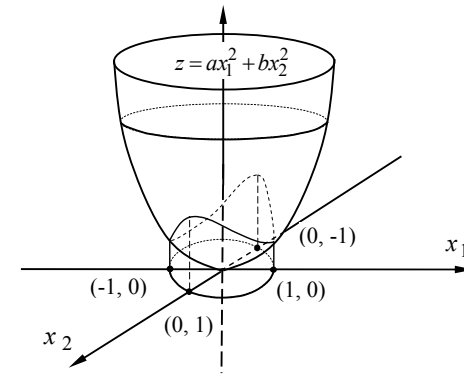


Рисунок 6.13 – Поверхня $f = ax_1^2 + bx_2^2$ при $b > a$

Розглядаючи рис. 6.13 і 6.14, неважко переконатися, що точки на площині x_1 0 x_2 з координатами $(1, 0)$ і $(-1, 0)$ є точками мінімуму поверхні $f = ax_1^2 + bx_2^2$, а точки з координатами $(0, 1)$ і $(0, -1)$ – її максимуму.

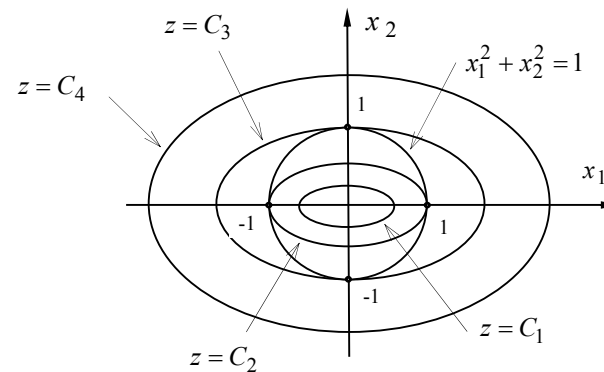


Рисунок 6.14 – Лінії рівнів поверхні $f(x_1, x_2) = C_k, k = 1, 2, \dots$

Маючи ясне уявлення про характер рішення даної задачі (його вдалося отримати завдяки двомірності задачі; зрозуміло, в разі більшого числа змінних виникли б істотні труднощі при здобутті рішення геометричним шляхом), приступимо до знаходження екстремальних точок аналітичними методами.

Скористаємося спочатку «прямою» процедурою: виразимо одну із змінних величин через іншу за допомогою умови (6.13)

$$x_1^2 = 1 - x_2^2, \quad (6.14)$$

і підставимо отриманий результат у вираз для поверхні $f(x_1, x_2)$:

$$g(x_2) = a(1 - x_2^2) + bx_2^2 = a + (b - a)x_2^2.$$

Функція $g(x_2) = a + (b - a)x_2^2$ досягає екстремуму у точці x_2 , що задовольняє рівнянню

$$\frac{dg(x_2)}{dx_2} = 2(b - a)x_2 = 0,$$

тобто при $x_2 = 0$. З (6.14) виходить, що інша координата екстремуму x_1 може набувати два значення

$$x_1 = \pm\sqrt{1 - x_2^2} = \pm 1.$$

Визначивши знаки другої похідної $\frac{d^2g(x_2)}{dx_2^2}$, неважко переконатися, що

точки з координатами $(\pm 1, 0)$ є точками мінімуму (що, втім, і так очевидно).

Відзначимо, що хоча у функції $g(x_2)$ у середині області її визначення (цією областю у випадку (6.13) є відрізок $[-1, 1]$) існує єдина екстремальна точка $x_2 = 0$, найбільші і найменші значення цієї функції можуть досягатися в граничних точках $x_2 = \pm 1$. Неважко перевірити, що в даному випадку мова йде про максимальні значення.

Таким чином, точки з координатами $(0, \pm 1)$ є точками максимуму функції $f = ax_1^2 + bx_2^2$ за додаткової умови (6.13). Вирішимо це ж завдання методом невизначених множників. Для цього введемо функцію

$$h(x_1, x_2) = f(x_1, x_2) + \lambda\varphi(x_1, x_2) = ax_1^2 + bx_2^2 + \lambda[x_1^2 + x_2^2 - 1].$$

Система рівнянь для визначення параметрів x_1, x_2, λ має вид

$$\left. \begin{aligned} \frac{dh}{dx_1} &= 2(a + \lambda)x_1 = 0 \\ \frac{dh}{dx_2} &= 2(b + \lambda)x_2 = 0 \\ \frac{dh}{d\lambda} &= x_1^2 + x_2^2 - 1 = 0 \end{aligned} \right\}.$$

Перші два рівняння даної системи можуть бути задоволені при $x_1 = 0$, $x_2 = 0$. Проте, третє рівняння показує, що x_1 і x_2 одночасно в нуль обертаються не можуть. Точніше кажучи, при $x_1 = 0$ повинно виконуватися: $x_2 = \pm 1$, і,

$$u(kT) = \begin{cases} 1, & k=0 \\ 0, & k=\pm 1, \pm 2, \dots \end{cases} \quad (9.22)$$

показаний на рис. 9.11, що грає роль δ -функції в теорії лінійних дискретних систем.

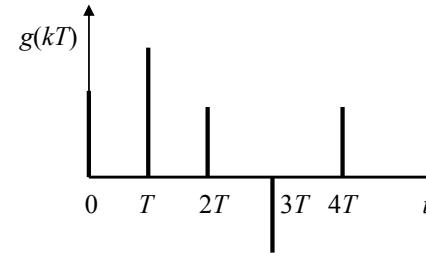


Рисунок 9.10 – Імпульсна реакція фільтра

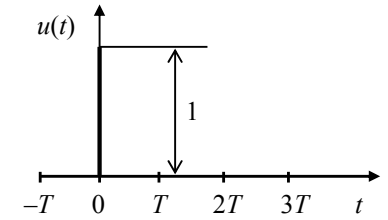


Рисунок 9.11 – Вхідний сигнал

Розглянемо для спрощення вхідний сигнал цифрового фільтра у виді скінченної послідовності відліків

$$x(0), x(T), \dots, x[(N - 1)T] \quad (9.23)$$

Зауважимо, що m -й член послідовності відліків (9.23) – $x(m)$ рівно у $x(m)$ раз більше (або менше) сигналу (9.22); крім того, він запізнюється на mT секунд у порівнянні із сигналом рис. 9.11. Отже, при подачі на вхід лінійного цифрового фільтра такого сигналу, на його виході одержимо відклік

$$x(mT)g(kT - mT), \quad k = 0, \pm 1, \dots, \quad (9.24)$$

де $g(k)$, $k = 0, \pm 1, \dots$, – імпульсна реакція розглянутого фільтра. При подачі ж на вхід цього ЦФ усієї послідовності (9.23) одержимо, спираючись на визначення лінійності, вихідний сигнал у виді суми відклів типу (9.24):

$$y(kT) = \sum_{m=0}^{N-1} x(mT)g(kT - mT), \quad k = 0, \pm 1, \dots \quad (9.25)$$

Співвідношення (9.25), що є дискретним аналогом рівності (1.17), називають дискретною згортою. Важливість його полягає в тому, що, як виходить з (9.25), довільна лінійна стаціонарна обробка сигналів може бути виконана в цифровій формі за допомогою наступних щодо простих операцій: зсувів, множень і додавань. Порядок, у якому повинні виконуватися ці операції, роз'яснює рис. 9.12, на якому показано «проходження» сигналу $x(kT)$, $k = 0, \pm 1, \dots$, який має три відліки, що відмінні від нуля (1-ий рядок рис. 9.12), через ЦФ з імпульсною реакцією $g(k)$, $k = 0, \pm 1, \dots$, із двома ненульовими відліками (2-ий рядок рис. 9.12). 3-ій, 4-ий і 5-ий рядки даного рисунка отримані відповідно до формули (9.24); а на 6-ому рядку показано вихідний сигнал ЦФ $y(kT)$, $k = 0, \pm 1, \dots$, що визначається як сума дискретних сигналів рядків за номерами 3, 4 і 5.

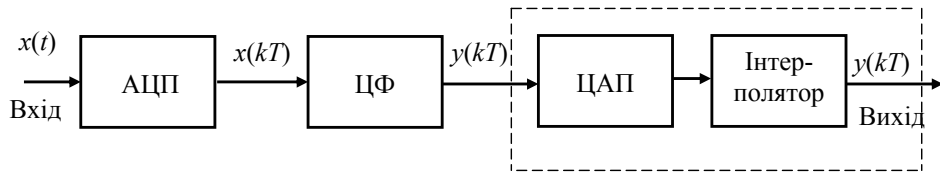


Рисунок 9.8 – Аналогово-цифрове та цифро-аналогове перетворення

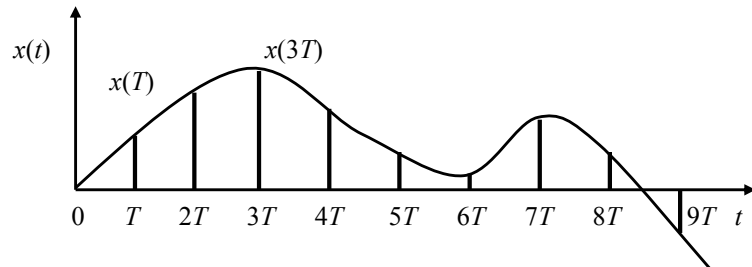


Рисунок 9.9 – Послідовність відліків сигналу $x(t)$

Послідовність відліків $x(kT)$, $k = 0, \pm 1, \dots$ з виходу АЦП надходить на вхід цифрового фільтра (ЦФ), що по заданому алгоритмі «перераховує» її в послідовність вихідних відліків $y(kT)$, $k = 0, \pm 1, \dots$ придушуючи або підсилюючи при цьому ті або інші частотні компоненти сигналу $x(t)$. Вихідна послідовність відліків ЦФ $y(kT)$, $k = 0, \pm 1, \dots$ також являє собою послідовність кодових комбінацій двійкового коду. За цими комбінаціями цифро-аналоговий перетворювач (ЦАП) відновлює імпульси прямокутної форми, чия амплітуда в точності дорівнює записаним у двійковій системі числення значенням $y(kT)$, $k = 0, \pm 1, \dots$ Аналоговий сигнал $y(t)$ знаходять за його відліками за допомогою фільтра-інтерполятора, який наближено реалізує співвідношення за теоремою Котельникова, що відновлює смугово-обмежений сигнал за його відліками. Штрихова лінія на рис. 9.8 поєднує ЦАП і інтерполятор у один вузол, що здійснює перетворення кодових комбінацій у відфільтрований аналоговий сигнал $y(t)$.

Якщо операції, що виконані АЦП, ЦАП і фільтром-інтерполятором є багато в чому стандартними і не залежними від призначення і властивостей цифрового фільтра, то сам ЦФ, точніше, його характеристики і параметри в значній мірі визначаються конкретною задачею фільтрації, що вирішує, даний ЦФ.

Які ж принципи побудови цього найбільше «інтелектуального» блоку схеми рис. 9.8?

Як відомо, теорія лінійних систем з дискретним часом подібна теорії лінійних аналогових систем. Для дискретних стаціонарних (тобто таких, що не змінюють з часом своїх характеристик) систем вводиться поняття імпульсної реакції (імпульсної характеристики), що являє собою відгук $g(kT)$, $k = 0, \pm 1, \dots$ (рис. 9.10) на вхідний сигнал виду

навіпаки, при $x_2 = 0$ має місце: $x_1 = \pm 1$. Отже, при $x_1 = 0$ і $x_2 = \pm 1$ необхідно для другого рівняння вибір $\lambda = -b$. При $x_2 = 0$ і $x_1 = \pm 1$ вибираємо $\lambda = -a$.

Як бачимо, метод неозначених множників приводить до того ж результату, що розглянуті вище підходи, включаючи геометричний аналіз досліджуваної поверхні.

Достоїнством неозначених множників є можливість рішення задачі на екстремум аналітично або за допомогою чисельних процедур (про них скажемо далі) в тих ситуаціях, коли систему (6.12) не вдається вирішити відносно яких-небудь m змінних.

Метод неозначених множників Лагранжа може бути перенесений також на задачі відшукування умовного екстремуму функціоналів.

Приведемо без доказу цей важливий результат. Поставлена вище ізометрична задача може бути зведена до наступної задачі на безумовний екстремум: знайти екстремум допоміжного функціоналу

$$J^*[y(x)] = \int_{x_0}^{x_1} [F(x, y, y') + \lambda G(x, y, y')] dx,$$

При $y(x_0) = y_0$, $y(x_1) = y_1$, де λ – множник Лагранжа, для визначення якого слід використовувати рівність

$$K[y(x)] = l.$$

Функція $J^*[y(x)]$ збігається з точністю до незалежного від вибору функції $y(x)$ доданку з наступним функціоналом, отримуваним формальним застосуванням до дослідження функціоналу $J[y(x)]$ викладеного вище методу неозначених множників для відшукування екстремумів функції:

$$J[y(x)] + \lambda \{K[y(x)] - l\} = \int_{x_0}^{x_1} [F(x, y, y') + \lambda G(x, y, y')] dx - \lambda l$$

і, як сказано вище, екстремалі функціоналу $J^*[y(x)]$ збігаються з екстремаліями функціоналу $J[y(x)]$ за умови $K[y(x)] = l$.

Дуже часто в додатках зустрічається наступна задача на умовний екстремум: знайти максимум (мінімум) функціоналу n змінних

$$J[y_1(x), y_2(x), \dots, y_n(x)] = \int_{x_0}^{x_1} [F(x, y_1, y_2, \dots, y_n, y_1', y_2', \dots, y_n')] dx \quad (6.15)$$

за наявності умов

$$\varphi_i(x, y_1, y_2, \dots, y_n) = 0, \quad i = 1, 2, \dots, m; \quad m < n, \quad (6.16)$$

де φ_i – завдання функції.

Ця задача рішається аналогічно ізопериметричній: складається допоміжний функціонал

$$\hat{J} = \int_{x_0}^{x_1} [F(x_1, y_1, y_2, \dots, y_n, y_1^1, y_2^1, \dots, y_n^1) + \sum_{i=1}^m \lambda_i(x) \varphi_i(x_1, y_1, y_2, \dots, y_n)] dx =$$

$$= \int_{x_0}^{x_1} \hat{F}(x_1, y_1, y_2, \dots, y_n, y_1^1, y_2^1, \dots, y_n^1) dx \quad (6.17)$$

де $\lambda_i(x)$, $i=1, 2, \dots, m$ – множники Лагранжа, цього разу вже функції, а не константи.

Функції $\lambda_i(x)$, $i=1, 2, \dots, m$, а також функції $y_i(x)$, $i=1, 2, \dots, n$, що задовольняють необхідним умовам екстремуму функціоналу (6.15) за умови (6.16), визначаються з рівнянь Ейлера для допоміжного функціоналу (6.17)

$$F_{y_i}^* - \frac{d}{dx} F_{y_i}^* = 0, \quad j=1, 2, \dots, n$$

і умов (6.17). Рівняння (6.16) можна також вважати за рівняння Ейлера для функціоналу (6.17), якщо за аргументи цього функціоналу вважати не тільки функції $y_1(x), y_2(x), \dots, y_n(x)$, але і $\lambda_1(x), \lambda_2(x), \dots, \lambda_m(x)$.

6.5 Достатні умови екстремуму

Доказ достатніх умов екстремуму функціоналу (6.6) є складним і вимагає залучення дуже тонких математичних міркувань. Приведемо в зв'язку з цим деякі класичні результати варіаційного числення без доказу.

Нижченаведене диференціальне рівняння відносно функції $u(x)$ називають рівнянням Якобі [3, 4]

$$F_{yy}u + F_{yy'} : u - \frac{d}{dx} (F_{yy}u + F_{y'y'}u') = 0 \quad (6.18)$$

де

$$F_{yy} = \frac{\partial^2 F(x, y(x), y'(x))}{\partial y^2};$$

$$F_{yy'} = \frac{\partial^2 F(x, y(x), y'(x))}{\partial y' \partial y};$$

$$F_{y'y'} = \frac{\partial^2 F(x, y(x), y'(x))}{\partial y' \partial y'}.$$

Причому в якості $y(x)$ у рівняння (6.18) підставляється рішення рівняння Ейлера для функціоналу

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y') dx,$$

що задовольняє заданим умовам на кінцях

$$(y(x_0) = y_0, \quad y'(x_1) = y_1).$$

Очевидно, що рівняння (6.18) – лінійне однорідне рівняння другого порядку відносно $u(x)$.

Завдяки своїй універсальності гармонійний коректор (синтезатор характеристик), винайдений у 30-х – роках Н. Вінером і Ю.В. Лі [4, 8], знайшов застосування в різних галузях радіоелектроніки, що далеко виходять за рамки задачі корекції лінійних спотворень. У техніці цифрової обробки сигналів пристрій за схемою рис. 9.7 називають цифровим КІХ-фільтром, тобто фільтром з кінцевою імпульсною характеристикою (реакцією); інша назва цієї структури – нерекурсивний цифровий фільтр (див. [9, 10]).

У техніці високошвидкісної передачі даних гармонійний синтезатор використовується не тільки у виді коректора лінійних спотворень, але і як пристрій більш загального призначення – оптимального фільтра, що здійснює притлумлення завад двох типів – флуктуаційних і міжсимвольних. Відомості про подібні результати є, наприклад, у монографії [11], а їх спрощений виклад – у навчальному посібнику [5]. В інших розробках гармонійний синтезатор використовується для зниження зазначених завад у сполученні з оптимальними нелінійними пристроями типу детектора Вітербі [12, 13]. Аналіз робіт за цим напрямком розміщено у навчальному посібнику [14] та істотно спирається на апарат математичної оптимізації – варіаційне числення.

9.4.4 Застосування оптимізації у задачах синтезу цифрових фільтрів

Розглянемо елементарні відомості про цифрову фільтрацію.

Цифрові фільтри за своєю суттю є спеціалізованими обчислювальними пристроями, що вирішують традиційні задачі фільтрації сигналів за допомогою методів і засобів обчислювальної математики й обчислювальної техніки. Оскільки цифрові обчислювальні пристрої довільного призначення, що здатні сприймати й обробляти інформацію лише в дискретні, віддалені один від одного на цілком визначену дистанцію моменти часу (позначимо їх так: $t = kT$, $k = 0, \pm 1, \dots$), то і цифрові фільтри мають таку особливість.

Практично цифрова фільтрація аналогових сигналів (функцій безперервного часу t) здійснюється у відповідності зі структурною схемою рис. 9.8. Аналоговий сигнал $x(t)$ надходить на вхід аналого-цифрового перетворювача (АЦП), на виході якого з'являються узяті відповідно до теореми Котельникова відліки $x(kT)$, $k = 0, \pm 1, \dots, n$; при цьому інтервал дискретизації T вибирається меншим або рівним величині $1/2f_b$, де f_b – верхня частота в спектрі сигналу $x(t)$ (рис. 9.9).

На практиці АЦП видає відліки $x(kT)$, $k = 0, \pm 1, \dots$ у цифровій формі, тобто у виді комбінацій двійкового коду, так, що кожна комбінація цього коду є представленням відповідного відліку в двійковій системі числення. Точність такого опису дискретних відліків сигналу $x(t)$ залежить від числа розрядів у кодових комбінаціях і, теоретично, може бути зроблена як завгодно високою. Саме тому надалі будемо зневажати ефектами, що викликані збереженням лише кінцевого числа розрядів у кодованому записі відліків $x(kT)$. Це скорочене представлення відлікових значень аналогового сигналу здійснюється в АЦП за допомогою операцій квантування або округлення [9, 10].

Передатна функція гармонійного коректора

$$G(\omega) = \sqrt{T/2\pi} \sum_{k=0}^n g_k e^{-i\omega kT}, \quad -\Omega < \omega < \Omega, \quad (9.21)$$

допускає достатньо просту реалізацію.

Дійсно, з теореми запізнювання виходить, що передатну функцію $e^{-i\omega kT}$ можна реалізувати за допомогою лінії затримки на час kT . Тому для синтезу електричної схеми з функцією передачі (9.21) з точністю до несуттєвого множника $\sqrt{T/2\pi}$ достатньо мати секційовану лінію затримки, до виходів секцій якої підключені підсилювачі з коефіцієнтами підсилення g_k , $k = 0, 1, \dots, n$, а виходи підсилювачів приєднані до загального суматора.

Зауваження 9.1.

Вираз (9.21) являє собою кінцеву суму ряду Фур'є, у якому коефіцієнти g_k , $k = 0, 1, \dots, n$, можуть вибиратися довільно. З відомої властивості рядів Фур'є [7, 8] виходить, що функція (9.21) при достатньо великому n і належному виборі коефіцієнтів g_k , $k = 0, 1, \dots, n$, дозволяє наблизити практично будь-яку запропоновану залежність у робочому діапазоні частот $(0, \Omega)$.

Зауваження 9.2.

Передатна функція гармонійного коректора (синтезатора) (9.19) є періодичною, тому що таку властивість мають його базисні функції

$$\sqrt{T/2\pi} e^{-i\omega kT} = \sqrt{T/2\pi} \cos \omega kT - i\sqrt{T/2\pi} \sin \omega kT, \quad k = 0, 1, \dots, n.$$

Серед інтервалів періодичності (довжиною 2Ω) на практиці найчастіше використовується проміжок $(-\Omega, \Omega)$. Якщо лінія затримки (див. рис. 9.7) виконана так, що з «доброю» точністю відтворює базисні функції (9.20) у смузі частот, наприклад, $30\Omega < \omega < 31\Omega$, то в зазначеному діапазоні гармонійний синтезатор забезпечить синтез практично довільних передатних функцій.

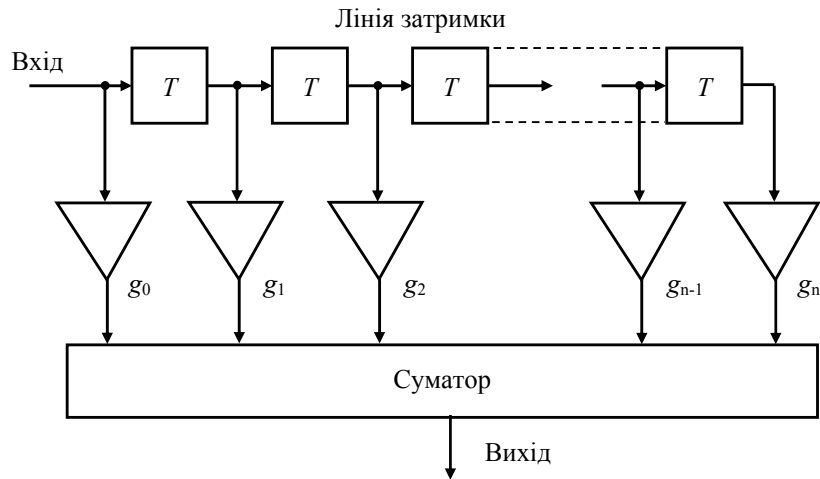


Рисунок 9.7 – Структурна схема гармонійного коректора

Якщо існує рішення рівняння Якобі $u(x)$, що обертається в 0 при $x = x_0$ і, яке більш не обертається в 0 ні в одній з точок відрізка $x_0 \leq x \leq x_1$,

то говорять, що виконана умова Якобі.

Наведемо деякі варіанти наборів умов, виконання яких вистачає для досягнення екстремуму:

1. Слабкий мінімум

а) $F_y - \frac{d}{dx} F_{y'} = 0$ ($y(x)$ є екстремаль),

б) умова Якобі виконана,

в) $F_{y'y'} > 0$ на екстремалі.

2. Сильний мінімум

а) $F_y - \frac{d}{dx} F_{y'} = 0$ ($y(x)$ є екстремаль),

б) умова Якобі виконана,

в) $F_{y'y'}(x, y, y') \geq 0$ для точок (x, y) , близьких до точок на екстремалі і для довільних значень y' .

При цьому передбачається, що функція $F(x, y, y')$ тричі диференційована за y' для будь-яких y' .

Відзначимо (також без доказу) наступні два важливих твердження.

1. Якщо у формулюванні умови Якобі відрізок $x_0 \leq x \leq x_1$ замінити інтервалом $x_0 < x < x_1$, то ця умова стає необхідною умовою екстремуму.

2. Стосовно функціоналів, у яких $F(x, y, y')$ не залежить від y' справедливе твердження: якщо існує рішення рівняння Ейлера, що задовольняє граничним умовам, воно неодмінно реалізує екстремум цього функціоналу. Це твердження витікає з наступного (грубого) достатнього критерію.

Достатньою умовою мінімуму є набір з наступних трьох співвідношень:

а) $F_y - \frac{d}{dx} F_{y'} = 0$;

б) $F_{y'y'} \geq 0$;

в) $F_{yy} \cdot F_{y'y'} - F_{y'y'}^2 \geq 0$.

Для отримання достатніх умов максимуму у всіх наведених вище варіантах умов, треба поміняти знаки нерівності на протилежні.

Контрольні питання та задачі до розділу 6

1. Чим характеризується задача про брахістохрону?
2. Що таке функціонал?
3. Що таке неперервний функціонал?
4. Що таке варіація функціоналу?
5. Який функціонал називається безперервним?

6. Чому при передачі імпульсних сигналів по смугово-обмеженому каналу зв'язку енергія позасмугових випромінювань не повинна перевищувати деякої нормативної величини?
7. Якими виразами можна представити сам сигнал та його спектр?
8. Як визначається інтенсивність випромінювання енергії сигналу поза діапазоном частот (ω_1, ω_2)?
9. Покажіть, що відносна частка енергії сигналу в заданому діапазоні частот (ω_1, ω_2) є функціоналом від форми цього сигналу?
10. Чому залежність імовірності помилки поодинокого елемента P_0 на виході дискретного каналу системи ПД від форми амплітудно-частотної характеристики $|K(\omega)|$ каналу зв'язку є функціоналом?
11. Які дві криві є близькими у розумінні близькості нульового порядку?
12. Які дві криві є близькими у розумінні близькості першого порядку?
13. В яких випадках функціонал перетворюється на функцію?
14. В чому полягає основна лема варіаційного числення?
15. В чому полягає задача з рухливими границями?
16. Знайти найкоротшу відстань між кривими $y = x^2$ і $x - y = 5$.
17. Для функціоналу знайти екстремаль і побудувати її графік:
 - а) $J(y) = \int_{-1}^1 (y'^2 + 4y^2 - 8xy + 2y^2) dx$; $y(-1) = 3$; $y(1) = 1$;
 - б) $J(y) = \int_0^1 (y\sqrt{1+y'^2}) dx$; $y(0) = 2$; $y(1) = 3$;
 - в) $J(y) = \int_{-1}^1 (y'^2 - 4y' \sin 2x - x^2) dx$; $y(0) = 1$; $y(2) = -1$;
 - г) $J(y) = \int_0^1 (\frac{\sqrt{1+y'^2}}{y}) y dx$; $y(0) = 2$; $y(1) = 1$;
 - д) $J(y) = \int_0^2 (y'^2 - 4y'e^{2x} + \sin^2 x) dx$; $y(0) = 1$; $y(2) = -2$.
18. Для функціоналу, що залежить від декількох функцій, знайти екстремаль і побудувати її графік

- а) $J(y) = \int_0^1 (2y'z' - y^2 + z^2 - 2ye^x) dx$; $y(0) = 0$; $z(0) = 1$;
 $y(1) = 1$; $z(1) = 0$;
- б) $J(y) = \int_0^1 (2y'z' + y^2 + z^2 - z \sin x) dx$; $y(0) = 0$; $z(0) = 1$;
 $y(1) = 1$; $z(1) = 0$;
- в) $J(y) = \int_0^1 (2y'z' + y^2 - z^2 + 2z \cos x) dx$; $y(0) = 0$; $z(0) = 1$;
 $y(1) = 1$; $z(1) = 0$;

Датчик синтезованої функції-зразка (наприклад, відліків імпульсної реакції $f(kT)$, якщо мова йде про використання середньоквадратичного критерію за формулою (9.19)), посилає сигнали на один вхід пристрою порівняння (ПП), на інший його вхід надходить імпульсна реакція системи «канал плюс коректор» (а саме, каскадного з'єднання каналу і коректора (див. рис. 9.1)), для чого на вхід каналу подають випробний сигнал – короткий імпульс, що моделює δ -функцію. Пристрій порівняння обчислює похибку D^2 , на підставі чого керуючий пристрій (КП) варіює положення органів налаштування коректора, що фізично моделюють параметри g_0, g_1, \dots, g_n доти, поки похибка коректування не досягне заданого (малого) значення.

Як практично підійти до питання вибору конкретного алгоритму налаштування поліномного коректора, тобто послідовності змін параметрів $g_k, k = 0, 1, \dots, n$, що приводить до мінімуму похибки коректування? Насамперед відзначимо, що, якщо передбачається «ручне» налаштування коректора, то бажано застосовувати просте правило для визначення черговості і кутів поворотів регуляторів (потенціометрів), що моделюють параметри g_0, g_1, \dots, g_n , а також контролювати похибку, наприклад, за стрілочним індикатором. При автоматичному налаштуванні коректора необхідно вилучити явища типу «зациклення», коли процес пошуку екстремуму переходить у блукання біля деякої точки, достатньо далекої від точки мінімуму похибки коректування. Зрозуміло, «зациклення» неприємне і при «ручному» налаштуванні, але в останньому випадку оператор може діагностувати його за занадто великою залишковою похибкою і зробити спробу повторного налаштування, трохи видозмінивши процедуру і вибравши інші початкові положення потенціометрів. З іншого боку, при автоматичному налаштуванні припустимі більш складні правила зміни параметрів g_0, g_1, \dots, g_n , у зв'язку з чим можуть бути реалізовані більш ефективні, тобто такі алгоритми, що швидше приводять до результату. Як при ручному, так і при автоматичному налаштуванні коректорів перевагу варто віддавати алгоритмам, що сходяться при довільно великій початковій похибці коректування.

Математичні методи мінімізації функцій багатьох змінних, що знайшли застосування, зокрема, при налаштуванні коректорів, розглянуті в попередніх розділах. Відзначимо, що згаданий матеріал істотно спирається на методи оптимізації функцій однієї змінної.

Серед поліномних коректорів найбільше поширення одержали такі, що мають ортонормовані, тобто ортогональні і нормовані, базисні функції. Умова ортонормованості має вигляд

$$\int_{-\Omega}^{\Omega} (\psi_k(\omega) \overline{\psi_l(\omega)}) d\omega = \begin{cases} 1, & k=l; \\ 0, & k \neq l. \end{cases}$$

Найбільш відомим прикладом коректорів цього типу, що одержали назву ортогональних, є гармонійний. Базисні функції гармонійного коректора (ГК) мають вид

$$\psi_k(\omega) = \sqrt{T/2\pi} e^{-i\omega kT}, \quad k=1, 0, \dots, n, \quad T = \pi/\Omega, \quad -\Omega < \omega < \Omega \quad (9.20)$$

і є періодичними функціями частоти (з періодом 2Ω).

9.4.3 Налаштування поліномних коректорів. Гармонійний коректор

Згідно з теоремою Релея-Парсеваля

$$D^2 = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |F(\omega) - \sum_{k=0}^n g_k \Phi_k(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |f(t) - \sum_{k=0}^n g_k \phi_k(t)|^2 dt, \quad (9.18)$$

де функції $f(t)$ і $\phi_k(t)$ визначаються виразами (9.10) і (9.11).

Більш того, якщо найвища частота в спектрі функцій $f(t)$ і $\phi_k(t)$ при $k = 0, 1, \dots, n$, не перевищує Ω , то похибка D^2 , як неважко показати спираючись на теорему Котельникова [6], може бути знайдена за відліковим значенням цих функцій, узятих з частотою, що удвічі перевищує $\Omega/2\pi$:

$$D^2 = \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} |f(kT) - \sum_{m=0}^n g_m \phi_m(kT)|^2, \quad T = \pi/\Omega. \quad (9.19)$$

Для фізичної реалізації процесу налаштування, наприклад, за критерієм D^2 , необхідний спеціальний прилад, що розраховує цю похибку за формулами (9.19) або (9.18). Вихідні дані цього приладу використовують для все нових і нових змін параметрів g_0, g_1, \dots, g_n у такий спосіб, щоб похибка коректування послідовно зменшувалася. Структурна схема всього комплексу, що складається безпосередньо з поліномного коректора і пристроєм для налаштування показана на рис. 9.6

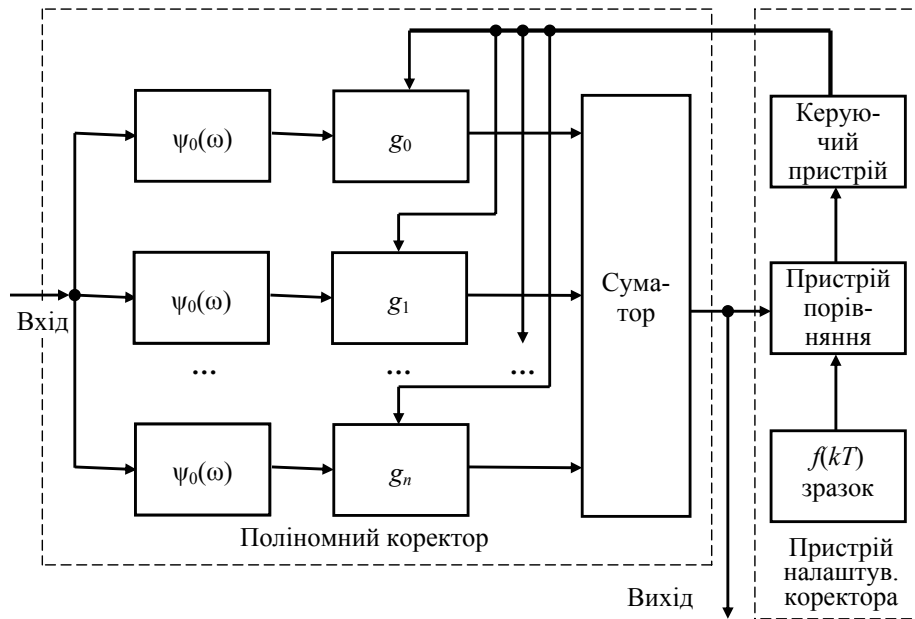


Рисунок 9.6 – Структурна схема поліномного коректора з пристроєм для налаштування

$$\begin{aligned} \text{г) } J(y) &= \int_0^1 (y^2 + z^2 + 2y'z' + ye^x) dx; & y(0) &= 0; & z(0) &= 1; \\ & & y(1) &= 1; & z(1) &= 0; \\ \text{д) } J(y) &= \int_0^1 (y^2 + 4yz + z^2 + y'^2 + z'^2 + 2ze^x) dx; & y(0) &= 0; & z(0) &= 1; \\ & & y(1) &= 1; & z(1) &= 0. \end{aligned}$$

Література до розділу 6

- 1 Михлин С.Г. Вариационные методы в математической физике / Михлин С.Г. – М.: Наука, 1970. – 512 с.
- 2 Захарченко Н.В. Выбор узлов аппаратуры передачи данных: учеб. пособие / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, – 1985. – Ч. 1. – 89 с.
- 3 Уайлд Д.Дж. Методы поиска экстремума / Уайлд Д.Дж. – М.: Наука, 1966. – 276 с.
- 4 Гельфанд И.М. Вариационное исчисление / И.М. Гельфанд, С.В. Фомин – М.: ГИФ-МЛ, 1961. – 228 с.
- 6 Левин Б.Р. Теоретические основы статистической радиотехники / Левин Б.Р. – М.: Сов. Радио, 1974. – 728 с.
- 7 Захарченко Н.В. Оптимизация и моделирование систем связи: Учеб. пособие. / Н.В. Захарченко, П.Я. Нудельман. – Одесса: ОЭИС им. А.С. Попова, 1988. – Ч. 1. – 86 с.

7 ОСНОВИ ТЕОРІЇ ГРАФІВ

7.1 Основні поняття теорії графів

Теорія графів є розділом математики, який має широке практичне застосування. Багато проблем, які виникають у таких дуже різних сферах знань, як психологія, управління, планування, хімія, електроніка, торгівля, освіта та інших, можуть бути сформульовані як задачі теорії графів. З огляду на це теорія графів є цікавою не лише сама по собі, але також і тим, що становить загальну основу, на якій результати, отримані в різних сферах знань, можуть бути зібрані, узагальнені та розповсюджені.

На відміну від інших наукових дисциплін теорія графів має цілком певну визначену дату народження. Перша робота з теорії графів, написана математиком Леонардом Ейлером, була опублікована у 1736 р. в працях Академії наук в Санкт-Петербурзі. Ним була сформульована й розв'язана задача про Кенігсберзькі мости [1].

До 50-х років нашого століття в теорії графів склались два суттєво різних напрямки: алгебричний та оптимізаційний. Останній набув широкого розвитку завдяки появі ЕОМ та у зв'язку з розробкою методів лінійного програмування. Розглянемо оптимізаційний напрямок теорії графів.

Граф G задається множиною точок або вершин x_1, x_2, \dots, x_n (яку позначимо X) та множиною ліній або ребер (гілок) a_1, a_2, \dots, a_m (яку позначимо A), що з'єднують між собою усі або частину цих точок. Таким чином, граф G повністю задається парою (X, A) [1], тобто $G = (X, A)$. Замість позначень X та A можна використати й інші.

Якщо ребра з множини A орієнтовані, що позначається стрілкою, то граф називається *орієнтованим графом* або *орграфом*. Ребра орграфа звичайно називають дугами. Орграф показано на рис. 7.1.

Якщо ребра не мають орієнтації, то граф називається *неорієнтованим* (рис.7.2). Граф, що має як орієнтовані, так і неорієнтовані ребра, називається *змішаним* (рис. 7.3). Якщо в графі між двома якими-небудь вершинами є декілька ребер, то такий граф називається *мультграфом* (рис. 7.4).

Звичайно вершинам графа приписуються номери (або позначення x_i, x_j), щоб розрізнити їх і мати можливість ними оперувати в процесі аналізу. Ребра графа позначаються $a = (x_i, x_j)$, де x_i, x_j – початкова та кінцева вершини, пов'язані даним ребром. Крім цього позначення застосовують також $a = (x_i, x_j)$ або $a(i, j)$, або a_{ij} . Іноді ребро (дугу) позначають (i, j) .

У використаних позначеннях i, j – номери вершин. Крім того, ребра можна позначати a_i , де i – номер ребра.

Шляхом μ_{ij} графа називається послідовність ребер, яка починається у вершині i й закінчується у вершині j . В цій послідовності кінцева вершина кожного ребра, відмінного від останнього, є початковою вершиною наступного.

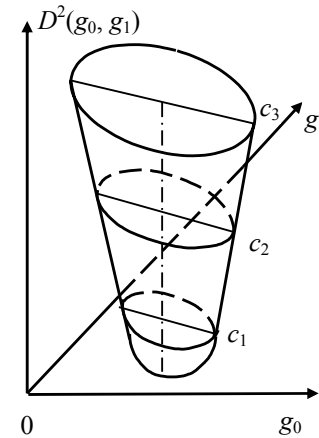


Рисунок 9.4 – Двовимірна параболічна поверхня

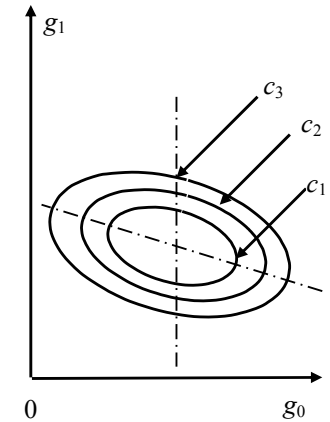


Рисунок 9.5 – Лінії рівня поверхні

Рівняння так званих ліній рівня цієї поверхні, а саме, ліній в області визначення (площини g_0g_1), на яких поверхня приймає фіксовані значення, має вид

$$D^2(g_0, g_1) = d - 2b_0g_0 - 2b_1g_1 + 2r_{01}g_0g_1 + r_{00}g_0^2 + r_{11}g_1^2 = c. \quad (9.17)$$

Можна довести, що дане рівняння в загальному випадку відповідає еліпсу (див. рис. 9.5, на якому показані лінії рівня поверхні (9.16), що зображена на рис. 9.4, при $c = c_1, c_2, c_3$, причому $c_1 < c_2 < c_3$).

Таким чином, задача корекції спотворень каналу зв'язку за допомогою поліномного коректора звелася до математичної задачі оптимізації (відшукування мінімуму) $(n + 1)$ -мірної параболічної поверхні.

Сам процес відшукування мінімуму похибки E (зокрема, D^2) при варіації параметрів g_0, g_1, \dots, g_n називають розрахунком, а у випадку апаратурної мінімізації – налаштуванням поліномного коректора. Важливо відзначити, що завдяки поданим вище міркуванням технічна задача з галузі зв'язку стала задачею чисто математичною.

Такі математичні задачі вирішуються методами пошуку екстремумів функцій багатьох змінних.

У наступному підрозділі розглянемо важливе з практичної точки зору питання апаратурного визначення величини $D^2(g_0, g_1, \dots, g_n)$. Це питання вирішується досить просто при налаштуванні за середньоквадратичним критерієм, тому що в цьому випадку похибку D^2 можна контролювати у часовій області.

Зі сказаного виходить, що $\text{Im}[F(\omega)\overline{\Phi_k(\omega)}]$ є непарною функцією частоти. Тому інтеграл від цієї функції в симетричних межах (другий доданок у правій частині рівності (9.9)) дорівнює нулю, звідки виходить справедливості співвідношення (9.8).

Спираючись на (9.8), введемо нові позначення

$$b_k = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega)\overline{\Phi_k(\omega)} d\omega, \quad k = 0, 1, \dots, n; \quad (9.12)$$

$$r_{kl} = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Phi_k(\omega) \cdot \overline{\Phi_l(\omega)} d\omega, \quad k, l = 0, 1, \dots, n. \quad (9.13)$$

З урахуванням (9.7) остаточно запишемо

$$D^2 = d - 2 \sum_{k=0}^n b_k g_k + \sum_{k=0}^n \sum_{l=0}^n r_{kl} g_k g_l. \quad (9.14)$$

Отже, середньоквадратична похибка D^2 надана в явному виді як функція $(n+1)$ -ї змінної:

$$D^2 = D^2(g_0, g_1, \dots, g_n). \quad (9.15)$$

Ця залежність є функцією другого степеня – $(n+1)$ -мірною параболою (вона невід’ємна, тому що є площею під невід’ємною кривою

$$\frac{1}{2\pi} \left| F(\omega) - \sum_{k=0}^n g_k \Phi_k(\omega) \right|^2,$$

і, як легко перевірити, необмежено зростає при прямуванні до нескінченності кожного з коефіцієнтів g_0, g_1, \dots, g_n ...

Приклад.

Для $n = 1$ $D^2(g_0, g_1)$ є двовимірною параболічною поверхнею, вид якої показаний на рис. 9.4:

$$\begin{aligned} D^2(g_0, g_1) &= d - 2b_0g_0 - 2b_1g_1 + r_{00}g_0^2 + r_{01}g_0g_1 + r_{10}g_1g_0 + r_{11}g_1^2 = \\ &= d - 2b_0g_0 - 2b_1g_1 + 2r_{01}g_0g_1 + r_{00}g_0^2 + r_{11}g_1^2. \end{aligned} \quad (9.16)$$

При виведенні останньої формули було прийнято до уваги, що

$$r_{01} = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Phi_0(\omega)\overline{\Phi_1(\omega)} d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Phi_1(\omega)\overline{\Phi_0(\omega)} d\omega = r_{10},$$

тому що функції $\Phi_0(\omega)\overline{\Phi_1(\omega)}$ і $\Phi_1(\omega)\overline{\Phi_0(\omega)} = \overline{\Phi_0(\omega)\overline{\Phi_1(\omega)}}$ мають однакові дійсні частини, а уявні частини відрізняються лише знаком, до того ж інтеграл у симетричних межах від уявної частини в силу її непарності дорівнює нулю.

На рис. 7.5 послідовність дуг $a_6, a_5, a_9, a_8, a_4, a_1, a_6, a_5, a_9$ є шляхи, які можна також записати у формі:

$$a(x_2, x_5), a(x_5, x_4), a(x_4, x_3), a(x_3, x_5), a(x_5, x_6);$$

$$a(x_1, x_2), a(x_2, x_5), a(x_5, x_4), a(x_4, x_3).$$

Більш простою є така форма запису:

$$a_{25}, a_{54}, a_{43}, a_{35}, a_{56};$$

$$a_{12}, a_{25}, a_{54}, a_{43}.$$

Шлях називається *простим*, чи *самонеперерізним*, якщо в ньому не повторюються ні вершини, ні ребра. У розглянутому прикладі другий шлях є простим.

Цикл – це шлях, початкова та кінцева вершини якого збігаються.

Граф називається *сполучним*, якщо для будь-яких різних вершин існує, принаймні, один з’єднуючий їх шлях.

В оргграфі шлях також називають *ланцюгом*.

Нагадаємо, що граф – це множина вершин та ребер, позначається $G = (X, A)$. Будь-яка підмножина множини G називається *підграфом* і позначається $\bar{G} = (\bar{X}, \bar{A})$.

Дерево графа визначається як сполучена підмножина \bar{G} , що не має циклів. Отже, для будь-яких двох вершин графа в дереві існує єдиний шлях, який з’єднує їх.

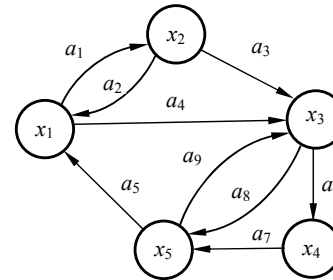


Рисунок 7.1 – Орієнтований граф

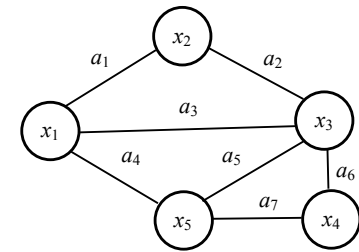


Рисунок 7.2 – Неорієнтований граф

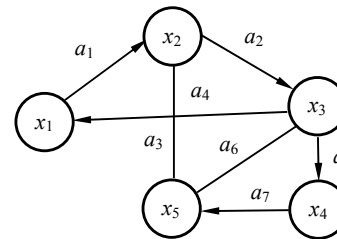


Рисунок 7.3 – Змішаний граф

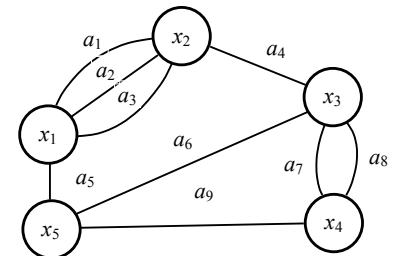


Рисунок 7.4 – Мультиграф

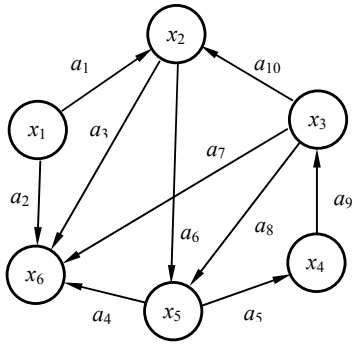


Рисунок 7.5 – Орієнтований граф (до пояснення елементів)

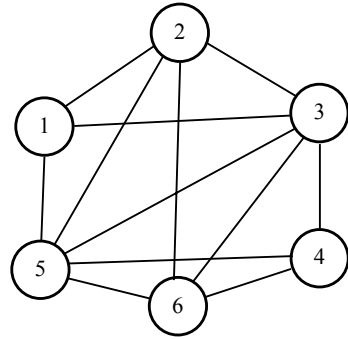


Рисунок 7.6 – Неорієнтований граф (до пояснення елементів)

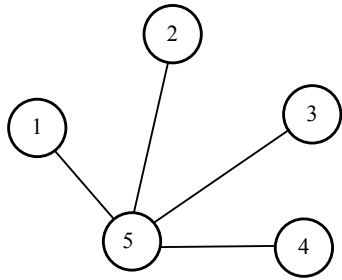


Рисунок 7.7 – Дерево графа

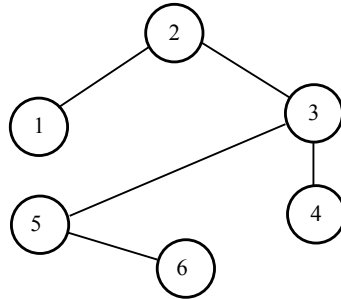


Рисунок 7.8 – Основне дерево графа

У графі, який містить n вершин, підграф із k вершин ($k \leq n$) є деревом, якщо виконані будь-які дві з таких умов:

- підграф є зв'язним;
- підграф не має циклів;
- число ребер у підграфі дорівнює $k - 1$.

Остовним деревом, або *остовом*, називається дерево, що містить всі вершини графа. Отже, якщо граф містить n вершин, то дерево з n вершинами та $n - 1$ ребром є остовом. Дерево для графа (рис. 7.11) показано на рис. 7.7, остовне дерево – на рис. 7.8. Тут для спрощення вершини графа позначимо номерами, а гілки лишимо не позначеними, їх будемо позначати за необхідністю. Зазначимо, що для графа можна побудувати декілька остовних дерев.

У подальшому для спрощення, якщо про це не домовлено окремо, остовне дерево будемо називати просто деревом.

$$\begin{aligned}
 &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} [F(\omega) \cdot \overline{F}(\omega) - F(\omega) \sum_{k=0}^n g_k \overline{\Phi}_k(\omega) - \overline{F}(\omega) \sum_{k=0}^n g_k \Phi_k(\omega) + \\
 &\quad + \sum_{k=0}^n g_k \Phi_k(\omega) \sum_{l=0}^n g_l \overline{\Phi}_l(\omega)] d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \{|F(\omega)|^2 - \\
 &\quad + \sum_{k=0}^n g_k [F(\omega) \overline{\Phi}_k(\omega) + \overline{F}(\omega) \Phi_k(\omega)] + \sum_{k=0}^n \sum_{l=0}^n g_k g_l \Phi_k(\omega) \overline{\Phi}_l(\omega)\} d\omega.
 \end{aligned}$$

Скористаємося тим, що операції інтегрування і підсумовування є перестановочні. Після перетворення останнього виразу одержимо

$$\begin{aligned}
 D^2 &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |F(\omega)|^2 d\omega - \sum_{k=0}^n g_k \frac{1}{2\pi} \int_{-\Omega}^{\Omega} 2\text{Re}[F(\omega) \overline{\Phi}_k(\omega)] d\omega + \\
 &\quad + \sum_{k=0}^n \sum_{l=0}^n g_k g_l \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Phi_k(\omega) \overline{\Phi}_l(\omega) d\omega.
 \end{aligned} \tag{9.6}$$

Для подальшого виявиться корисним ввести позначення для першого доданка в (9.6)

$$d = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |F(\omega)|^2 d\omega \tag{9.7}$$

Для перетворення другого доданка в (9.6) скористаємося тим, що

$$\int_{-\Omega}^{\Omega} 2\text{Re}[F(\omega) \overline{\Phi}_k(\omega)] d\omega = \int_{-\Omega}^{\Omega} F(\omega) \overline{\Phi}_k(\omega) d\omega. \tag{9.8}$$

Справедливість останньої рівності обґрунтовується наступним. Відомо, що спектри дійсних сигналів мають парні дійсні і непарні уявні частини. Але інтеграл, що стоїть в правій частині рівності (9.8), надамо у виді

$$\int_{-\Omega}^{\Omega} F(\omega) \overline{\Phi}_k(\omega) d\omega = \int_{-\Omega}^{\Omega} \text{Re}[F(\omega) \overline{\Phi}_k(\omega)] d\omega + i \int_{-\Omega}^{\Omega} \text{Im}[F(\omega) \overline{\Phi}_k(\omega)] d\omega. \tag{9.9}$$

Вираз $F(\omega) \overline{\Phi}_k(\omega)$ можна розглядати як передатну функцію ланки у виді каскадного з'єднання двох схем з дійсними імпульсними реакціями:

$$f(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) e^{i\omega t} d\omega \tag{9.10}$$

– бажаної («ідеальної») імпульсної реакції відкоректованого каналу і $\Phi_k(-t)$ – інвертованої за часом імпульсної реакції ланки з передатною функцією $\Phi_k(\omega)$, тобто

$$\Phi_k(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \Phi_k(\omega) e^{i\omega t} d\omega. \tag{9.11}$$

Функції $f(t)$ і $\Phi_k(-t)$ є дійсними, звідки виходить дійсність імпульсної реакції каскадного з'єднання ланок з передатними функціями $F(\omega)$ і $\overline{\Phi}_k(\omega)$.

фізичне моделювання виразу (9.4) дозволяє спостерігати величину E в виді показань, наприклад, стрілочного індикатора.

Надалі будемо розглядати лише середньоквадратичний критерій точності коректування (9.4), поклавши $W(\omega) = 1$, $\omega_n \leq \omega \leq \omega_b$, де $\omega_n = 0$, $\omega_b = \Omega < \infty$.

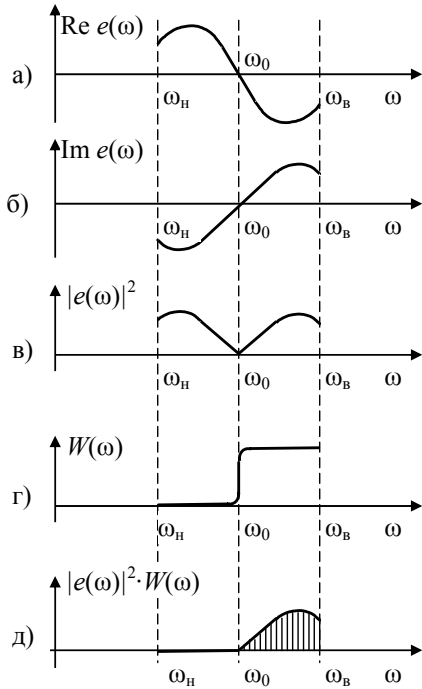


Рисунок 9.3 – Ілюстрація розуміння середньоквадратичного критерію точності корекції каналу зв'язку

Крім того, традиційно, похибку $E[e(\omega)]$ будемо оцінювати на симетричному інтервалі $(-\Omega, \Omega)^*$, поділивши його на 2π , що можливо оскільки $E[e(\omega)]$ приймає однакові значення на інтервалах $(-\Omega, 0)$ і $(0, \Omega)$. Для такого випадку (згадаємо вираз для похибки $e(\omega)$) величину (9.4) позначимо повному:

$$D^2 = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \left| F(\omega) - \sum_{k=0}^n g_k \Phi_k(\omega) \right|^2 d\omega \quad (9.5)$$

Виконаємо деякі перетворення виразу (9.5). Позначимо рискою операцію переходу від числа $a + ib$ до комплексно-сполученого числа $a - ib$ та знайдемо

$$D^2 = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \left[F(\omega) - \sum_{k=0}^n g_k \Phi_k(\omega) \right] \cdot \left[\bar{F}(\omega) - \sum_{l=0}^n \bar{g}_l \bar{\Phi}_l(\omega) \right] d\omega =$$

Якщо для позначення деяких природних обмежень, що накладаються на ребра (дуги) графа, зіставити числа, що називаються вагою ребер, або їх кількісні характеристики, що відповідають, наприклад, відстані, вартості, часу або іншим адитивним величинам, то можна ввести поняття мінімального (чи максимального) дерева. Вага дерева визначається як сума ваг його ребер (звідси й вимога до адитивності характеристик ребер графа).

Найкоротшим (максимальним) деревом називається дерево з мінімальною (максимальною) вагою. Іноді таке дерево називають деревом мінімальної (максимальної) вартості.

Дерева та остови визначаються як для орієнтованих, так і для неорієнтованих графів.

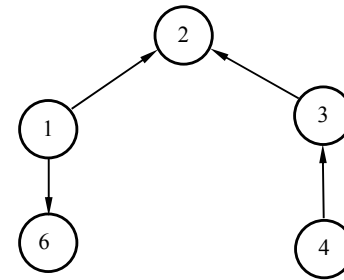


Рисунок 7.9 – Дерево графа, позначеного на рис.7.5

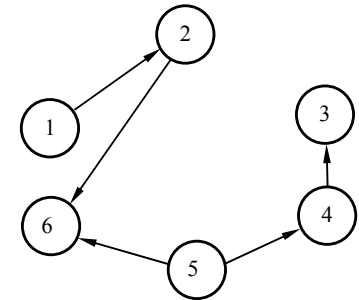


Рисунок 7.10 – Остов графа, позначений на рис.7.5

На рис. 7.9 показано дерево графа, показаного на рис. 7.5, а на рис. 7.10 – остов цього графа.

Перерізом графа називається сукупність ребер, при видаленні яких граф розпадається на два незв'язних підграфи.

Переріз σ_{st} , який розділяє вершини s і t , є множиною ребер (A, \bar{A}) , де $s \in A$, $t \in \bar{A}$. Число $c(A, \bar{A})$ називається пропускнуою здатністю (величиною) перерізу. Пропускна здатність перерізу визначається як сума пропускну здатностей ребер, які складають переріз. На графах перерізи звичайно позначаються лініями, що перерізають ребра, які складають переріз.

На рис. 7.11 переріз σ_1 включає ребра $(1, 2)$ і $(1, 6)$ та поділяє граф на два підграфи, в одному з яких вершина 1, в другому – вся решта вершин та ребер. Переріз σ_2 (ребра $(2, 3)$, $(6, 3)$, $(6, 5)$) поділяє граф на два підграфи, в одному з яких вершини 1, 2, 6 та ребра $(1, 2)$, $(1, 6)$, в другому – вершини 3, 4, 5 та ребра $(3, 4)$, $(3, 5)$ і $(4, 5)$. Переріз σ_3 поділяє граф на два підграфи, в одному з яких вершини 1, 2, 5, 6 і ребра $(1, 2)$, $(1, 6)$, $(6, 5)$, в другому – вершини 3 і 4 та ребр $(3, 4)$.

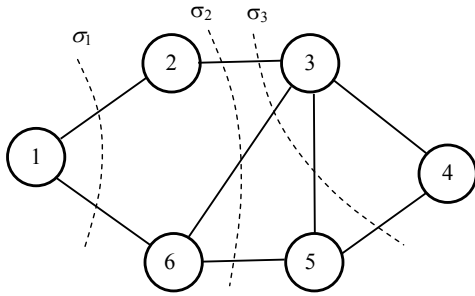


Рисунок 7.11 – Перерізи графа

7.2 Модель мережі зв'язку

Структура мережі зв'язку, її топологія – це сукупність пунктів (вузлів, станцій тощо) і ліній (каналів), які їх з'єднують.

Для математичного опису структури мережі зв'язку зручно користуватися мережною моделлю. Як модель використовується граф $G = (X, A)$, де $X = \{x_i\}$ – сукупність вершин графа, які ставляться у відповідність пунктам мережі, а $A = \{a_{ij}\}$ – сукупність ребер графа, які ставляться у відповідність лініям зв'язку. Відповідно до того, що канали мереж можуть бути однобічними та двобічними, ребра графа можуть бути орієнтованими та неорієнтованими. Таким чином, як модель мережі зв'язку можуть бути використані орієнтовані і неорієнтовані, змішані графи, а також мультиграфи.

Мережні моделі широко використовуються на практиці при проектуванні систем електрозв'язку, систем космічного та радіозв'язку, телетрансляційних мереж, обчислювальних комплексів, транспортних мереж тощо.

Мережний аналіз відіграє все більш зростаючу роль, тому що за допомогою мереж (графів) можна досить просто побудувати модель системи. Розширення сфери використання мереж пов'язане з тим, що методи мережного аналізу дозволяють [2]:

- побудувати модель складної системи як сукупність простих її елементів;
- скласти формальні процедури для визначення якісних та кількісних характеристик системи;
- показати механізм взаємодії компонентів системи з метою опису останньої в термінах її основних характеристик;
- визначити, які дані необхідні для дослідження системи.

Основна перевага мережного підходу полягає в тому, що він може бути успішно застосований до розв'язання практично будь-якої задачі, коли дослідник має необхідні знання та здібність точно побудувати мережну модель.

В усіх практичних випадках зробити похибку $e(\omega)$ рівною нулю не вдається, та у цьому і немає необхідності.

Замість похибки $e(\omega)$ звичайно зводять до мінімуму або припустимої величини деяке число E , що є «прийнятною» характеристикою $e(\omega)$, точніше, мірою того, наскільки функція $e(\omega)$ відхиляється від нуля

$$E = E[e(\omega)]. \quad (9.3)$$

Введення числової характеристики для оцінки «рівня» у загальному випадку комплексної функції $e(\omega)$ є досить, універсальним для різних прикладних задач спрощеним підходом.

9.4.2 Середньоквадратична похибка коректування

Найбільше часто на практиці

$$E[e(\omega)] = \int_{\omega_n}^{\omega_b} |e(\omega)|^2 W(\omega) d\omega, \quad (9.4)$$

де $W(\omega)$ – дійсна невід'ємна обмежена функція, названа вагою, котру в багатьох випадках вважають постійною, а ω_n і ω_b як і раніше рівні нижній і верхній частотам смуги пропускання каналу зв'язку. Вираз (9.4) називають середньоквадратичною похибкою корекції каналу.

Рис. 9.3 пояснює розуміння середньоквадратичного критерію точності корекції каналу зв'язку (9.4). Нехай $e(\omega) = \text{Re } e(\omega) + i \text{Im } e(\omega)$ (див. рис. 9.3,а,б).

Тоді $|e(\omega)|^2 = [\text{Re } e(\omega)]^2 + [\text{Im } e(\omega)]^2$ (рис. 9.3,в). Заштрихована на рис. 9.3,д площа саме дорівнює величині $E[e(\omega)]$. У даному випадку вагова функція (рис. 9.3,г) «підкреслює» роль високих частот, зневажаючи відмінність реальної передатної функції від ідеальної в області низьких частот можна припускати, що розроблювач коректора, що задав вагову функцію $W(\omega)$ рівну нулю лівіше частоти ω_0 , впевнений у відсутності в переданому сигналі енергії в області частот (ω_n, ω_0) , у зв'язку з чим вагова функція рис. 9.3,г «зосереджує» здатність коректора, що виправляє, винятково в діапазоні частот (ω_0, ω_b) .

Звернемо увагу на ту обставину, що, якщо $e(\omega) = 0$ при $\omega_n \leq \omega \leq \omega_b$, то вираз (9.4) дорівнює нулю, якщо ж функція $|e(\omega)|^2 W(\omega)$ більше нуля в довільному кінцевому діапазоні частот смуги (ω_n, ω_b) , то число $E[e(\omega)]$ виявляється більшим нуля. Таким чином, число $E[e(\omega)]$ за формулою (9.4) дійсно можна розглядати як деякий показник якості корекції каналу зв'язку.

Знаходять застосування також середня абсолютна похибка

$$E = \int_{\omega_n}^{\omega_b} |e(\omega)| W(\omega) d\omega$$

і максимальна абсолютна (чебишевська) похибка

$$E = \max_{\omega_n \leq \omega \leq \omega_b} W(\omega) |e(\omega)|.$$

Введення числової оцінки похибки синтезу E досить зручно з практичної точки зору. У той час як контроль і, тим більше, візуальне спостереження комплексної функції $e(\omega)$ у всьому інтервалі (ω_n, ω_b) викликає великі труднощі,

$$e(\omega) = F(\omega) - F_p(\omega) = F(\omega) - H(\omega) G_p(\omega),$$

є комплексною функцією частоти ω .

Широке застосування знаходять поліномні коректори з передатними функціями у виді узагальненого багаточлена (полінома)

$$G_p(\omega) = \sum_{k=0}^n g_k \psi_k(\omega), \quad (9.2)$$

де $\psi_k(\omega)$, $k = 0, 1, \dots, n$ – заздалегідь реалізовані (так звані базисні) функції, а g_k , $k = 0, 1, \dots, n$ – варійовані параметри (дійсні числа), зміною яких здійснюється синтез необхідної передатної функції коректора [4, 5, 6]. Формулу (9.2) відповідає структурна схема, що показана на рис. 9.2.

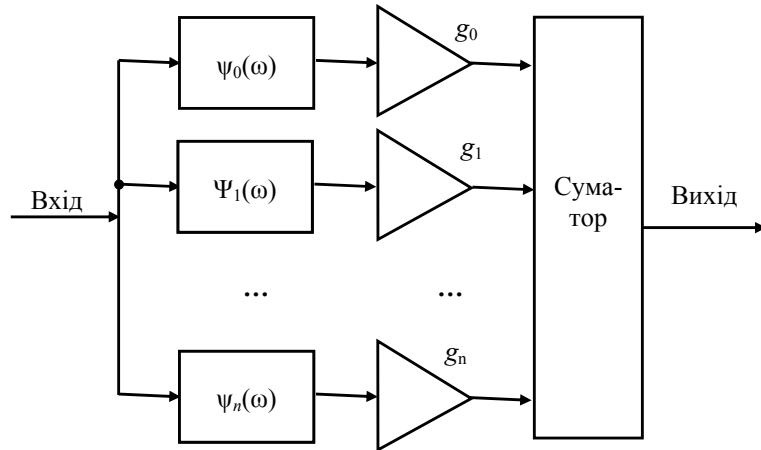


Рисунок 9.2 – Структурна схема поліномного коректора

При корекції частотних характеристик каналів за допомогою поліномного коректора вираз для похибки синтезу необхідної передатної функції каналу приймає вид

$$e(\omega) = F(\omega) - H(\omega) G_p(\omega) = F(\omega) - H(\omega) \sum_{k=0}^n g_k \psi_k(\omega) = F(\omega) - \sum_{k=0}^n g_k \Phi_k(\omega),$$

де $\Phi_k(\omega) = \psi_k(\omega) H(\omega)$, $k = 0, 1, \dots, n$ – базисні функції «еквівалентного» поліномного синтезатора, що формує (синтезує) передатну функцію відкоректованого каналу зв'язку.

Розглянемо докладніше синтез необхідної частотної характеристики поліномним синтезатором з передатною функцією

$$\sum_{k=0}^n g_k \Phi_k(\omega).$$

7.3 Матричне подання графів

Для алгебричного завдання графів зручно користуватися топологічними матрицями та матрицями характеристик ребер графа (гілок мережі).

Топологічні матриці. Матриця суміжності. Матриця суміжності (сполучення) графа G – це квадратна матриця $A = [a_{ij}]$ розміру n (n – число вершин графа). Визначається таким чином:

$$a_{ij} = \begin{cases} 1, & \text{якщо } G \in \text{ребро } (i, j); \\ 0, & \text{якщо } G \text{ немає ребра } (i, j). \end{cases}$$

Елементи головної діагоналі матриці A звичайно приймають рівними нулю – $a_{ij} = 0$, – за винятком випадків, коли в деяких вершинах є "петлі".

Матриця суміжності графа, що показана на рис. 7.12, має вид

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \quad (7.1)$$

У вершинах 2 і 6 є "петлі", тому відповідні елементи головної діагоналі $a_{22} = 1$ і $a_{66} = 1$. В подальшому будемо розглядати графи, які не мають "петель" у вершинах. Відзначимо, що матриця A для орієнтованого графа несиметрична відносно головної діагоналі, симетричною вона буде для неорієнтованого графа.

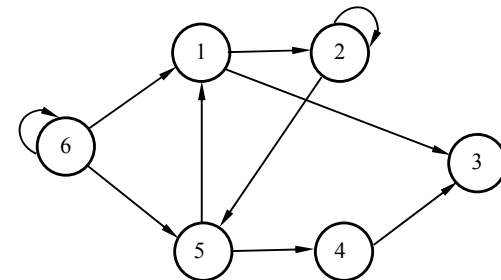


Рисунок 7.12 – Граф мережі (до (7.1))

Структурна матриця. Для спрощення запису структури мережі ребрам графа присвоюються позначення, наприклад $a, b, c \dots$. Ці позначення використовуються як елементи структурної матриці.

Структурна матриця графа G – квадратна матриця $B = [b_{ij}]$ розміру n .

Визначається так:

$$b_{ij} = \begin{cases} 1, & \text{якщо } i = j; \\ 0, & \text{якщо } G \text{ немає ребра } (i, j); \\ x, & \text{якщо ребро } (i, j) \text{ в } G \text{ є, } i < j; \\ x, & \text{якщо ребро } (i, j) \text{ в } G \text{ є, } i > j, \end{cases}$$

де x – позначення ребра на графі.

Для графа, показаного на рис. 7.13, структурна матриця B має вид

$$B = \begin{bmatrix} 1 & a & 0 & h & e \\ \bar{a} & 1 & b & 0 & g \\ 0 & \bar{b} & 1 & c & f \\ 0 & 0 & \bar{c} & 1 & d \\ \bar{e} & 0 & \bar{f} & \bar{d} & 1 \end{bmatrix}. \quad (7.2)$$

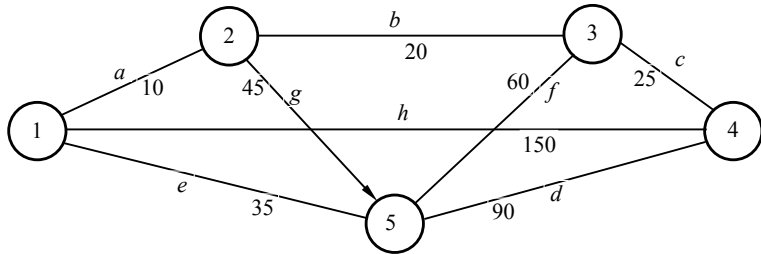


Рисунок 7.13 – Змішаний граф (для пояснення структурної матриці B)

Крім розглянутих топологічних матриць, можуть бути використані матриці інцидентій "вершини-дуги", "дуги-дуги" [1, 2, 3].

Матриці кількісних характеристик ребер графа. Для різних кількісних оцінок кожному ребру графа приписується деяка вага – число, яке характеризує яку-небудь властивість даного ребра, наприклад, довжину, вартість, пропускну здатність, каналну ємність, час передачі інформації, надійність тощо.

Зазначені характеристики ребер графа подаються у формі відповідних квадратних матриць розміру n – довжин, вартостей і т. ін. Для побудови, наприклад, матриці довжин $L = (L_{ij})$ користуються правилами

$$L_{ij} = \begin{cases} 1, & \text{якщо } i = j; \\ \infty, & \text{якщо } G \text{ немає ребра } (i, j); \\ \text{довжині ребра } (i, j), & \text{якщо ребро } (i, j) \text{ є в } G. \end{cases}$$

Вважаючи, що наведені біля ребер графа числа (рис. 7.13) є довжинами ребер, матриця довжин L має вид

Таблиця 9.1 – Результати обчислень за методом найшвидшого спуску

k	$x_1^{(k)}$	$x_2^{(k)}$	$f(x^{(k)})$
1	-1,2	2,16	24,480
2	-0,6086	1,2192	7,4227
3	-0,3364	0,6728	2,4227
4	-0,1857	0,3714	0,6897

9.4 Корекція лінійних спотворень каналів зв'язку

9.4.1 Постановка задачі корекції

Корекцію (виправлення, компенсацію) лінійних спотворень каналів зв'язку можна здійснити каскадним включенням у тракт передачі лінійних коригувальних пристроїв (коректорів), що забезпечують у заданій смузі частот $\omega_n \leq \omega \leq \omega_v$, виконання умови

$$H(\omega) G(\omega) = F(\omega) \quad (9.1)$$

де $H(\omega)$, $G(\omega)$, $F(\omega)$ – передатні функції відповідно каналу зв'язку зі спотворенням, коректора і «ідеально» відкоректованого тракту передачі (рис. 9.1) [3]. Функція $F(\omega)$ вважається заданою

$$F(\omega) = A e^{-i\omega\tau_0}, \quad A = \text{const}, \quad \tau_0 = \text{const}.$$

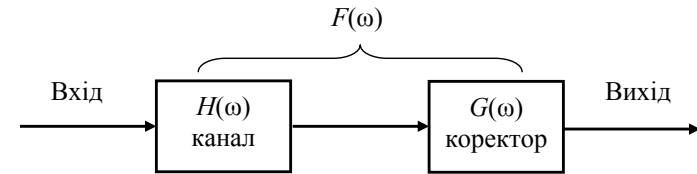


Рисунок 9.1 – Включення у тракт передачі коректора

Якщо передатна функція $H(\omega)$ заздалегідь відома, коректор може бути розрахований і синтезований у виді пристрою з фіксованими параметрами, але в багатьох практичних задачах функція $H(\omega)$ заздалегідь не визначена (наприклад, на мережах зв'язку, що комутуються), а іноді може мінятися під час передачі (як, наприклад, у багатопроменевих радіоканалах). Тоді задача корекції зводиться до синтезу лінійного чотириполюсника, що переналаштовується, з передатною функцією $G(\omega)$ такою, що в кожному конкретному випадку виконувалася умова (9.1). Необхідна передатна функція коректора $G(\omega)$ завжди реалізується з деякою похибкою; позначаючи через $G_p(\omega)$ і $F_p(\omega)$ передатні функції технічно реалізованого коректора і тракту «канал плюс коректор» відповідно

$$F_p(\omega) = H(\omega) G_p(\omega).$$

Вираз для похибки коректування

Умова досягнення точки мінімуму з заданою похибкою ϵ не виконана, пошук має бути продовжено.

Ітерація 2. Знайдемо координати другої точки за формулою (5.8):

$$\begin{aligned}x_1^{(2)} &= x_1^{(1)} - \alpha^{(1)} \frac{\partial f}{\partial x_1} = x_1^{(1)} - \alpha^{(1)} (16x_1^{(1)} + 4x_2^{(1)}) = \\ &= -1,2 - 0,056[16(-1,2) + 4 \cdot 2,16] = -0,6086; \\ x_2^{(2)} &= x_2^{(1)} - \alpha^{(1)} \frac{\partial f}{\partial x_2} = x_2^{(1)} - \alpha^{(1)} (4x_1^{(1)} + 10x_2^{(1)}) = \\ &= 2,16 - 0,056[4(-1,2) + 10 \cdot 2,16] = 1,2192.\end{aligned}$$

Отже, $x^{(2)} = [-0,6086, 1,2192]^T$.

Розрахуємо значення функції $f(x)$ у другій точці $x^{(2)}$

$$\begin{aligned}f(x^{(2)}) &= 8x_1^2 + 4x_1x_2 + 5x_2^2 = \\ &= 8 \cdot (-0,6086)^2 + 4 \cdot (-0,6086) \cdot 1,2192 + 5 \cdot (1,2192)^2 = 7,4227.\end{aligned}$$

Умова досягнення точки мінімуму з заданою похибкою ϵ не виконана, пошук має бути продовжено.

Ітерація 3. Знайдемо координати третьої точки за формулою (5.8):

$$\begin{aligned}x_1^{(3)} &= x_1^{(2)} - \alpha^{(2)} \frac{\partial f}{\partial x_1} = x_1^{(2)} - \alpha^{(2)} (16x_1^{(2)} + 4x_2^{(2)}) = \\ &= -0,6086 - 0,056[16(-0,6086) + 4 \cdot 1,2192] = -0,3364; \\ x_2^{(3)} &= x_2^{(2)} - \alpha^{(2)} \frac{\partial f}{\partial x_2} = x_2^{(2)} - \alpha^{(2)} (4x_1^{(2)} + 10x_2^{(2)}) = \\ &= 1,2192 - 0,056[4(-0,6086) + 10 \cdot 1,2192] = 0,6728.\end{aligned}$$

Отже, $x^{(3)} = [-0,3364, 0,6728]^T$.

Розрахуємо значення функції $f(x)$ у третій точці $x^{(3)}$

$$\begin{aligned}f(x^{(3)}) &= 8x_1^2 + 4x_1x_2 + 5x_2^2 = \\ &= 8 \cdot (-0,3364)^2 + 4 \cdot (-0,3364) \cdot 0,6728 + 5 \cdot (0,6728)^2 = 2,2633.\end{aligned}$$

Умова досягнення точки мінімуму з заданою похибкою ϵ не виконана, пошук має бути продовжено.

Ітерація 4. Далі знаходяться координати кожної наступної точки, доки не будуть виконані умови досягнення точки мінімуму з заданою похибкою ϵ , після чого пошук має бути закінчено.

Після 4-ї ітерації пошук закінчено, оскільки точка мінімуму досягнута з заданою похибкою пошуку, а саме $f(x^{(4)}) = 0,6897$, що менше $\epsilon = 0,7$.

Отже $x^{(4)} = [-0,1857, 0,3714]^T$.

Результати обчислень після 4-х проведених ітерацій за методом найшвидшого спуску заносяться в табл. 9.1

$$L = \begin{bmatrix} 0 & 10 & \infty & 150 & 35 \\ 10 & 0 & 20 & \infty & 45 \\ \infty & 20 & 0 & 25 & 60 \\ \infty & \infty & 25 & 0 & 90 \\ 35 & \infty & 60 & 90 & 0 \end{bmatrix}. \quad (7.3)$$

Матрицю канальних ємностей ребер $C = [c_{ij}]$ отримують за правилом:

$$c_{ij} = \begin{cases} \infty, & \text{якщо } i = j; \\ 0, & \text{якщо } G \text{ немає ребра } (i, j); \\ \text{числу каналів ребра } (i, j), & \text{якщо ребро } (i, j) \in G. \end{cases}$$

Вважаючи, що позначені біля ребер графа числа (рис. 7.13) є числами каналів ребер, відповідна матриця канальних ємностей C має вид

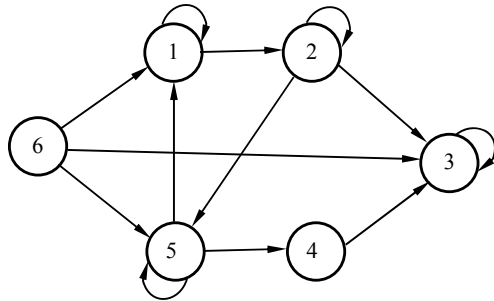
$$L = \begin{bmatrix} \infty & 10 & 0 & 150 & 35 \\ 10 & \infty & 20 & 0 & 45 \\ 0 & 20 & \infty & 25 & 60 \\ 0 & 0 & 25 & \infty & 90 \\ 35 & 0 & 60 & 90 & \infty \end{bmatrix}. \quad (7.4)$$

Аналогічно отримують інші матриці характеристик ребер графа. Якщо G – неорієнтований граф, то усі матриці симетричні відносно головної діагоналі, тобто їх елементи $x_{ij} = x_{ji}$.

Контрольні питання та задачі до розділу 7

1. Що таке граф і як він позначається?
2. Який граф називається орієтованим графом?
3. Як називаються ребра орграфа?
4. В якому випадку граф не є орієтованим?
5. Який граф називається змішаним?
6. Який граф називається мультиграфом?
7. Що таке шлях і яким він може бути?
8. Який шлях називається циклом?
9. Який граф називається сполучним?
10. Що таке підграф?
11. Який граф називається деревом?
12. Що називається остовним деревом?
13. Що таке вага дерева?
14. Яке дерево називається з мінімальною (максимальною) вагою?
15. Як визначаються дерева та остови для орієтованих і неорієтованих графів?
16. Що називається перерізом графа?
17. Що називається пропускною здатністю перерізу?

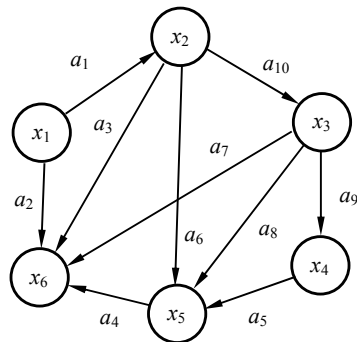
19. Що таке структура мережі зв'язку і як вона описується математично?
20. Дайте приклади моделей мережі зв'язку за допомогою графів?
21. Що таке матриця суміжності графа?
22. Для чого потрібно алгебричне завдання графів?
23. Для чого потрібна структурна матриця і як вона визначається?
24. Побудувати матрицю суміжності для графа



25. За матрицею суміжності A побудувати граф

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

26. Знайти дерево та остов для графа



27. Для вершин x_1 і x_5 графа із завдання 26 знайти варіанти простих шляхів.

Отже

$$\text{grad}^2 f(\bar{x}) = \begin{bmatrix} +4 & -10 \\ -10 & +2 \end{bmatrix}.$$

Матриця $\text{grad}^2 f(\bar{x})$ є невизначеною, оскільки другі похідні заданої функції двох змінних $f(x)$ приймають позитивні і від'ємні значення. Отже точка \bar{x} не є ні точкою мінімуму, ні точкою максимуму, а являє собою сідлову точку.

9.3.2 Пошук точки екстремуму функції двох змінних з використанням методу найшвидшого спуску

Розглянемо рішення задачі мінімізації функції двох змінних $f(x)$ з використанням методу найшвидшого спуску

$$f(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2.$$

Умови досягнення точки мінімуму визначаються заданою похибкою пошуку $\varepsilon = 0,7$.

Рішення. Насамперед обчислимо компоненти градієнта функції двох змінних $f(x)$, це є перші похідні по ∂x_1 та ∂x_2 :

$$\frac{\partial f}{\partial x_1} = 16x_1 + 4x_2,$$

$$\frac{\partial f}{\partial x_2} = 4x_1 + 10x_2.$$

Виберемо координати початкової точки

$$x^{(0)} = [10, 10]^T.$$

Для реалізації послідовного наближення до точки екстремуму методом найшвидшого спуску скористаємося математичним описом переходу від k -ї точки з координатами $(x_0^{(k)}, x_1^{(k)}, \dots, x_n^{(k)})$ до $(k+1)$ -ї точки з координатами $(x_0^{(k+1)}, x_1^{(k+1)}, \dots, x_n^{(k+1)})$ за формулою (5.8):

$$x_l^{(k+1)} = x_l^{(k)} + \alpha^{(k)} \frac{\partial f(x_0^{(k)}, x_1^{(k)}, \dots, x_n^{(k)})}{\partial x_l}, \quad l = 0, 1, \dots, n.$$

Ітерація 1. Для кожної з двох змінних, з урахуванням попередньо вибраного коефіцієнта $\alpha^{(0)} = 0,056$, запишемо для першої точки:

$$\begin{aligned} x_1^{(1)} &= x_1^{(0)} - \alpha^{(0)} \frac{\partial f}{\partial x_1} = x_1^{(0)} - \alpha^{(0)} (16x_1^{(0)} + 4x_2^{(0)}) = \\ &= 10 - 0,056(16 \cdot 10 + 4 \cdot 10) = -1,2; \end{aligned}$$

$$\begin{aligned} x_2^{(1)} &= x_2^{(0)} - \alpha^{(0)} \frac{\partial f}{\partial x_2} = x_2^{(0)} - \alpha^{(0)} (4x_1^{(0)} + 10x_2^{(0)}) = \\ &= 10 - 0,056(4 \cdot 10 + 10 \cdot 10) = 2,16. \end{aligned}$$

Отже, $x^{(1)} = [-1,20, 2,16]^T$.

Розрахуємо значення функції $f(x)$ у першій точці $x^{(1)}$

$$f(x^{(1)}) = 8x_1^2 + 4x_1x_2 + 5x_2^2 = 8 \cdot (-1,2)^2 + 4 \cdot (-1,2) \cdot 2,16 + 5 \cdot (2,16)^2 = 24,48.$$

Крок 2.

$$x_1 = 5 - \frac{(5-1)19,36}{19,36+12} = 2,53.$$

Крок 3. $f'(x_1) = 7,62 > 0$; покласти $b = 2,53$.

Ітерація 2

Крок 1. $b = 2,53$, $a = 1$, $f'(b) = 7,62$, $f'(a) = -12$.

Крок 2.

$$x_2 = 2,53 - \frac{(2,53-1)7,62}{7,62+12} = 1,94.$$

Крок 3. $f'(x_1) = 3,51 > 0$; покласти $b = 1,94$.

Ітерації продовжуються доти, поки не буде виконуватися нерівність $|f'(x_n)| \leq \varepsilon$.

9.3 Приклади розв'язання екстремальних задач функцій багатьох змінних

9.3.1 Визначення критеріїв оптимальності функції багатьох змінних

Розглянемо рішення задачі визначення критеріїв оптимальності функції двох змінних $f(x)$

$$f(x) = 2x_1^2 + 4x_1x_2^3 - 10x_1x_2 + x_2^2.$$

Необхідно класифікувати точку $\bar{x} = [0, 0]^T$.

Рішення. Насамперед обчислимо перші похідні по ∂x_1 та ∂x_2 заданої функції двох змінних $f(x)$:

$$\frac{\partial f}{\partial x_1} = 4x_1 + 4x_2^3 - 10x_2,$$

$$\frac{\partial f}{\partial x_2} = -10x_1 + 12x_1x_2^2 + 2x_2.$$

Для заданої точки $\text{grad } f(\bar{x}) = [0, 0]^T$, отже точка $\bar{x} = [0, 0]^T$ є стаціонарною.

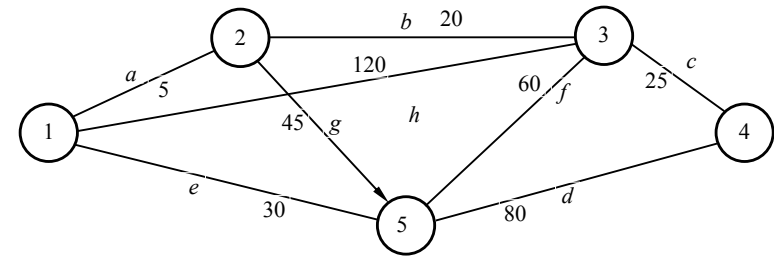
Обчислимо другі похідні по ∂x_1 , ∂x_2 та $\partial x_1 \partial x_2$ заданої функції двох змінних $f(x)$:

$$\frac{\partial^2 f}{\partial x_1^2} = +4,$$

$$\frac{\partial^2 f}{\partial x_2^2} = 24x_1x_2 + 2 = +2,$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 12x_2^2 - 10 = -10.$$

28. Побудувати структурну матрицю для графа



29. Побудувати матрицю довжин для графа із завдання 28.

30. Побудувати матрицю каналних ємностей ребер для графа із завдання 28.

Література до розділу 7

1 Кристофидес Н. Теория графов. Алгоритмический подход / Кристофидес Н. – М.: Мир, 1981. – 323 с.

2 Филипс Л. Методы анализа сетей / Л. Филипс, А. Гарсиа-Диас – М.: Мир, 1984. – 496 с.

3 Форд Л. Поток в сетях / Л. Форд, Д. Фалкерсон – М.: Мир. – 1966. – 276 с.

8 АНАЛІЗ, СИНТЕЗ І ОПТИМІЗАЦІЯ МЕРЕЖ ЗВ'ЯЗКУ

8.1 Задачі аналізу, синтезу і оптимізації мереж зв'язку

Телекомунікаційні мережі – це великі, складні й динамічні системи. Їх математична модель, як правило, будується на основі теорії *систем масового обслуговування* (СМО). В загальному випадку СМО обслуговують вимоги, що надходять до системи через випадкові інтервали часу, причому тривалість обслуговування також може бути випадковою. Методами теорії СМО досліджується вплив випадкових факторів на процеси функціонування системи.

Одним із класів СМО є *системи розподілу інформації* (СРІ), які характеризуються наявністю розподільчої мережі, подібно транспортним системам або системам енергопостачання. При передачі інформації аналогом розподільчої мережі є телекомунікаційна мережа, яка складається з каналів передачі інформації та вузлів комутації. Сукупність цих засобів зв'язує джерела інформації з їх споживачами. По каналах передається інформація, що безпосередньо є предметом передачі й розподілу, і допоміжна, яка необхідна в процесі керування роботою всієї системи. Вузли комутації забезпечують з'єднання каналів передачі інформації і в них за певними алгоритмами обслуговуються повідомлення телекомунікаційних служб мережі. При цьому обслуговування повідомлення ототожнюється з вимогою на його передачу або обробку і прикладами таких вимог можуть бути виклики телефонної станції або пакети пакетного комутатора.

Кількісна сторона процесів обслуговування потоків вимог (трафіка) в СРІ є предметом теорії телетрафіка, ще називаної теорією розподілу інформації. Дана теорія являє собою набір імовірнісних методів аналізу, синтезу та оптимізації СРІ і, таким чином, вирішення проблем проектування нових та експлуатації діючих мереж зв'язку. В теорії телетрафіка в якості СРІ може розглядатися не тільки мережа зв'язку в цілому, а й пучок каналів або ліній, окремий комутатор або весь комутаційний вузол. Без вирішення задач аналізу, синтезу і на цій основі оптимізації телекомунікаційних мереж та систем неможливий їх подальший розвиток.

Задача аналізу – це встановлення залежностей і значень величин, які характеризують якість обслуговування, від характеристик і параметрів вхідного потоку вимог, схеми і дисципліни обслуговування. Задача аналізу виникає в тих випадках, коли телекомунікаційна мережа або система вже побудована і функціонує. Цілями аналізу є отримання реальних характеристик СРІ, порівняння їх із проектними характеристиками, надання об'єктивних оцінок якості роботи системи. Аналіз дозволяє визначити причини зниження якості обслуговування і видати рекомендації з усунення цих причин. Іноді аналіз робиться після внесення змін у систему або після підключення нових джерел навантаження (реконструкції). Розробка методів оцінки якості функціонування телекомунікаційних мереж та систем є основною метою теорії телетрафіка.

Ітерація 3. $I_3 = (0,236, 0,618)$, $L_3 = 0,382 = \tau^2$. Наступне обчислення значення функції проводиться у точці, розміщеній на відстані $\tau \times$ (довжина отриманого інтервалу від лівої граничної точки інтервалу), або на відстані $(1 - \tau) \times$ (довжина інтервалу) від правої граничної точки. Таким чином,

$$w_4 = 0,618 - (1 - \tau)L_3 = 0,618 - \tau^2 L_3 = 0,618 - \tau^2(\tau^2) = 0,618 - \tau^4 = 0,472, \\ f(w_4) = (40 - 90w_4)^2 = (40 - 90 \cdot 0,472)^2 = 6,15.$$

Оскільки $f(w_4) < f(w_2)$ і $w_4 > w_2$, інтервал $w \leq w_2$ вилучається.

У результаті отримано наступний інтервал невизначеності $0,382 \leq w \leq 0,618$ для змінної w або $94,4 \leq x \leq 115,6$ для змінної x .

При переході до змінної x отримаємо мінімальні (після трьох ітерацій) межі інтервалу невизначеності $94,4 \leq x \leq 115,6$ для змінної x , які і є рішенням задачі.

Порівняємо ефективність методів пошуку границь інтервалу у завданнях 2 і 3. Якщо в процесі пошуку методом золотого перетину проведено шість обчислень значень функції, то довжина результуючого інтервалу для змінної w дорівнює

$$\tau^{N-1} = \tau^5 = 0,09,$$

що відповідає інтервалу довжини 8,1 для змінної x . Для порівняння нагадаємо, що в аналогічній ситуації метод розподілу інтервалу навпіл призвів до отримання інтервалу з довжиною 11,25.

9.2 Приклади розв'язання задач з використанням похідних

9.2.1 Приклад розв'язання задач з використанням методу Ньютона-Рафсона

Розглянемо задачу:

$$\text{мінімізувати } f(x) = 2x^2 + (16/x).$$

Для того, щоб визначити стаціонарну точку функції $f(x)$, скористаємося методом Ньютона-Рафсона [2], поклавши $x_1 = 1$:

$$f'(x) = 4x + (16/x^2), \quad f''(x) = 4 + (32/x^3).$$

$$\text{Ітерація 1. } x_1 = 1, \quad f'(x_1) = -12, \quad f''(x_1) = 36, \quad x_2 = 1 - (-12/36) = 1,33.$$

$$\text{Ітерація 2. } x_2 = 1,33, \quad f'(x_2) = -3,73, \quad f''(x_2) = 17,6,$$

$$x_3 = 1,33 - (-3,73/17,6) = 1,54.$$

Ітерації продовжуються доти, поки не буде виконана нерівність $|f'(x_k)| < \varepsilon$, де ε – заздалегідь установлена величина припустимого відхилення.

9.2.2 Приклад розв'язання задач з використанням методу січних

Розглянемо задачу: мінімізувати $f(x) = 2x^2 + (16/x)$ у інтервалі $1 \leq x \leq 5$.

$$f'(x) = \frac{df(x)}{dx} = 4x - \frac{16}{x^2}.$$

Ітерація 1

$$\text{Крок 1. } b = 5, \quad a = 1, \quad f'(b) = 19,36, \quad f'(a) = -12.$$

Ітерація 2

$$a = 82,5, \quad b = 127,5, \quad \text{довжина інтервалу } L = 127,5 - 82,5 = 45, \\ x_m = (82,5 + 127,5)/2 = 105, \\ f(x_m) = f(105) = (100 - 105)^2 = 25, \\ x_1 = a + (L/4) = 82,5 + (45/4) = 93,75, \\ x_2 = b - (L/4) = 127,5 - (45/4) = 116,25.$$

Розрахуємо і порівняємо значення $f(x_1)$ з $f(x_m)$ та $f(x_2)$ з $f(x_m)$:

$$f(x_1) = f(93,75) = (100 - 93,75)^2 = 39,06 > f(x_m) = 25, \\ f(x_2) = f(116,25) = (100 - 116,25)^2 = 264,06 > f(x_m) = 25.$$

Отже, вилучаються інтервали $(82,5, 93,75)$ і $(116,25, 127,5)$.

Інтервал невизначеності дорівнює $(93,75, 116,25)$.

Довжина інтервалу пошуку зменшується з 45 до 22,5.

Ітерація 3

$$a = 93,75, \quad b = 116,25, \quad \text{довжина інтервалу } L = 116,25 - 93,75 = 22,5, \\ x_m = (93,75 + 116,25)/2 = 105, \\ f(x_m) = f(105) = (100 - 105)^2 = 25, \\ x_1 = a + (L/4) = 93,75 + (22,5/4) = 99,375, \\ x_2 = b - (L/4) = 116,25 - (22,5/4) = 110,625,$$

Розрахуємо і порівняємо значення $f(x_1)$ з $f(x_m)$ та $f(x_2)$ з $f(x_m)$:

$$f(x_1) = f(99,375) = (100 - 99,375)^2 = 0,3906 < f(x_m) = 25.$$

Оскільки $f(x_1) < f(x_m)$, то у цьому разі $f(x_2)$ розраховувати не слід.

Отже, вилучається інтервал $(105, 116,25)$. Новий інтервал невизначеності дорівнює $(93,75, 105)$, його середня точка є $99,375$ (точка x_1 на ітерації 3).

Відзначимо, що за три ітерації (шість обчислень значення функції) початковий інтервал пошуку з довжини 90 зменшився до величини $90 \times (1/2)^3 = 11,25$.

9.1.3 Метод золотого перетину

Розглянемо задачу за прикладом 9.1.2, у якому потрібно мінімізувати $f(x) = (100 - x)^2$ у інтервалі $60 < x < 150$.

Для того щоб перейти до інтервалу одиничної довжини, проведемо заміну змінної x , поклавши $w = (x - 60)/90$. Отже, задача набуває наступного виду:

$$\text{мінімізувати } f(w) = (40 - 90w)^2 \\ \text{при обмеженні } 0 \leq w \leq 1.$$

Ітерація 1. $I_1 = (0, 1)$; $L_1 = 1$. Проведемо два перших обчислення значень функції:

$$w_1 = \tau = 0,618, \quad f(w_1) = (40 - 90w_1)^2 = (40 - 90 \times 0,618)^2 = 244,0, \\ w_2 = 1 - \tau = \tau^2 = 0,382, \quad f(w_2) = (40 - 90w_2)^2 = (40 - 90 \times 0,382)^2 = 31,6.$$

Оскільки $f(w_2) < f(w_1)$ і $w_2 < w_1$, інтервал $w \geq w_1$ вилучається.

Ітерація 2. $I_2 = (0, 0,618)$; $L_2 = 0,618 = \tau$. Наступне обчислення значення функції проводиться у точці

$$w_3 = \tau - \tau^2 = \tau(1 - \tau) = \tau^3 = 0,236, \\ f(w_3) = (40 - 90w_3)^2 = (40 - 90 \times 0,236)^2 = 352.$$

Оскільки $f(w_3) > f(w_2)$ і $w_3 < w_2$, інтервал $w < w_3$ вилучається.

Задача синтезу – це визначення структурних параметрів мережі або, наприклад, схеми комутаційного вузла цієї мережі при заданих потоках, дисципліні і якості обслуговування. Задача синтезу певною мірою є зворотною до задачі аналізу. Синтез (проекування) телекомунікаційних мереж може складатися з кількох етапів. З позицій системної методології, основними етапами рішення проблеми синтезу мереж та систем зв'язку є: аналіз проблеми; визначення системи; визначення цілей, критеріїв, ресурсів; визначення альтернативних варіантів; оцінка, порівняння і вибір варіантів; реалізація рішення. Задачі проектування і планування телекомунікаційних мереж виникають з необхідності завчасного вибору технічних засобів, що забезпечують задоволення потреб у передаванні інформаційних повідомлень. Метою проектування є оптимальна структура мережі на тривалу перспективу з урахуванням поточного стану розвитку телекомунікаційної техніки і технологій.

Задачі оптимізації є близькими до задач аналізу і синтезу. Як правило, при проектуванні телекомунікаційних мереж та систем вони формулюються в такий спосіб: визначити структурні параметри або алгоритми функціонування мережі (системи), для яких:

- при заданих потоках, якості і дисципліні обслуговування вартість або обсяг мережі (системи) мінімальні;
- при заданих потоках, дисципліні обслуговування і вартості якісні показники функціонування мережі (системи) оптимальні.

При експлуатації телекомунікаційних мереж та систем задача оптимізації формулюється як задача керування потоками вимог або структурою мережі для досягнення найкращих показників якості функціонування. Через великі обчислювальні труднощі задачі оптимізації телекомунікаційних мереж та систем вирішуються за допомогою ЕОМ.

Аналіз, синтез і оптимізація СРІ виконуються із застосуванням теорії імовірностей, математичної статистики, комбінаторних і алгебраїчних методів, теорії множин, теорії графів, принципів системного підходу (системотехніки) та інших.

Основними методами вирішення задач у теорії телетрафіка є аналітичний, числовий та метод статистичного моделювання.

Аналітичні методи дозволяють вирішувати задачі теорії телетрафіка в тих випадках, коли структура системи, характеристики потоку і дисципліна обслуговування відносно прості. При цьому розглядаються всі можливі стани системи, обумовлені, наприклад, положенням кожної точки комутації або кількістю зайнятих каналів. Такі стани називаються мікростанами системи. Щораз, коли надходить нова вимога, закінчується яка-небудь фаза роботи керуючого пристрою з встановлення з'єднання або закінчується з'єднання, система змінює свій мікростан. Для кожного мікростану записується рівняння статистичної рівноваги. Розв'язуючи систему таких рівнянь, знаходять точне рішення задачі в межах прийнятої моделі.

Числові методи використовують спеціальні алгоритми, що дозволяють знаходити наближені рішення ітераційними або іншими методами. Вони застосовуються для складних систем, де кількість мікростанів настільки велика,

що розв'язати систему рівнянь статистичної рівноваги неможливо навіть за допомогою швидкодіючих ЕОМ. Тому застосовується так званий макропідхід. У складній системі з дуже великою кількістю мікростанів ϵ та або інша ознака, за якою мікростани поєднуються в класи-макростани. Шляхом усереднення визначаються інтенсивності переходів з одних макростанів в інші. Для кожного макростана записується рівняння статистичної рівноваги. В результаті розв'язання системи таких рівнянь виводяться наближені формули для імовірностей макростанів.

Методи статистичного моделювання є найбільш універсальними методами, які придатні для вирішення задач практично будь-якої складності. Метод полягає в побудові математичної моделі системи, реалізація якої здійснюється у виді програми для ЕОМ. Моделювання дозволяє одержати числові результати, що характеризують якість обслуговування при заданих параметрах потоку, схеми і дисципліни обслуговування. Однак у силу специфіки методу він менш зручний у порівнянні з аналітичним і числовим методами при визначенні неявних закономірностей функціонування або залежностей між окремими характеристиками системи.

Для детального аналізу досліджуваних СРІ можливе поєднання аналітичних і числових методів з методом статистичного моделювання. Наприклад, якщо при малих значеннях параметрів системи вдається одержати рішення точними аналітичними методами і проаналізувати граничні випадки при асимптотичній поведінці характеристик досліджуваної системи, то потім отримані відомості доповнюються результатами статистичного моделювання в області реальних значень параметрів системи.

8.2 Моделі систем розподілу інформації

Теорія телетрафіка оперує не із самими системами розподілу інформації, а з їхніми математичними моделями. Для повного опису СРІ необхідно вказати імовірнісні процеси, що описують вхідний потік вимог, структуру системи та дисципліну обслуговування. Таким чином математична модель СРІ містить наступні основні елементи:

1. *Вхідний потік вимог на обслуговування (трафік)* – класифікується за ознаками стаціонарності, ординарності та післядії. Основними характеристиками потоку вимог є його параметр та інтенсивність.

2. *Структура системи розподілу інформації* – це інформація про кількість обслуговуючих пристроїв або серверів (пристроїв, що надають послуги), їх взаємне з'єднання (схему) та доступність для вхідних вимог.

3. *Дисципліна обслуговування потоку вимог* – характеризує взаємодію потоку вимог із системою розподілу інформації. В теорії телетрафіка дисципліна обслуговування описується:

- способом обслуговування вимог;
- порядком обслуговування вимог;
- режимами пошуку виходів схеми (наприклад, довільний або груповий);
- законами розподілу тривалості обслуговування;

9 ПРИКЛАДИ РОЗВ'ЯЗАННЯ ЗАДАЧ З ОПТИМІЗАЦІЇ

9.1 Приклади розв'язання задач пошуку точки екстремуму усередині заданого інтервалу

9.1.1 Визначення границь інтервалу

Розглянемо задачу мінімізації $[1, 2]$, а саме, пошуку мінімального значення функції $f(x) = (100 - x)^2$ при заданій початковій точці $x_0 = 30$ і величині кроку $|\Delta| = 5$.

Знак Δ визначається на підставі порівняння значень

$$f(x_0) = f(30) = 4900,$$

$$f(x_0 + |\Delta|) = f(35) = 4225,$$

$$f(x_0 - |\Delta|) = f(25) = 5625.$$

Оскільки

$$f(x_0 - |\Delta|) \geq f(x_0) \geq f(x_0 + |\Delta|),$$

то величина Δ має бути позитивною, а координата точки мінімуму x^* має бути більше 30.

$$\text{Маємо } x_2 = x_1 + 2\Delta = 35 + 2 \times 5 = 45,$$

$$f(x_2) = f(45) = (100 - 45)^2 = 3025 < f(x_1),$$

отже $x^* > 35$.

$$x_3 = x_2 + 2^2\Delta = 65, \quad f(x_3) = f(65) = (100 - 65)^2 = 1225 < f(x_2),$$

отже $x^* > 45$.

$$x_4 = x_3 + 2^3\Delta = 105, \quad f(x_4) = f(105) = (100 - 105)^2 = 25 < f(x_3),$$

отже $x^* > 65$.

$$x_5 = x_4 + 2^4\Delta = 185, \quad f(x_5) = f(185) = (100 - 185)^2 = 7225 > f(x_4),$$

остаточно, $x^* < 185$.

Таким чином, шість кроків обчислень x^* дозволили виявити інтервал $65 \leq x^* \leq 185$, у якому розміщена точка x^* . Отже, пошук граничних точок може бути завершеним.

9.1.2 Метод розподілу інтервалу навпіл

Мінімізувати $f(x) = (100 - x)^2$ в інтервалі $60 \leq x \leq 150$. Тут $a = 60$, $b = 150$ і $L = 150 - 60 = 90$.

$$x_m = (60 + 150)/2 = 105.$$

$$f(x_m) = f(105) = (100 - 105)^2 = 25$$

Ітерація 1

$$x_1 = a + (L/4) = 60 + (90/4) = 82,5,$$

$$x_2 = b - (L/4) = 150 - (90/4) = 127,5,$$

Розрахуємо і порівняємо значення $f(x_1)$ з $f(x_m)$ та $f(x_2)$ з $f(x_m)$:

$$f(x_1) = f(82,5) = (100 - 82,5)^2 = 306,25 > f(x_m) = 25,$$

$$f(x_2) = f(127,5) = (100 - 127,5)^2 = 756,25 > f(x_m) = 25.$$

Таким чином, вилучаються інтервали $(60, 82,5)$ і $(127,5, 150)$. Довжина інтервалу пошуку зменшується з 90 до 45.

Література до розділу 8

1. Крылов В.В. Теория телеграфика и её приложения / В.В.Крылов, С.С. Самохвалова // СПб.: БХВ-Петербург, 2005. – 288 с.: ил.
 2. Шнепс М.А. Системы распределения информации. Методы расчета: справ. пособие / М.А. Шнепс // М.: Связь, 1979. – 344 с., ил.
 3. Корнышев Ю.Н. Теория распределения информации: учеб. пособие для вузов / Ю.Н. Корнышев, Г.Л. Фань // М.: Радио и связь, 1985. – 184 с., ил.
 4. Ложковский А.Г. Экспериментальная оценка модели потока вызовов на современных телефонных сетях / А.Г. Ложковский, Н.В. Захарченко, С.М. Горохов // Наукові праці ОНАЗ ім. О.С. Попова. – 2001. – №2. – С. 40-43.
 5. Клейнрок Л. Теория массового обслуживания; пер. с англ. / Л. Клейнрок // М.: Машиностроение, 1979. – 432 с., ил.
 6. Вентцель Е.С. Теория случайных процессов и её инженерные приложения / Е.С. Вентцель, Л.А. Овчаров // М.: Наука. Ред. физ.-мат. лит., 1991. – 384 с.
 7. Кениг Д. Методы теории массового обслуживания; пер. с нем / Д.Кениг, Д.Штойян // М.: Радио и связь, 1981. – 128 с., ил.
 8. Ложковский А.Г. Спрощений метод розрахунку багатоканальної системи з чергою в моделі $M/D/m/\infty$ / А.Г. Ложковский // Наукові праці ОНАЗ ім. О.С. Попова. – 2008. – № 2. – С. 69-76.
 9. Ложковский А.Г. Моделирование многоканальной системы обслуживания с организацией очереди / А.Г. Ложковский, Н.С. Салманов, О.В. Вербанов // Восточно-европейский журнал передовых технологий. – 2007. – №3/6(27). – С.72-76.
 10. Ложковский А.Г. Нова методика оцінювання імовірності втрат викликів, наближена до реальних умов / А.Г. Ложковский // Зв'язок. – 2004. – №3. – С. 52-53.
 11. Ложковский А.Г. Метод расчета систем обслуживания с ожиданием при произвольном потоке вызовов / А.Г. Ложковский // Зв'язок. – 2006. – № 1. – С. 57-60.
 12. Ложковский А.Г. Сравнительный анализ методов расчета характеристик качества обслуживания при самоподобных потоках в сети / А.Г. Ложковский // Моделирование та інформаційні технології: зб. наук. пр. ПМЕ ім. Г. С. Пухова НАН України. – Вып. 47. – К.: 2008. – С. 187-193.
 13. Ложковский А.Г. Оценка параметров качества обслуживания самоподобного трафика энтропийным методом / А.Г. Ложковский, Р.А. Ганифаев // Наукові праці ОНАЗ ім. О.С. Попова. – 2008. – № 1. – С. 57-62.
 14. Ложковский А.Г. Расчет одноканальных систем с бесконечной очередью при экспоненциальной длительности обслуживания / А.Г. Ложковский // Наукові праці ОНАЗ ім. О.С. Попова. – 2009. – № 2. – С. 10-13.
- наявністю переваг (пріоритетів) в обслуговуванні деяких вимог;
 – наявністю обмежень при обслуговуванні (наприклад, за тривалістю очікування або обслуговування, кількості вимог, що очікують);
 – законами розподілу імовірностей виходу з ладу елементів схеми.
- У загальному випадку вхідний потік вимог на обслуговування описується функцією розподілу імовірностей інтервалів часу між сусідніми вимогами $A(z)$:
- $$A(z) = P(\leq z), \quad (8.1)$$
- де $P(\leq z)$ – імовірність того, що час між послідовними вимогами $\leq z$.
- Якщо інтервали часу між послідовними вимогами є незалежними та однаково розподіленими випадковими величинами, то вхідний потік вимог утворює стаціонарний процес відновлення. Це є ознакою незмінності в часі імовірнісних характеристик випадкових процесів, що у більшості випадків добре віддзеркалює реальні процеси в СРІ за короткі проміжки часу. Таким чином, функція розподілу інтервалів $A(z)$ є достатньою для опису потоку вимог.
- Час, протягом якого вимога перебуває в сервері, описується функцією розподілу імовірностей тривалості обслуговування $B(x)$:
- $$B(x) = P(\leq x). \quad (8.2)$$
- де $P(\leq x)$ – імовірність того, що час обслуговування $\leq x$.
- Для опису інтервалу часу між послідовними вимогами або тривалості обслуговування застосовні різні закони. Найбільше з них використовуються ті розподіли, що подано далі та позначено певними літерами:
- M – експонентний (M – марковська модель);
 - H – гіпер-експонентний (*Hyper-exponential*);
 - D – детермінований (*Determined*);
 - U – рівномірний (*Uniform*);
 - E – розподіл Ерланга;
 - G – довільний або узагальнений (*General*).
- Дисципліна обслуговування потоку вимог визначає правила обслуговування та долю вимог при їх надходженні до системи на обслуговування. Розрізняють такі типи СРІ, які визначаються способом обслуговування вимог:
1. *Системи із втратами* – вимоги, які при надходженні до системи не знаходять в ній жодного вільного серверу, отримують відмову в обслуговуванні та втрачаються.
 2. *Системи з чергами* – вимоги, які не можуть бути обслужені відразу із-за зайнятості всіх серверів системи, стають до черги, і за допомогою деякої дисципліни обслуговування черги визначається, у якому порядку вимоги, що очікують, вибираються із черги для обслуговування. Найбільш поширеними дисциплінами обслуговування черги є:
 - FF (*FIFO – first in first out*) – вимоги з черги обслуговуються в порядку їхнього надходження (упорядкована черга);
 - LF (*LIFO – last in first out*) – щораз перевагу для обслуговування має вимога, що надійшла до черги останньою;

– SR (*SIRO* – *service in random order*) – наступна вимога для обслуговування із черги вибирається випадково (випадкова черга).

3. *Комбіновані системи* з чергами та втратами (системи з чергою при обмеженнях). Наприклад, очікувати може тільки кінцева кількість вимог, обумовлена кількістю місць очікування, меншою за нескінченність. Можливо й так – вимога втрачається тоді, коли час очікування або перебування в черзі перевищує задані межі.

4. *Пріоритетні системи* – для вимог передбачено різні пріоритети в обслуговуванні. Якщо вимога, що надійшла, має високий пріоритет, а всі сервери зайняті, то вона або займає одне з перших місць у черзі, або тимчасово припиняє обслуговування вимоги низького пріоритету і займає її місце в сервері. При цьому можуть бути застосовані такі пріоритетні правила:

- абсолютний пріоритет з перериванням (*preemptive discipline*) – вимога високого пріоритету перериває обслуговування вимоги низького пріоритету. Може бути: абсолютний пріоритет із втратами (*preemptive loss discipline*), абсолютний пріоритет з дообслуговуванням (*preemptive resume discipline*) і абсолютний пріоритет з обслуговуванням заново (*preemptive repeat different discipline*);
- відносний пріоритет (*head of the line priority discipline*) – вимога високого пріоритету посідає перше місце в черзі і переривань немає.

Змішані пріоритети зумовлюють вибір абсолютного або відносного пріоритетного правила залежно від уже реалізованої частини тривалості обслуговування, а динамічні – залежно від типу поточних вимог і співвідношення кількості вимог різних пріоритетів, що є в серверах та в черзі.

Основні характеристики, що представляють структуру СРІ такі:

- кількість обслуговуючих пристроїв (серверів, ліній, каналів, портів);
- кількість місць очікування або максимальна довжина черги (ємність пам'яті, у якій накопичуються вимоги, що очікують);
- доступність – спосіб включення серверів, за якого кожній вимозі доступні всі або не всі (хоча всім вимогам у сукупності доступні всі) сервери. Відповідно до цього структура СРІ може бути повнодоступною або неповнодоступною;
- взаємне з'єднання (схема) – спосіб включення серверів, за якого кожна вимога обслуговується одним сервером або декількома, але поетапно. Відповідно до цього схеми СРІ поділяються на однокаскадні та багатокаскадні (ланцюгові).

Структурні характеристики СРІ частково впливають на дисципліну обслуговування потоку вимог. Наприклад, значення кількості місць очікування $r = 0$ визначає систему із втратами, значення $0 < r < \infty$ – комбіновану систему з чергою та із втратами, а $r = \infty$ – чисту систему з чергою, тобто з нескінченною кількістю місць очікування.

Для стислого запису досліджуваної системи Д. Кендаллом запропоновано спеціальне *умовне позначення базової моделі*, в якому зі всіх наведених параметрів математичної моделі СРІ представлено чотири елементи: $A/B/m/r$.

24. Якими методами досліджується функціонування СРІ за умови обслуговування пуассонівського потоку вимог?
25. Які ознаки пуассонівського потоку вимог?
26. Що визначає поняття „стан системи”?
27. Для розрахунку якої системи призначена *B*-формула Ерланга?
28. Для розрахунку якої системи призначена *C*-формула Ерланга?
29. Що визначає формула Літтла?
30. Що визначає формула Поллачека-Хінчина?
31. Як відрізняються характеристики якості обслуговування в моделях $M/M/1/\infty$ та $M/D/1/\infty$?
32. Що таке „гіперекспонентний розподіл”?
33. Як відрізняються між собою інтенсивність навантаження та її дисперсія для пуассонівського та реального (гіперекспонентного) потоку вимог?
34. Що таке „коефіцієнт скупченості” або „підфактор” навантаження СРІ?
35. Чим характерний трафік пакетних мереж зв'язку і якою моделлю він може бути описаний?
36. Які значення може приймати коефіцієнт Херста для самоподібних та несамоподібних потоків трафіка.
37. Що визначає ентропія імовірнісного розподілу?
38. Чому в моделі $G/M/1/\infty$ (за умов довільного потоку) імовірність очікування може не співпадати з імовірністю зайнятості або із коефіцієнтом загрузки системи?
39. Визначити частку втрачених вимог для системи $M/M/1$, до якої надходить потік вимог з інтенсивністю $\lambda = 2$, а інтенсивність обслуговування $\mu = 4$.
40. Визначити середню кількість вимог в черзі в системі $M/M/1/\infty$, до якої надходить потік вимог з інтенсивністю $\lambda = 2$, а інтенсивність обслуговування $\mu = 2,5$.
41. Визначити імовірність очікування та імовірність втрати вимоги для системи $M/M/1/5$, до якої надходить потік вимог з інтенсивністю $\lambda = 2$, а інтенсивність обслуговування $\mu = 4$.
42. На скільки відрізняється середня кількість вимог у системах $M/M/1/\infty$ та $M/D/1/\infty$, до яких надходить потік вимог з інтенсивністю 2 вимоги на секунду за середньої тривалості обслуговування вимог $\bar{x} = 0,35$ с.
43. Скільки в середньому (в секундах) будуть очікувати вимоги в черзі системи $M/M/80/\infty$ за інтенсивності навантаження $\Lambda = 70$ Ерл та середньої тривалості обслуговування 35 с?
44. Визначити імовірність очікування в системі $M/M/80/\infty$ за інтенсивності навантаження $\Lambda = 70$ Ерл.
45. Довести, що імовірність очікування в системі $M/M/m/r$ завжди менша за цю ж імовірність в системі $M/M/m/\infty$.

результати розрахунків виявляються трохи заниженими. При цьому найбільша похибка обчислень за формулами моделі $G^*/D/1/\infty$ у порівнянні з результатами моделювання спостерігається у випадку самоподібного потоку вимог, згенерованому за розподілом Парето. Наприклад, при завданні параметра форми розподілу $a = 1,3$ (відповідає коефіцієнтові самоподібності трафіка $H = 0,85$ [13] і $v_2 \geq 30$) похибка обчислень не перевищує 20 % для всіх параметрів QoS . При збільшенні a і пов'язаним з цим зменшенням H , наприклад, до значень 1,9 і 0,55 відповідно (ще самоподібний процес), дана похибка не перевищує вже 10 %.

Слід зазначити, що точність розрахунків декілька підвищується при зменшенні інтенсивності навантаження ρ і в багатьох випадках при $\rho < 0,5$ похибка розрахунків не перевищує 5%.

Контрольні питання та задачі до розділу 8

1. В чому полягає сутність задач аналізу телекомунікаційних систем?
2. Яка значимість задач синтезу телекомунікаційних систем?
3. В чому полягає мета задач оптимізації телекомунікаційних систем?
4. Назвіть основні методи вирішення задач теорії телетрафіка.
5. Які складові входять до математичної моделі систем розподілу інформації?
6. Як описується дисципліна обслуговування вимог у СРІ?
7. Як описується вхідний потік вимог на обслуговування у СРІ?
8. Які види імовірнісних розподілів застосовуються для опису випадкових процесів, що відбуваються у СРІ (вхідний потік, тривалість обслуговування, стани системи)?
9. Назвіть види СРІ за способом обслуговування вимог.
10. Назвіть дисципліни обслуговування черги в СРІ.
11. Назвіть основні правила обслуговування вимог за пріоритетами.
12. Назвіть основні характеристики, що представляють структуру СРІ.
13. Поясніть структуру умовного позначення базової моделі СРІ за Кендаллом.
14. Що таке „інтенсивність потоку вимог“?
15. Що таке „інтенсивність обслуговування вимог“?
16. Що таке „навантаження“ СРІ?
17. Що таке „інтенсивність навантаження“ і якими способами її можна визначити?
18. В чому різниця між вхідним та обслугованим навантаженням СРІ?
19. Які характеристики властиві пуассонівському потоку вимог?
20. Які типи реального трафіка визначено для телекомунікаційних мереж і в чому різниця між ними?
21. Назвіть характеристики якості обслуговування для системи з втратами.
22. Назвіть характеристики якості обслуговування для системи з чергами.
23. Що таке „пропускна здатність“ СРІ?

Елемент A характеризує потік вимог і певною літерою з наведених вище видів розподілів позначається функція розподілу імовірностей інтервалів часу між сусідніми вимогами.

Елемент B характеризує випадкові послідовності тривалості обслуговування на окремих серверах системи і аналогічно до попереднього може використовувати такі ж розподіли.

Елементи m та r характеризують відповідно кількість обслуговуючих пристроїв і місць очікування в системі.

Умовне позначення базової моделі крім цих основних позначень може містити ще й додаткові символи, які вказуються після знаку „:“ і можуть уточнювати особливості системи.

Наприклад. Запис $M/D/120/r = \infty$ значить, що СРІ обслуговує найпростіший потік вимог (M – експонентний розподіл імовірностей інтервалів часу між сусідніми вимогами) за допомогою $m = 120$ обслуговуючих пристроїв, де кожна вимога має постійну тривалість обслуговування (D – детермінований розподіл). В системі передбачено нескінчену кількість місць очікування ($r = \infty$), що й визначає дисципліну обслуговування з чергами. Запис $G/M/120:Loss$ значить, що СРІ обслуговує довільний потік вимог з експонентним розподілом їх тривалості. Ємність накопичувальної пам'яті, у якій вимоги могли б очікувати у разі зайняття всіх 120 серверів системи, дорівнює нулю (не записується), через те дана система має дисципліну обслуговування із втратами ($Loss$ – втрати).

Отже, базова математична модель СРІ позначається послідовністю символів: перший – вказує функцію розподілу інтервалів часу між вимогами, другий – функцію розподілу тривалості обслуговування, третій і наступний (необов'язковий) символи – схему і дисципліну обслуговування.

Побудова математичної моделі, що адекватно відображає реальну систему розподілу інформації, у багатьох випадках є непростою задачею. Від правильного вибору моделі залежить точність вирішення задач аналізу, синтезу або оптимізації.

8.3 Математична модель потоку вимог

Сукупність (послідовність) подій надходження до системи в моменти t_1, \dots, t_n вимог на обслуговування утворюють *потік вимог* (рис. 8.1).

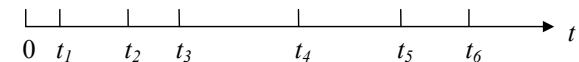


Рисунок 8.1 – Потік вимог на обслуговування

Інтервали часу $z_n = t_n - t_{n-1}$ між подіями надходження вимог можуть бути однаковими або випадковими й відповідно потік буде *детермінованим* або *випадковим*. В телекомунікаціях переважна більшість потоків є випадковими і детермінований потік може розглядатись як окремих випадок випадкового.

Випадковий потік вимог може бути описаний функцією розподілу імовірностей інтервалів часу між сусідніми вимогами $F(t)$:

$$F(t) = P(z_n \leq t), \quad (8.3)$$

де $P(z_n \leq t)$ – імовірність того, що час між послідовними вимогами $z_n \leq t$.

Основна характеристика такого потоку – це середнє значення інтервалів z , що для випадкової величини є математичним сподіванням \bar{z} . Параметр, зворотній до математичного сподівання \bar{z} , визначається як *інтенсивність потоку* надходження вимог λ за одну одиницю часу, якими вимірюється \bar{z} :

$$\lambda = \frac{1}{\bar{z}}. \quad (8.4)$$

Наприклад, при $\bar{z} = 0,1$ с інтенсивність потоку $\lambda = 10$ вимог на секунду, а при $\bar{z} = 100$ мс – інтенсивність потоку $\lambda = 0,01$ вимоги на мілісекунду.

Найпоширенішою математичною моделлю потоку вимог в телефонних мережах зв'язку є модель експонентного розподілу інтервалів часу між вимогами (викликами АТС) з параметром λ :

$$P(z_n \leq t) = 1 - e^{-\lambda t}. \quad (8.5)$$

Густина (щільність) цього розподілу дозволяє розрахувати імовірність будь-якої тривалості $z_n = t$ випадкової величини z (інтервалів між вимогами) за заданої інтенсивності надходження вимог λ :

$$p(t) = \lambda e^{-\lambda t}. \quad (8.6)$$

Середнє значення випадкової величини t , розподіленої за експонентним законом (8.6), дорівнює λ^{-1} і тому з (8.4) випливає, що параметр даного розподілу λ – це теж середня кількість вимог на одиницю часу, в яких вимірюється \bar{z} .

На рис. 8.2 представлено два графіки експонентного розподілу (8.6), які позначено $p(z)$ та $p^1(z)$ з параметрами $\lambda = 0,5$ та $\lambda^1 = 0,25$ відповідно.

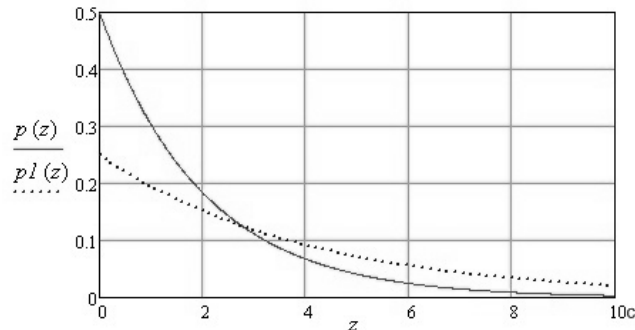


Рисунок 8.2 – Експонентний розподіл інтервалу z

З графіків видно, що для потоку, заданого функцією $p(z)$ є значно більшою імовірність (частка) коротких інтервалів між вимогами, ніж для потоку, заданого функцією $p^1(z)$. Це свідчить про більшу інтенсивність потоку вимог, яка у першому випадку складає $\lambda = 0,5$, а у другому – $\lambda^1 = 0,25$ вимог на

Таблиця 8.3 – Основні параметри якості обслуговування

Хар-ка QoS	Модель			
	M/D/1/∞	G/M/1/∞	G/G/1/∞	G*/D/1/∞
			точно	приблизно
P_{zn}	ρ	ρ	$\rho, \frac{Q}{t_q P_{w>0}}$	
$P_{w>0}$	ρ	$1 - \frac{\rho}{N}$	$\frac{W}{t_q}, \frac{Q}{t_q \rho}$	
Q	$\frac{\rho^2}{2(1-\rho)}$	$\frac{\rho P_{w>0}}{1-P_{w>0}}$	$\frac{\rho P_{w>0}}{1-P_{w>0}} (t_q - W)$	$\frac{\rho P_{w>0}}{2(1-P_{w>0})}$
W	$\frac{\rho}{2(1-\rho)}$	$\frac{P_{w>0}}{1-P_{w>0}}$	$\frac{P_{w>0}}{1-P_{w>0}} (t_q - W)$	$\frac{P_{w>0}}{2(1-P_{w>0})}$
t_q	$\frac{1}{2(1-\rho)}$	$\frac{1}{1-P_{w>0}}$	$\frac{1}{1-P_{w>0}} (t_q - W)$	$\frac{1}{2(1-P_{w>0})}$
N	$\rho + \frac{\rho^2}{2(1-\rho)}$	$\frac{\rho}{1-P_{w>0}}$	$\rho + \frac{\rho P_{w>0}}{1-P_{w>0}} (t_q - W)$	$\frac{\rho(2-P_{w>0})}{2(1-P_{w>0})}$
T	$1 + \frac{\rho}{2(1-\rho)}$	$\frac{1}{1-P_{w>0}}$	$1 + \frac{P_{w>0}}{1-P_{w>0}} (t_q - W)$	$\frac{2-P_{w>0}}{2(1-P_{w>0})}$

Якщо припустити, що умова (8.104) здійснена і при $P_{w>0} \neq \rho$, то з тих же формул моделі G/G/1/∞ можна одержати відповідні формули розрахунку характеристик QoS моделі G/D/1/∞ (стовпчик 5 табл. 8.3). Але оскільки, як показує імітаційне моделювання, отримані у такий спосіб формули дають не зовсім точні і різні результати для різних вхідних потоків, то краще назвати таку модель з умовно загальним розподілом випадкової величини інтервалу часу між вимогами і позначити G*/D/1/∞.

Для моделі G*/D/1/∞ різниця (8.104) не завжди дорівнює 0,5. За результатами імітаційного моделювання, виконаного за допомогою алгоритму [9], встановлено, що для потоків, у яких коефіцієнт варіації тривалості інтервалу часу між вимогами $v_z < 1$ (більш вирівняний потік у порівнянні з пуассонівським) $t_q - W < 0,5$. Наприклад, за рівномірного розподілу інтервалу часу між вимогами (модель U/D/1/∞) $v_z = 0,58$. Для пуассонівського потоку з експонентним розподілом зазначеного інтервалу при $v_z \equiv 1$ або з розподілом Вейбулла при підбраному коефіцієнті форми розподілу так, щоб $v_z = 1$, умова (8.18) виконується. Для гіпер-експонентного, логарифмічно нормального або розподілів Парето та Вейбулла при параметрах розподілів, що забезпечують значення v_z від одиниць до декількох десятків (дуже нерівномірні потоки), різниця t_q і W завжди трохи більше 0,5.

У випадку $t_q - W < 0,5$ результати розрахунків трохи вище результатів моделювання, при $t_q - W = 0,5$ – вони цілком збігаються, а при $t_q - W > 0,5$ –

При аналізі пакетних комутаторів варто враховувати наявність у кожного пакета заголовка фіксованої довжини, що вимагає обліку в тривалості обслуговування деякої постійної добавки, навіть якщо припустити, що розподіл довжин пакетів є експонентним. Отже, розробка методу розрахунку основних характеристик якості обслуговування заявок у системах, представлених моделлю $G/D/1/\infty$ (D – детермінований інтервал часу) також є актуальною.

У п. 8.5 показано, що для моделі $G/M/1/\infty$ з геометричним розподілом r_k (розподіл кількості заявок у системі в моменти надходження нових заявок) при експонентному законі розподілу тривалості обслуговування середній час очікування в системі W дорівнює:

$$W = \frac{P_{w>0}}{1 - P_{w>0}}. \quad (8.101)$$

З урахуванням цього результату та формули Літгла в табл. 8.2 наведено формули розрахунку характеристик якості обслуговування для моделі $G/M/1/\infty$ і відомий результат для моделі $M/D/1/\infty$, що виходить з формули Поллачека-Хінчина.

В формули розрахунку характеристик QoS моделі $G/G/1/\infty$ можна ввести величину $t_q - W$ (табл. 8.3). Правильність цього кроку легко перевірити підстановкою в кожен з них справедливого за визначенням для будь-якої моделі рівняння $P_{w>0} = \frac{W}{t_q}$.

Для розрахунку характеристик QoS моделі $G/D/1/\infty$ можна використати виявлену в п. 8.5 властивість моделі $G/M/1/\infty$, а саме: $t_q = T$. Оскільки з визначення середнього часу перебування вимоги в будь-якій системі (в тому числі й в моделі $G/G/1/\infty$) виходить, що:

$$T = W + 1. \quad (8.102)$$

то з урахуванням наведеної властивості моделі $G/M/1/\infty$:

$$t_q - W = 1. \quad (8.103)$$

Видно, що при виконанні умови (8.103), кожна з формул розрахунку характеристик QoS моделі $G/G/1/\infty$ збігається з аналогічною формулою моделі $G/M/1/\infty$. У той же час відомо, що для моделі $M/D/1/\infty$:

$$t_q - W = 0,5. \quad (8.104)$$

Тому при виконанні умови (8.104) разом з умовою $P_{w>0} = \rho$, властивої за пуассонівського потоку, кожна з формул моделі $G/G/1/\infty$ перетвориться у відповідну формулу моделі $M/D/1/\infty$ (стовпчики 4 та 2 табл. 8.3).

єдиницю часу. Відповідно до цього середнє значення інтервалу часу між вимогами складає $\bar{z} = 2$ та $\bar{z}1 = 4$ одиниць часу.

Випадковий потік вимог можна описати й функцією $P_i(t)$ – розподілом імовірностей кількості вимог i за умовну одиницю часу t . Наприклад, якщо діаграму процесу надходження вимог, яку показано на рис. 8.1, умовно поділити на однакові проміжки часу тривалістю t , що у рази або десятки разів перевищує середнє значення інтервалів \bar{z} , то на кожний з таких умовних інтервалів припаде випадкова кількість вимог i . Функція розподілу випадкової величини i й буде описувати потік вимог, що надходить до системи на обслуговування.

Відомо, якщо інтервал часу між подіями (вимогами) z розподілений за експонентним законом, то кількість таких подій на умовну одиницю часу тривалістю t буде розподілена за законом Пуассона:

$$P_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}. \quad (8.7)$$

Величина λt є параметром розподілу Пуассона. За цим розподілом можна розрахувати імовірність надходження в систему точно i вимог за умовну одиницю часу тривалістю t за заданої інтенсивності надходження вимог λ .

Для наведеного вище прикладу експонентного потоку з інтенсивністю $\lambda = 0,5$ побудовано розподіл Пуассона випадкової кількості вимог, що припадає на умовні інтервали часу, наприклад, тривалістю $t = 20$ с.

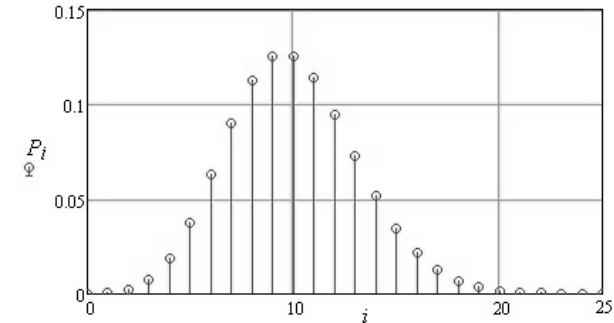


Рисунок 8.3 – Розподіл Пуассона при $\lambda = 0,5$ та $t = 20$ с

Середнє значення випадкової величини i , розподіленої за законом (8.7), визначається як $\bar{i} = \lambda t$, і у цьому випадку $\lambda t = 10$. З рис. 8.3 видно, що імовірність саме такого значення кількості вимог i за умовну одиницю часу t є найбільшою, і це є імовірність середнього значення \bar{i} . При зростанні i від нуля до \bar{i} імовірність $P_i(t)$ поступово зростає, а далі – зменшується. Даний графік достатньо симетричний, а форма апроксимуючої кривої наближується до форми кривої нормального (Гауса) закону розподілу випадкової величини.

Математичну модель потоку вимог, що надходить на обслуговування, можна відобразити за допомогою імовірнісних функцій розподілу:

- інтервалів часу між сусідніми вимогами z , наприклад, (8.6);
- кількості вимог i за умовну одиницю часу t , наприклад, (8.7).

У першому випадку, як правило, застосовуються неперервні закони, а у другому – дискретні.

Свою назву кожний із видів модельованого потоку запозичує від назви закону розподілу імовірнісних характеристик потоку – інтервалів часу між вимогами або їх кількості за умовну одиницю часу. Тому описана модель потоку, яка визначається розподілом (8.7) і тривалий час використовується в телефонних мережах, називається пуассонівським потоком.

Очевидно, що оцінка якості обслуговування або пропускної здатності СРІ потребує врахування всіх елементів її моделі. Найбільш складним при цьому є врахування математичної моделі вхідного потоку вимог. Саме з цієї причини весь пакет задач аналізу й синтезу СРІ для будь-яких із її схем та дисциплін обслуговування вирішено тільки для випадку найпростішої моделі трафіка – моделі пуассонівського потоку. Для цієї моделі відомі всі аналітичні формули розрахунку основних характеристик якості обслуговування в системах розподілу інформації [1–3].

Стрімкий розвиток телекомунікаційних технологій, нові принципи побудови мереж зв'язку, зміна структурного складу абонентів і спектра надаваних послуг – все це дуже позначається на характері трафіка в мережах. Ці фактори, насамперед, збільшують нерівномірність інтенсивності потоків вимог, яка вимірюється дисперсією інтенсивності. Результати статистичних вимірів, виконуваних на різних етапах еволюції розвитку мереж і послуг, дають змогу виділити 3 типи трафіка, до яких слід вживати певні математичні моделі:

I тип – в моносервісних мережах з однорідним трафіком. Такими є суто телефонні мережі з єдиною послугою телефонного зв'язку, що й зумовило однорідність трафіка. Найпростіша модель пуассонівського потоку, в основному, відповідає таким умовам, а значення інтенсивності трафіка та її дисперсії співпадають або достатньо близькі.

II тип – в мультисервісних мережах з різнорідним трафіком. Інтегральний характер мультисервісної мережі з розширеним спектром надаваних послуг зумовлює різнорідність трафіка, яка сильно змінює його параметри та математичну модель. Реальним потокам властива підвищена нерівномірність трафіка, за якої дисперсія інтенсивності трафіка перевищує її математичне сподівання від 2 до 15 раз. Іноді дане перевищення буває й більшим, але це відбувається або за межами ГНН, або на невеликих пучках каналів [4].

III тип – в пакетних мережах з мультисервісним трафіком. Трафік має довгострокові залежності в інтенсивності та ще більш суттєво відрізняється від пуассонівського потоку. Адекватною моделлю потоків в таких мережах є самоподібні процеси. В мультисервісних пакетних мережах трафік є різнорідним і з певними вимогами до QoS . Тут передачу потоків різних служб забезпечує одна і та ж сама мережа з єдиними протоколами та законами управління. Оскільки джерела кожної служби можуть мати різні швидкості передавання інформації або змінювати її в процесі сеансу зв'язку, то

Таблиця 8.2 – Залежності між параметрами QoS в системі $G/M/1/\infty$

Хар-ка QoS	Характеристика QoS				
	Q	W	t_q	N	
P_{zn}	ρ	ρ	ρ	ρ	
$P_{w>0}$	$\frac{Q}{\rho+Q}$	$\frac{W}{1+W}$	$1-\frac{1}{t_q}$	$1-\frac{\rho}{N}$	$\frac{Q}{N}$
Q	–	$\rho \cdot W$	$\rho \cdot (t_q - 1)$	$N - \rho$	$N \cdot P_{w>0}$
W	$\frac{Q}{\rho}$	–	$t_q - 1$	$\frac{N}{\rho} - 1$	
t_q	$1 + \frac{Q}{\rho}$	$1 + W$	–	$\frac{N}{\rho}$	
N	$\rho + Q$	$\rho \cdot (1 + W)$	$t_q \rho$	–	$\frac{Q}{D_{w>0}}$
T	$1 + \frac{Q}{\rho}$	$1 + W$	t_q	$\frac{N}{\rho}$	

По-перше, середній час очікування в черзі t_q чисельно збігається із середнім часом перебування вимоги в системі T . Це значить, що середній час очікування в системі W менше середнього часу очікування в черзі t_q на величину середньої тривалості обслуговування (одиниця часу \bar{x}):

$$W = t_q - 1. \tag{8.99}$$

По-друге, імовірність очікування можна визначити як

$$P_{w>0} = \frac{Q}{N}. \tag{8.100}$$

За визначенням (8.12) імовірність очікування $P_{w>0}$ – це співвідношення кількості затриманих вимог та загальної кількості вимог, що надійшли. В цьому випадку, частку вимог, що очікують, можна визначити і як відношення середньої кількості вимог у черзі Q до середньої кількості вимог у системі N . Звідси з урахуванням формули Літтла виходить кілька важливих співвідношень між параметрами QoS , справедливих вже для моделей $G/G/1/\infty$ і $G/G/m/\infty$ (наведено у крайньому правому стовпці табл.8.2).

8.7.6 Система із необмеженою чергою $G/D/1/\infty$

В телекомунікаційних системах часто використовується незмінна довжина пакетів при постійній тривалості їхнього обслуговування, наприклад, за технології АТМ. Для таких випадків у досліджуваній математичній моделі можна обмежитися тільки детермінованим законом розподілу тривалості обслуговування. Однак і за змінної довжини пакетів, наприклад, експонентне наближення для розподілу тривалості обслуговування виявляється недостатнім.

навантаження ($\rho < 1$ – умова ергодичності процесу і наявності стаціонарного розподілу імовірностей станів системи).

Для будь-якої односерверної системи $\rho = 1 - p_0$, де p_0 – імовірність вільності системи (стан системи p_0 – зайнято 0 каналів). Отже, ρ – чисельно збігається з імовірністю зайнятості системи P_{zn} (стан системи p_1 – зайнятий єдиний канал, відповідає частці часу зайнятості каналу). З урахуванням вимог, що знаходяться в черзі, у стаціонарному режимі існує стаціонарний розподіл кількості вимог у системі p_k , де k – кількість вимог. Цей розподіл не залежить від моменту прибуття вимоги до системи.

За пуассонівського потоку імовірність очікування $P_{w>0}$ збігається з імовірністю зайнятості системи P_{zn} (див. 8.61). Для односерверної моделі $M/G/1/\infty$ з довільним розподілом тривалості обслуговування дані імовірності однакові і $P_{w>0} = \rho$. Однак для моделі $G/M/1/\infty$ такої рівності нема, тобто за цим параметром моделі не інваріантні. В [5, с. 272] показано, що система $G/M/1/\infty$ призводить до геометричного розподілу кількості вимог у системі в моменти надходження нових вимог r_k , де k – кількість вимог. Розподіл p_k відрізняється від розподілу r_k тим, що $p_0 = 1 - P_{zn}$ (або $p_0 = 1 - \rho$), у той час як $r_0 = 1 - P_{w>0}$. Для системи $M/G/1/\infty$ виконується рівняння $p_k = r_k$.

Заявка повинна очікувати обслуговування з імовірністю $P_{w>0} = 1 - r_0$. Тому при експонентному розподілі тривалості обслуговування безумовний розподіл тривалості очікування визначиться так:

$$W(t) = 1 - P_{w>0} e^{-\mu(1-P_{w>0})t}, \quad \text{при } t \geq 0. \quad (8.98)$$

З цього можна розрахувати середній час очікування в системі W і всі інші параметри якості обслуговування:

$$P_{zn} = \rho; \quad P_{w>0} = 1 - \frac{\rho}{N};$$

$$W = \frac{P_{w>0}}{1 - P_{w>0}}; \quad Q = \frac{\rho \cdot P_{w>0}}{1 - P_{w>0}}; \quad t_q = T = \frac{1}{1 - P_{w>0}}; \quad N = \frac{\rho}{1 - P_{w>0}}.$$

Оскільки $\mu = 1$, тобто середня тривалість обслуговування $\bar{x} = 1/\mu$ приймається за умовну одиницю часу, то W , t_q і T оцінюються в одиницях середньої тривалості обслуговування.

В табл. 8.2 наведені залежності між основними параметрами QoS для моделі $G/M/1/\infty$. З таблиці виходить, що за наявності лише одного відомого параметра (наприклад, відомий параметр Q , W , t_q або N) всі інші параметри розраховуються через наведені в таблиці співвідношення.

Для моделі $G/M/1/\infty$ виявлено й перевірено за допомогою імітаційного моделювання важливі властивості односерверної системи, які виконуються тільки за експонентної тривалості обслуговування (виділені рамкою в табл.8.2).

об'єднаному потоку пакетів властива так звана «пачковість» трафіка (*burstness*), вимірювана коефіцієнтом пачковості [1]. Ця пачковість обумовлює ще більшу нерівномірність трафіка, за якої дисперсія інтенсивності трафіка перевищує її математичне очікування від 20 до 60 раз і більше.

Незалежно від способу представлення математичної моделі потоку вимог вибрана модель обов'язково має бути адекватною реальним потокам трафіка телекомунікаційних мереж, оскільки від цього суттєво залежить точність розрахунку характеристик якості обслуговування та пропускної здатності СРІ при їх аналізі, синтезі і оптимізації.

8.4 Навантаження та його види

Сумарний час обслуговування всіх вхідних вимог для серверів СРІ є певним *навантаженням* (*traffic*). Чим більший цей час – тим більше навантаження „переносять” або „оброблюють” сервери системи, і навпаки. В теорії телетрафіка розрізняють *вхідне* (*traffic offered*), *обслужене* (*traffic carried*) та *надлишкове* (*overflow traffic*) навантаження. Надлишкове навантаження – це різниця між вхідним і обслуженим навантаженнями, яка для систем із втратами буде *втраченим* навантаженням.

Навантаження на сервери системи є результатом спільного процесу надходження та обслуговування вимог. Оскільки вимоги надходять до системи через випадкові інтервали часу (або кількість вимог на умовну одиницю часу є випадковою) і тривалість обслуговування також може бути випадковою, то й навантаження є випадковою величиною.

Навантаження вимірюється в *годино-заняттях*. Наприклад, навантаження в одне годино-заняття (1 г.-зан.) утворюється безперервним обслуговуванням вимог одним сервером на протязі однієї години або двома серверами терміном по півгодини кожний і т.д. В силу цього параметр „навантаження” не дає чітких відомостей, щодо напруженості роботи системи, оскільки невідомо, за який час виконане це навантаження. Адже сумарний час обслуговування всіх вимог, який дорівнює, наприклад, 20 г.-зан. може свідчити про роботу в системі 1 сервера на протязі 20 годин або 20 серверів на протязі 1 години кожний. Тому вводиться поняття *інтенсивності обслуженого навантаження* Y (*traffic carried intensity*), яка визначається як приведене навантаження або сумарний час обслуговування всіх вимог x_i , поділений на певний інтервал часу $t_2 - t_1$, що може дорівнювати, наприклад, 1 годині:

$$Y = \frac{\sum x_i}{t_2 - t_1}. \quad (8.8)$$

Миттєве значення обслуженого навантаження визначається кількістю зайнятих серверів або кількістю одночасно обслуговуваних вимог. Через те що ця кількість також є випадковою величиною, то основною характеристикою обслуженого навантаження буде її математичне сподівання, що так само є *інтенсивністю обслуженого навантаження* Y . Якщо в багатосерверній системі за час її роботи від t_1 до t_2 з будь-якою періодичністю, наприклад 1 с, n раз

обчислити (сканувати) кількість зайнятих серверів C_i і у підсумку розрахувати її середнє значення

$$Y = \frac{1}{n} \sum_{i=1}^n C_i, \quad (8.9)$$

то воно однозначно співпадає зі значенням, розрахованим для цієї ж системи за формулою (8.8).

Отже інтенсивність обслуженого навантаження – це приведений час сумарного обслуговування всіх вимог (8.8) або середнє значення кількості зайнятих серверів (8.9). Обидва способи можуть використовуватись в засобах вимірювання параметрів навантаження на діючих мережах зв'язку. Середню кількість зайнятих серверів ще називають *завантаженістю системи*.

Оскільки навантаження є результатом спільного процесу надходження та обслуговування вимог, то *інтенсивність вхідного навантаження* Λ (не плутати з інтенсивністю потоку надходження вимог λ) нормується по відношенню до середнього часу обслуговування вимог \bar{x} . Тому ця інтенсивність визначається у відповідності до другого способу подання математичної моделі вхідного потоку вимог, але при цьому за умовну одиницю часу t береться середня тривалість обслуговування вимог \bar{x} . Отже, інтенсивність вхідного навантаження

$$\Lambda = \lambda \bar{x} \quad (8.10)$$

Інтенсивність навантаження вимірюється в *Ерлангах* (названо на честь засновника теорії телетрафіка А.К. Ерланга). Для випадку, показаного на рис. 8.3, величину $\lambda t = 10$ можна вважати інтенсивністю навантаження в 10 Ерл тільки за умови, якщо середня тривалість обслуговування $\bar{x} = t = 20$ с. Отже *інтенсивність вхідного навантаження* Λ , вимірювана в *Ерлангах*, це середня кількість вимог, що надходить до системи за середній час обслуговування однієї вимоги або інтенсивність потоку надходження вимог λ , віднесена до середньої тривалості обслуговування \bar{x} .

8.5 Характеристики якості обслуговування

Для будь-якої телекомунікаційної системи важливою є оцінка ступеня задоволення потреби в обслуговуванні, або якість обслуговування (*QoS – Quality of Service*). В теорії телетрафіка якість обслуговування потоку вимог характеризується можливістю негайного обслуговування вимоги або тривалістю очікування початку обслуговування. З математичної моделі СРІ випливає, що ці можливості визначаються обраною дисципліною обслуговування вимог. Через те для кожної дисципліни обслуговування вимог властивий певний набір основних і допоміжних характеристик якості обслуговування.

Системи з втратами.

З економічних причин СРІ можуть проектуватися з дисципліною обслуговування із втратами, за якої вимозі, що надходить до системи в момент відсутності вільних обслуговуючих пристроїв відмовляється в обслуговуванні і вона відразу ж залишає її та втрачається. Основною

всїх вимог $W = 1,611$, що перевищує відповідні значення для моделі $M/M/1/\infty$ усього на 3% (на стільки ж відмінність і значень ентропії). Такий же збіг основних параметрів якості обслуговування СМО з чергою спостерігається у всїх інших точках, для яких однакові значення ентропії розподілу станів системи, незалежно від закону розподілу тривалості обслуговування. Тому для цих розрахунків можна застосовувати формули Поллачека-Хінчина з табл. 8.1 моделі $M/G/1/\infty$.

Алгоритм застосування ентропійного методу розрахунку *QoS* такий:

1. Для встановленого закону розподілу станів системи визначається ентропія розподілу H_{JBM} (за відомими формулами).

2. Зміною коефіцієнта варіації v_x для моделі, наприклад $M/H/1/\infty$, досягається збіг значень ентропії $H_{MHL} = H_{JBM}$.

3. За допомогою знайденого коефіцієнта варіації v_x визначається середня кількість пакетів в системі N за формулою Поллачека-Хінчина (табл. 8.1).

4. Через відомі співвідношення визначаються інші характеристики *QoS*.

Таким чином, розрахунок параметрів *QoS* в моделі з самоподібним трафіком за будь-якого закону розподілу тривалості обслуговування здійснимий. Необхідною умовою такого розрахунку є визначення ентропії розподілу станів системи.

8.7.5 Система із необмеженою чергою $G/M/1/\infty$

В мультисервісних пакетних мережах зв'язку вхідні інформаційні потоки можуть мати постійну (CBR), змінну (VBR) і змішану бітову швидкість, від чого математична модель потоку може бути від найпростішої пуассонівської до складної моделі фрактальних процесів (самоподібний трафік). Закон розподілу інтервалу часу між вимогами в цих потоках може бути довільний і тому в узагальненій моделі резонно досліджувати узагальнений (G) вид розподілу випадкової величини цього інтервалу. Довжина пакетів кожної зі служб загальної для них мультисервісної мережі (отже, і тривалість обслуговування) може бути різною – для одних служб постійною, а для інших – змінною. У такому випадку також бажано досліджувати загальний вид розподілу випадкової величини тривалості обслуговування.

Системи з чергою типу $G/G/m/\infty$ є найважливішими моделями, що розглядаються у теорії телетрафіка. При їх дослідженні застосовувалися різні методи й отримані численні наближені результати. Окремий випадок – система з чергою й одним сервером ($m = 1$) розглядається, наприклад в [7], де отримано тільки приблизні результати. Однак дотепер у загальному випадку не існує досить простих і точних, безпосередньо застосовуваних на практиці формул для розрахунку характеристик *QoS* у стаціонарному режимі.

В багатьох випадках односерверної системи хорошим наближенням є експонентна функція розподілу тривалості обслуговування вимог. Тому визначимо характеристики *QoS* для моделі $G/M/1/\infty$ [14].

Коефіцієнт використання системи ρ визначається як відношення інтенсивності вхідного потоку вимог λ до інтенсивності обслуговування μ . Для односерверної системи цей коефіцієнт в межах $0 \leq \rho < 1$ і є інтенсивністю

видозмінюється потік вимог і відповідна функція розподілу імовірності інтервалу часу між вимогами, але не змінюється функція розподілу тривалості обслуговування. Зрозуміло, при $H=0,5$ потік втрачає властивості самоподібності, але в цьому випадку результати (8.95–8.97) мають корелюватися з результатами для деякої моделі з постійною тривалістю обслуговування, а не експонентною [12].

Випадковий процес (ВП) надходження пакетів до СРІ на обслуговування, утворює потік пакетів (трафік), який характеризується певним законом розподілу, що встановлює зв'язок між значенням випадкової величини (кількістю пакетів) і імовірністю появи цього значення. У більшості випадків для розрахунку параметрів QoS досить знати про закон розподілу тільки деякі його числові характеристики – моменти розподілу різних порядків. Для розрахунку в умовах пуассонівського розподілу достатньо математичного сподівання Λ , а для нормального розподілу – необхідні Λ і дисперсія D .

Основні характеристики випадкового процесу Λ і D , хоча й досить важливі, у той же час не є вичерпними, а іноді й марними для прогнозування значення випадкової величини. Іноді ВП характеризуються однаковими значеннями Λ і D , але внутрішня структура цих процесів різна. Одні можуть мати плавно мінливі реалізації, а інші – яскраво виражену коливальну структуру при стрибкоподібній зміні окремих значень випадкової величини (наприклад, різке зростання кількості пакетів в мережі, що призводить до „пачковості” трафіка). Для „плавних” процесів характерна велика передбачуваність реалізацій, а для „пачкових” – дуже мала імовірнісна залежність між двома випадковими величинами ВП. У таких випадках закон розподілу, що характеризує ВП, несе в собі деяку невизначеність і дозволяє з більшою або меншою надійністю прогнозувати значення випадкової величини.

Таким чином, використовувані імовірнісні закони розподілу, що описують трафік у пакетних мережах, не дають такої кількісної оцінки невизначеності стану системи масового обслуговування, як ентропія розподілу:

$$H(m) = - \sum_{j=1}^m P_j \log P_j.$$

Ентропія не залежить від значень, що набуває випадкова величина, а тільки від їхніх імовірностей.

Оцінки параметрів якості обслуговування самоподібного трафіка можлива ентропійним методом, який зводиться до використання методів розрахунку відомих розподілів, ентропія яких співпадає або найбільш близька до ентропії станів системи при обслуговуванні самоподібного трафіка [13].

Результатами моделювання встановлено, що в тих точках, де однакова ентропія розподілу станів системи, однакові й досліджувані параметри якості обслуговування, такі як середня довжина черги Q та середня тривалість очікування вимог W . Наприклад, для моделей $M/M/1/\infty$ і $fBM/D/1/\infty$ при коефіцієнті Херста $H=0,8$ та $\rho=0,6$ ентропії розподілів станів системи досить близькі і дорівнюють 1,683 та 1,719 відповідно. При цьому для моделі $fBM/D/1/\infty$ середня довжина черги $Q=0,982$ і середня тривалість очікування

кількісною оцінкою якості обслуговування в цьому випадку є імовірність втрати вимоги P_B (B – втрати, *blocking*). Імовірність P_B на відрізку часу (t_1, t_2) визначається як відношення кількості втрачених за цей відрізок часу вимог $C_B(t_1, t_2)$ до загальної кількості вимог, що надійшли за той же час до системи $C(t_1, t_2)$:

$$P_B = \frac{C_B(t_1, t_2)}{C(t_1, t_2)}. \quad (8.11)$$

Допоміжними характеристиками QoS є імовірність втрати навантаження (кількість вимог на умовну одиницю часу) та імовірність втрат за часом, які використовуються рідко. Імовірність втрати навантаження визначається як відношення інтенсивності втраченого навантаження до вхідного, а імовірність втрат за часом – це сумарна частка часу з проміжку часу (t_1, t_2) , за якої були зайняті всі сервери системи.

Середня кількість вимог у системі N характеризує ступінь завантаженості системи і співпадає з інтенсивністю обслуговуваного навантаження Y , тому що інші вимоги зі всіх, що надходять, втрачаються і до системи не потрапляють.

Середній час перебування вимоги в системі T співпадає з середнім часом обслуговування вимоги \bar{x} .

Системи з чергами.

Для кількісної оцінки якості обслуговування систем із чергою розраховують такі основні характеристики:

- імовірність очікування $P_{w>0}$ або середню частку затриманих вимог;
- середню довжину черги Q ;
- середню тривалість очікування для затриманих вимог t_q ;
- середню тривалість очікування для будь-якої вимоги W .

Імовірність $P_{w>0}$ на відрізку часу (t_1, t_2) визначається як відношення кількості вимог, що потрапили за цей відрізок часу до черги $C_Q(t_1, t_2)$ до загальної кількості вимог, що надійшли за той же час до системи $C(t_1, t_2)$:

$$P_{w>0} = \frac{C_Q(t_1, t_2)}{C(t_1, t_2)}. \quad (8.12)$$

Довжина черги є ключовим параметром якості обслуговування (та показником ефективності функціонування СРІ) і визначається кількістю вимог, що очікують обслуговування. Довжина черги залежить від того, коли і скільки вимог надійшло в систему, скільки часу витрачено на обслуговування вимог, що надійшли, і т.ін. Оскільки довжина черги є випадковою величиною, то як показник довжини черги використовується її математичне сподівання Q .

Середній час очікування в черзі t_q утворюється за рахунок затримки вимог в черзі. Він залежить від кількості вимог, що знаходяться в даний момент у черзі, часу закінчення обслуговування всіх попередніх вимог і т.ін.

Середній час очікування в системі W являє собою середнє значення часу очікування, віднесене до всіх вимог – затриманих і незатриманих. Цей параметр

вводиться через те, що не всі вимоги потрапляють до черги, а частина з них за наявності вільних серверів системи обслуговується негайно.

Допоміжними характеристиками QoS є середня кількість вимог у системі N та середній час перебування вимоги в системі T . Вони є допоміжними тому, що їх можна розрахувати із основних характеристик.

Середня кількість вимог у системі N визначає ступінь завантаженості системи і за необмеженої черги складається з середньої кількості вимог, що надходять до системи Λ , та тих, що очікують в черзі Q :

$$N = \Lambda + Q. \quad (8.13)$$

Середній час перебування вимоги в системі T – це час, проведений одною вимогою в системі й усереднений по всіх вимогах (затриманих і незатриманих). Він складається із середнього часу обслуговування \bar{x} і середнього часу очікування вимог у системі W :

$$T = \bar{x} + W. \quad (8.14)$$

Комбіновані системи (з чергами та втратами).

Система з чергою, в якій введено обмеження на максимальну кількість вимог, що можуть перебувати в черзі, або на максимальний час очікування початку обслуговування є системою з комбінованою дисципліною обслуговування. За обмеженої кількості місць очікування (максимальна довжина черги) у випадку надходження вимоги в момент, коли всі сервери і місця очікування зайняті попередніми вимогами, дана вимога втрачається. За обмеженого часу очікування, якщо вимога перебуває в черзі понад припустимий час, то їй відмовляється в обслуговуванні і вона теж втрачається. Тому окрім характеристик QoS чистої системи з чергами розраховуються й такі:

- імовірність втрати вимоги P_B (із-за обмеженої довжини черги);
- імовірність очікування понад припустимий час $P_{W>t}$.

Імовірність $P_{W>t}$ залежить від дисципліни обслуговування черги і найбільш простим для розрахунку є випадок упорядкованої черги $FIFO$. Цю імовірність ще називають *умовними втратами*, оскільки вимога, що очікує понад припустимий час t може втратити свою актуальність для користувача.

Пріоритетні системи.

Комбіновані системи з обмеженнями на довжину черги та час очікування є найбільш поширеними в телекомунікаціях. В умовах різномірного трафіка (мова, відео, дані) цю систему доповнюють механізмом пріоритетів, в якому всі вимоги поділяють на категорії і вимоги більш високої категорії при обслуговуванні мають певні переваги (пріоритети) перед вимогами більш низької категорії. Для кількісної оцінки якості обслуговування систем із пріоритетами розраховуються такі ж характеристики, як і для системи з чергами, але для кожного з введених пріоритетів окремо. Наприклад, середній час очікування в черзі вимоги k -го пріоритету.

В будь-якій з наведених СРІ якість обслуговування істотно впливає на такі характеристики системи, як пропускна здатність і продуктивність. При цьому критеріями якості обслуговування для систем із втратами є імовірність

В стані статистичної рівноваги при $\rho = \frac{\lambda}{\mu} < 1$ імовірність того, що кількість вимог у системі N перевищить задану величину x , представлено у виді функції:

$$\Pr(N > x) = \Pr\left(\sup_{t>0} \left(Z(t) - \frac{\mu - \lambda}{\sqrt{a\lambda}} t\right) > \frac{x}{\sqrt{a\lambda}}\right) = f\left(\left(\frac{\mu - \lambda}{\sqrt{a\lambda}}\right)^{(1-H)/H} \frac{\mu - \lambda}{\sqrt{a\lambda}}\right).$$

Для випадку, коли ця імовірність дорівнює наперед заданій величині $\Pr(N > x) = \varepsilon$, із вище наведеного випливає, що

$$\frac{1 - \rho}{\rho^H} \mu^{\frac{H-0.5}{H}} \frac{1-H}{x^H} = \frac{0.5}{f(\varepsilon)} = \text{const},$$

а це значить, що

$$x = \frac{(1 - \rho)^{\frac{H}{H-1}}}{\rho^{\frac{0.5}{H-1}}}. \quad (8.95)$$

Величину x знайдено з попереднього в припущенні, що $\text{const} = 1$. Імовірність, яка дорівнює одиниці – це достовірна подія і, отже, x – це кількість вимог у системі, яку не може бути перевищено, тобто це може бути верхня оцінка середньої кількості вимог N у системі $fBM/D/1/\infty$.

Оскільки з формули Літгла випливає, що $T = N/\lambda$, то середній час перебування вимоги в системі при $\mu = 1$, тобто в одиницях часу тривалості обслуговування визначається формулою:

$$T = \frac{\frac{H}{(1 - \rho)^{H-1}}}{\rho^{H-1}} \frac{1}{\rho} = \frac{\frac{H}{(1 - \rho)^{H-1}}}{\rho^{H-1}}. \quad (8.96)$$

Виходячи з того, що для будь-якої односерверної системи середня довжина черги $Q = N - \rho$, то з урахуванням (5) отримуємо:

$$Q = \frac{\frac{H}{(1 - \rho)^{H-1}}}{\rho^{H-1}} - \rho. \quad (8.97)$$

Результат (8.95), (8.96) і (8.97) вважається аналітичним рішенням для системи $fBM/D/1/\infty$. Однак при аналізі цього рішення можна помітити, що при завданні коефіцієнта Херста $H = 0,5$ (несамоподібний процес) маємо відомий результат для середньої кількості вимог, середньої тривалості перебування і середньої довжини черги в системі типу $M/M/1/\infty$. Це є нелогічним результатом, оскільки спочатку досліджувалася система з детермінованим часом обслуговування $fBM/D/1/\infty$. При зміні коефіцієнта Херста від значення $H = 1$ (максимальне значення) до $H = 0,5$ (мінімальне значення), безсумнівно,

– із (8.61) та (8.91) для заданих Λ і σ^2 визначається первинна імовірність $P_{w>0}$ (як би для випадку, коли вимоги із черги не йдуть в систему і не збільшують навантаження на неї);

– для розрахованих t_q і $P_{w>0}$ відповідно до (8.58) і (8.58) визначаються первинні значення W і Q ;

– для розрахованих із (8.93) і (8.94) значень Λ_2 та σ_2 відповідно до (8.61) і (8.91) визначається уточнена імовірність $P_{w>0}$, тобто з урахуванням впливу додаткового навантаження на сервери системи із черги. (При цьому довжина черги більш реальна, оскільки вимоги, що не йдуть із системи негайно, сприяють росту черги);

– за уточненим $P_{w>0}$ із (8.58) і (8.59) уточнюються значення W та Q .

Шляхом імітаційного моделювання встановлено, що реалізація цього алгоритму у великому діапазоні варіювання параметрів Λ , S і m дає завжди занижену оцінку імовірності $P_{w>0}$, однак при цьому відносна помилка ніколи не перевищує -10%. Тому, оскільки на останньому кроці знову уточнюються значення W і Q , то можна ще раз перерахувати $P_{w>0}$ з більш точними значеннями Λ_3 і σ_3 . Перевірка цього кроку показала, що результати розрахунків після третьої ітерації завжди дають верхню оцінку імовірності $P_{w>0}$, яка не перевищує +10% [11].

8.7.4 Система із необмеженою чергою $fBM/D/1/\infty$

Трафік мультисервісних мереж з комутацією пакетів характеризується наявністю довгострокових залежностей в інтенсивності навантаження й істотною відмінністю статистичних властивостей потоків пакетів від пуассонівського потоку. Адекватною моделлю потоків в таких мережах вважаються самоподібні (*self-similarity*) процеси, де вхідний потік описується фрактальним броунівським рухом (модель fBM). Однак дослідження характеристик якості обслуговування СРІ в цих умовах є дуже складною математичною задачею. Причиною цього є слабка формалізованість моделі самоподібних потоків, внаслідок чого й неможливо отримати аналітично обґрунтовані результати для оцінки параметрів QoS в системах розподілу інформації.

Для односерверної системи з нескінченною чергою та постійним часом обслуговування (модель $fBM/D/1/\infty$) наближене рішення наведено в [1], де показано, що кількість вимог у розглянутій системі в будь-який момент часу t може бути представлено випадковою величиною

$$N(t) = \sup_{s \leq t} (A(t) - A(s) - \mu(t - s)),$$

де

$$A(t) = \lambda t + \sqrt{a\lambda} Z(t).$$

Випадковий процес $Z(t)$ є нормалізованим фрактальним броунівським рухом з параметром Херста H ($H = 0.5, \dots, 1$), а позитивний коефіцієнт a є деяким множником масштабу.

втрати вимоги, а для систем з чергами – імовірність очікування. Чим більше припустима норма втрат, тим менша якість обслуговування.

Пропускна здатність – це максимальна інтенсивність навантаження, що може надходити до системи при забезпеченні заданої якості обслуговування. За малої імовірності втрат інтенсивність обслуженого навантаження близька до інтенсивності вхідного навантаження (пропускною спроможністю часто вважають інтенсивність обслуженого навантаження). Однак при великих втратах ця відмінність суттєва і тому, чим менша якість обслуговування, тим більшою буде пропускна спроможність (і навпаки). Крім того, чим вище норма якості обслуговування, тим більше серверів необхідно для забезпечення певної пропускної здатності.

Продуктивність – це гранична, статистично усереднена кількість вимог, які будуть обслужені системою в одиницю часу при заданій якості обслуговування. Ця характеристика використовується, як правило, для оцінки систем управління та керуючих пристроїв.

Пропускна здатність та продуктивність СРІ залежать не тільки від імовірності втрат, але й від структури системи (кількості серверів і схеми їх включення) та дисципліни обслуговування і закону розподілу тривалості обслуговування. Крім того, на пропускну здатність та якість обслуговування суттєво впливає вид потоку вимог або його математична модель.

8.6 Аналіз систем в умовах пуассонівського потоку вимог

Дослідження функціонування СРІ під впливом випадкових факторів можливе тільки за допомогою випадкових процесів. Випадковий процес – це функція $X(t)$, значення якої випадкові величини. Вибір випадкових процесів, використуваних для опису й аналізу систем, залежить від структури й типу системи, від припущення про незалежність або залежність випадкових величин в системі, від виду їхніх функцій розподілів.

Якщо всі функції розподілу, що характеризують поведінку елементів системи, експонентні, то систему можна описати за допомогою однорідних безперервних марковських ланцюгів або його підвиду однорідних процесів *розмноження та загибелі*, як моделі СРІ. При цьому аналітичне визначення величин, що характеризують систему, є відносно простим.

В умовах випадкових потоків вимог розрахунки ступеня завантаженості або пропускної здатності та характеристик якості обслуговування виконуються на основі імовірнісних функцій розподілу станів системи, які й визначають ці характеристики. За пуассонівського потоку вимог для визначення *стаціонарних* імовірностей станів системи використовується математичний апарат марковського процесу, який потребує складання та розв'язання системи лінійних алгебраїчних рівнянь рівноваги. *Стан системи* – це поточна кількість вимог системи, що обслуговуються в серверах або очікують в черзі. Зміна станів системи – це випадковий процес, який є результатом спільного процесу надходження та обслуговування вимог. Закон розподілу станів системи, на

відміну від середньої кількості вимог в ній (завантаженості), більш повно характеризує функціонування СРІ під впливом випадкових факторів.

Процес є марковським (без післядії), якщо для будь-якого моменту часу поведінка системи в майбутньому залежить тільки від теперішнього стану системи і не залежить від того, коли і як система цього стану дісталася. Марковський процес з *дискретним простором* станів називається ланцюгом Маркова. Множина $\{S\}$ утворює ланцюг Маркова, якщо імовірність того, що наступний стан буде S_{k+1} залежить тільки від поточного стану S_k або вплив всієї передісторії процесу цілком зосереджено на теперішньому стані.

Таким чином в ланцюзі Маркова потік випадкових величин визначається тільки імовірністю переходу від попереднього значення випадкової величини (стану системи) до наступного. Для *однорідного* ланцюга Маркова імовірності цих переходів не залежать від номера кроку, на якому система досягає певного стану. Крім того, згідно *теоремі Маркова*, якщо кількість станів системи S кінцева, і з кожного стану можна перейти, за певну кількість кроків в будь-який інший стан, то граничні імовірності станів існують і не залежать від початкового стану системи.

Умовою застосування ланцюгів Маркова для визначення імовірностей станів системи є експонентний розподіл інтервалів часу між вимогами потоку. Експонентний розподіл єдиний зі всіх неперервних законів, якому властива *відсутність післядії*, яка полягає в наступному: якщо тривалість інтервалу між вимогами експонентна, то з будь-якого моменту від початку інтервалу залишок часу до кінця інтервалу не залежить від подій до цього моменту і буде мати експонентний розподіл з тим же параметром, що і весь інтервал.

Оскільки експонентний розподіл інтервалів часу між вимогами призводить до пуассонівського розподілу кількості вимог на умовну одиницю часу, то властивість відсутності післядії переноситься й на пуассонівський потік вимог, який разом з тим є:

- *стаціонарним*, тому що імовірність кількості вимог на відрізок часу залежить тільки від тривалості цього відрізка і не залежить від того, де саме на осі часу він розташований;
- *ординарним*, тому що імовірність надходження більше однієї вимоги за нескінченно малий відрізок часу є нескінченно малою в порівнянні з імовірністю надходження точно однієї вимоги (це означає, що вимоги надходять тільки по одній);
- *без післядії*, тому що для будь-яких відрізків часу, що не перетинаються, кількість вимог одного відрізка не залежить від того, скільки вимог надійшло на інший.

Для такого потоку інтенсивність λ , тобто середня кількість вимог на одиницю часу, є величина незмінна.

Якщо в моделі не всі розподіли експонентні, шукають такі аналітичні прийоми, які призводять досліджувані процеси до *марковських*, або шукають такі моменти часу, в яких процес стає *марковським*. Далі простими методами дискретних *марковських* ланцюгів обчислюються шукані величини або

$$t_q \approx \frac{1}{m - \Lambda} \cdot S \cdot \frac{m}{m + \Lambda + 1 + \Lambda/m} = \frac{S}{(m + 1) \cdot [1 - (\Lambda/m)^2]} \quad (8.92)$$

Для розрахунку $P_{w>0}$ застосуємо такі аргументи. Імовірність $P_{w>0}$ – це імовірність зайнятості всіх m серверів (8.61).

Аналогічно до моделі $M/D/m/\infty$, при кінцевому числі m і необмеженій кількості місць очікування вимоги обслуговуються без втрат. На сервери системи надходять вимоги з первинного потоку з інтенсивністю Λ та із черги з інтенсивністю $\Lambda \cdot P_{>0} \cdot t_w$ (див. 8.48). Тому загальна інтенсивність навантаження на систему збільшується до величини $\Lambda_2 = \Lambda + Q$ (див. п.8.6.4).

В умовах гіперекспонентного потоку функція розподілу кількості вимог в системі (обслуговуються та у черзі) або станів системи P_j відрізняється від функції розподілу P_i кількості вимог, що надходять. На рис. 8.7 надано розподіли стану системи за гіперекспонентного потоку вимог з параметрами $\Lambda = 100$ Ерл та $S = 4$ для ємності системи $m = 105, 110$ та 120 серверів ($m > \Lambda$).

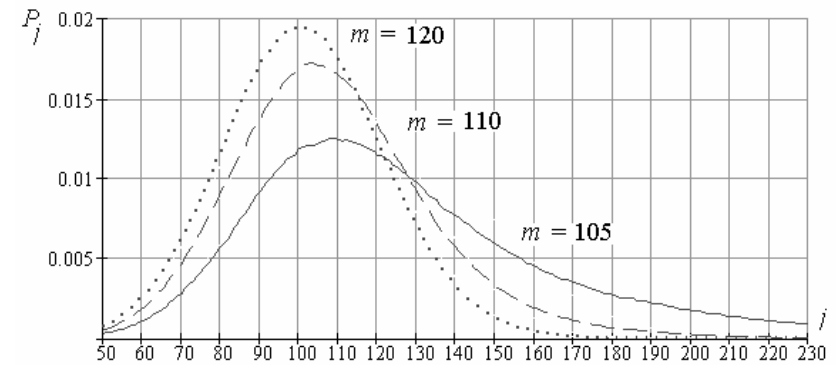


Рисунок 8.7 – Розподіл станів системи $H/D/m/\infty$

З рисунка видно, що при зменшенні m розкид окремих значень функції розподілу станів системи від середнього значення збільшується. З чого виходить, що додатковий потік вимог із черги збільшує загальну інтенсивність навантаження Λ_2 та її дисперсію σ_2^2 . Помітно, що вже при $m = 120$ (пунктирна лінія) функція розподілу станів системи доволі симетрична, що дозволяє її цілком апроксимувати нормальним законом розподілу. Апроксимація цих функцій нормальним законом (8.91) з параметрами

$$\Lambda_2 = \Lambda + Q, \quad (8.93)$$

$$\sigma_2 \approx \sigma + Q/2 \quad (8.94)$$

дає достатньо точні результати на відрізку, обумовленому межами підсумовування у (8.61), тобто від 0 до $m - 1$.

З цього виходить простий ітераційний алгоритм розрахунку основних характеристик якості обслуговування системи $H/D/m/\infty$:

- із (8.92) для заданих Λ, S та m розраховується t_q ;

8.7.3 Система із необмеженою чергою $H/D/m/\infty$

До повнодоступної m -серверної системи з необмеженою чергою надходить гіперекспонентний потік вимог з інтенсивністю Λ , пікфактором $S > 1$ та нормальним розподілом кількості вимог на одиницю часу x , де x – постійна тривалість обслуговування вимоги. Вибір із черги – в порядку надходження. Необхідно визначити основні характеристики QoS :

- імовірність очікування $P_{w>0}$;
- середню довжину черги Q ;
- середню тривалість очікування вимог у системі W ;
- середню тривалість очікування вимог у черзі t_q .

Розподіл імовірностей P_i випадкової кількості вимог i , що надходять до системи на одиницю часу x , описується *нормальним* законом розподілу:

$$P_i = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(i-\Lambda)^2}{2\sigma^2}}. \quad (8.91)$$

Незалежно від виду потоку вимог W та Q мають розраховуватись за формулами (8.58) та (8.59) на підставі визначення та формули Літтла відповідно. З цих формул видно, що дані характеристики QoS залежать від $P_{w>0}$ та t_q , які можуть бути визначені з функцій розподілу станів системи P_j та розподілу часу очікування вимог у черзі $P(t_q)$. Однак для непуассонівського потоку не існує загального методу отримання таких функцій і формули для них не є простими.

Для розрахунку t_q вживаємо такі, встановлені раніше, результати:

- з C -формули Ерланга виходить, що в системі $M/M/m/\infty$ середня тривалість очікування вимог у черзі $t_{q(M)} = 1 / (m - \Lambda)$;
- з формули Поллачека-Хінчина виходить, що в системі $M/D/1/\infty$ середня тривалість очікування вимог у черзі $t_{q(D)} = t_{q(M)} / 2$.

Очевидно, що в шуканому виразі для розрахунку t_q системи $H/D/m/\infty$ мають враховуватися дані наслідки. Перший – тому, що пуассонівський потік (M) є окремим випадком гіперекспонентного (H). Другий – тому, що одноструверна система ($m = 1$) є окремим випадком багатоструверної.

В [11] для системи $H/D/m/\infty$ показано, що $t_{q(D)} > t_{q(M)}$ у $S/$ ($k = 2$) разів за кількості серверів $m = \Lambda$. Це добре узгоджується з наведеними вище співвідношеннями – через пікфактор S враховано відмінність гіперекспонентного потоку від пуассонівського, і відмінність у два рази середньої тривалості очікування при постійній та експонентній тривалості обслуговування, але віднесено це до характерної точки $m = \Lambda$.

З ростом кількості серверів системи m коефіцієнт $k = 2$ убиває приблизно зі швидкістю $k(m) \approx (m + \Lambda) / m$. За результатами імітаційного моделювання системи $H/D/m/\infty$ встановлено, що точність розрахунку t_q підвищується при заміні даної залежності на $k(m) \approx (m + \Lambda + 1 + \Lambda / m) / m$. Остаточна формула для розрахунку t_q системи $H/D/m/\infty$ має вигляд:

імовірності, що характеризують стан системи, після чого вони перетворюються у відповідні величини вихідного процесу. Для розв'язання задач такого типу застосовуються методи напівмарковських процесів або вкладених ланцюгів Маркова. Однак реальні результати з застосуванням цих методів отримані для невеликого класу систем (наприклад, формула Поллачека-Хінчина для системи типу $M/G/1/\infty$), або отримано тільки часткові та наближені результати (для систем типу $M/G/m/\infty$, $GI/G/1/\infty$ та інших).

8.6.1 Система із втратами $M/M/m$

Повнодоступна схема СРІ з m серверами обслуговує пуассонівський потік вимог за дисципліною із втратами. Інтервал часу між вимогами та тривалість їх обслуговування мають експонентні закони розподілу. Параметр потоку вимог λ , а середнє значення тривалості обслуговування – \bar{x} . Необхідно визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0 \dots m$) та характеристики якості обслуговування (QoS).

Для ординарного потоку *інтенсивність надходження* вимог співпадає з параметром потоку λ . Сукупність моментів завершення обслуговування вимог утворюють потік звільнення серверів системи з параметром μ . Цей параметр називається *інтенсивністю обслуговування* вимог і розраховується як зворотна величина до тривалості обслуговування:

$$\mu = \frac{1}{\bar{x}}. \quad (8.15)$$

Дискретні стани системи S_k змінюються при кожній події надходження вимоги або закінченні її обслуговування. Ланцюг Маркова з кінцевою кількістю станів відображається у виді діаграми переходів (рис. 8.4).

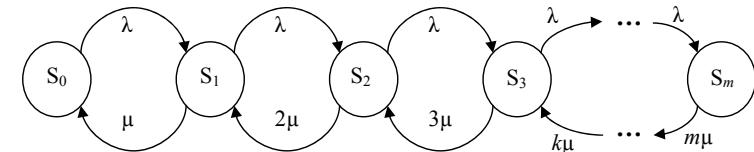


Рисунок 8.4 – Діаграма переходів m -серверної системи із втратами

Для пуассонівського потоку інтенсивність надходження вимог λ є величина незмінна. Якщо система знаходиться в стані S_k і надходить вимога, то система переходить з однаковою інтенсивністю λ в стан S_{k+1} за будь-якого k . В разі закінчення обслуговування вимоги інтенсивність, з якою система переходить із стану S_k в стан S_{k-1} залежить від кількості зайнятих серверів або поточного стану системи. Якщо, наприклад, система знаходиться в S_1 , то по закінченні обслуговування вимоги, яка займає один сервер, система переходить в стан S_0 з інтенсивністю μ . Якщо система в стані S_2 , то вона може перейти в стан S_1 з інтенсивністю вже 2μ , тому що може закінчитися обслуговування будь-якої з двох вимог, які займають два сервери системи, і т.ін. Таким чином, якщо обслуговуванням зайняті k серверів, то потік звільнення серверів, що переводить систему із станів S_k в S_{k-1} , буде в k раз інтенсивніше.

Стационарний розподіл імовірностей станів системи P_k для однорідного дискретного ланцюга Маркова визначається системою лінійних алгебраїчних рівнянь

$$P_k = \sum_i P_i p_{i,k}, \quad (8.16)$$

де $p_{i,k}$ – імовірність переходу системи зі стану i в стан k [5]. При цьому розподіл (8.16) має задовольняти умові нормування, за якої $\sum_{k=0}^m P_k = 1$.

Сума в (8.16) обчислюється для всіх можливих варіантів станів i , з яких можна перейти до стану k . За ординарного потоку в цей стан можна перейти тільки із станів $k-1$, k та $k+1$. Тому за рахунок зменшення кількості можливих імовірностей $p_{i,k}$ система рівнянь (8.16) скоротиться:

$$P_k = P_{k-1}p_{k-1,k} + P_k p_{k,k} + P_{k+1}p_{k+1,k}. \quad (8.17)$$

Перехід із стану $k-1$ в стан k можливий в разі надходження вимоги, імовірність чого пропорційна параметру λ , а перехід із $k+1$ в стан k можливий в разі закінчення обслуговування вимоги, імовірність чого за експонентного розподілу тривалості обслуговування x пропорційна параметру μ . Таким чином імовірність $p_{k-1,k} = \lambda$, а імовірність $p_{k+1,k} = \mu(k+1)$ (див. рис. 8.4).

Імовірність переходу системи із стану k в стан k очевидна – система не має піти із стану k ні в стан $k-1$, ні в стан $k+1$, бо її стан незмінний:

$$p_{k,k} = 1 - p_{k,k-1} - p_{k,k+1} = 1 - \mu k - \lambda. \quad (8.18)$$

Отже, із (8.17) з урахуванням (8.18) у підсумку маємо

$$P_k = \lambda P_{k-1} + P_k [1 - \mu k - \lambda] + \mu(k+1)P_{k+1}. \quad (8.19)$$

Після розкриття дужок і об'єднання послідовно отримуємо:

$$\begin{aligned} P_k &= \lambda P_{k-1} + P_k - \mu k P_k - \lambda P_k + \mu(k+1)P_{k+1}, \\ 0 &= \lambda P_{k-1} - (\lambda + \mu k)P_k + \mu(k+1)P_{k+1}, \end{aligned}$$

а після переносу:

$$(\lambda + \mu k)P_k = \lambda P_{k-1} + \mu(k+1)P_{k+1}. \quad (8.20)$$

Система рівнянь (8.20) описує стаціонарний режим в ланцюзі Маркова і є математичною моделлю процесу обслуговування вимог потоку. В цьому режимі виконується закон збереження інтенсивності імовірностей – інтенсивність потоку імовірностей в стан k дорівнює інтенсивності потоку із цього стану в стани $k-1$ або $k+1$. Тому (8.20) є рівнянням рівноваги або балансу.

Розв'язання рівняння балансу розпочнемо зі значення $k=0$, за якого із (8.20) отримуємо

$$\lambda P_0 = \lambda P_{-1} + \mu P_1.$$

Імовірність $P_{-1} \equiv 0$ як імовірність неіснуючого стану і тому $\lambda P_0 = \mu P_1$, звідки виходить, що

$$P_1 = \frac{\lambda}{\mu} P_0.$$

При $k=1$ з (8.20) отримуємо $(\lambda + \mu)P_1 = \lambda P_0 + 2\mu P_2$ і тому

В [10] показано, що регулярний закон розподілу тривалості обслуговування, з точки зору якості обслуговування, є найгіршим і тому формула (8.89) є не що інше, як верхня оцінка можливих значень імовірності відмови в обслуговуванні. Для інших законів розподілу тривалості обслуговування вимог до формули (8.89) застосовується спеціальна апроксимаційна функція, яка отримана за результатами імітаційного моделювання [9], і в якій коефіцієнтом h задається вид закону розподілу тривалості обслуговування:

$$P_B = \frac{S}{\sum_{k=0}^m \exp\left[\frac{-(k-2\Lambda+m)(k-m)}{2\Lambda S}\right]} \left[1 - \frac{(S^2-1)(\sigma-\Lambda+m)}{\sigma(S^2 h - h + 5)}\right], \quad (8.90)$$

де h – коефіцієнт, що дорівнює 4.25, 3.55 і 2.85 для рівномірного, експонентного і логарифмічно нормального законів розподілу тривалості обслуговування відповідно. Видно, що при $S=1$ (тобто $\sigma^2 = \Lambda$) дана формула перетворюється у формулу (8.88) для випадку $i=m$ – стан системи m (зайняті всі сервери).

На рис. 8.6 показані залежності імовірності втрат P_B від кількості серверів в системі $H/G^*/m$ та виду закону розподілу тривалості обслуговування при $\Lambda=100$ Ерл та $S=8$ (G^* – регулярний, рівномірний, експонентний та логарифмічно нормальний закони розподілу тривалості обслуговування).

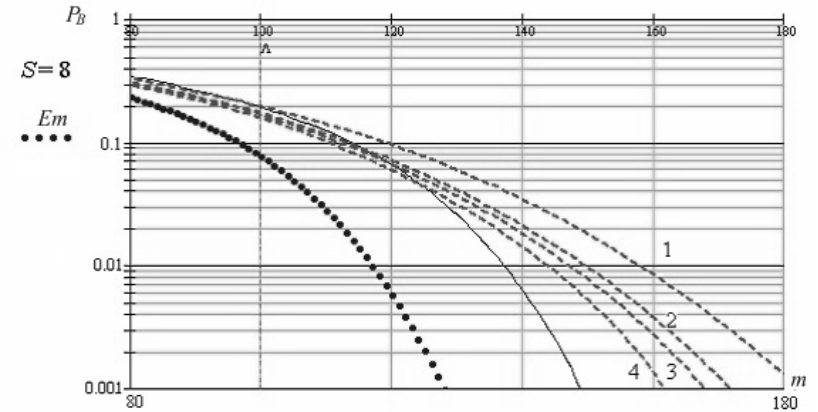


Рисунок 8.6 – Залежність P_B від m та закону розподілу тривалості зайняття

Пунктирна крива $E_m(\Lambda)$ представляє залежність втрат від m , розраховану за формулою Ерланга. Штриховими лініями 1, 2, 3 та 4 показана залежність імовірності втрат вимог для регулярного, рівномірного, експонентного та логарифмічно нормального законів розподілу тривалості зайняття відповідно.

З рисунка видно, що за умов значної ($S=8$) дисперсії інтенсивності навантаження краща якість обслуговування буде там, де в законі розподілу більша частка (імовірність) коротших за тривалістю обслуговування вимог.

система була у початковий момент (було прийнято, що в початковий момент усі сервери вільні). Відомо, що властивість ергодичності мають *марковські* процеси і для будь-якого такого процесу після досить тривалого часу функціонування системи обов'язково наступить стаціонарний режим, де імовірність того, що система буде в i -му стані, не залежить від того, у якому стані вона знаходилася в початковий момент часу. Відповідно до теореми Маркова умовою ергодичності процесу є (8.24) і тоді система працює в стані статистичної рівноваги.

Підставивши в (8.87) замість імовірностей $Q(H_i)$ щільність розподілу Пуассона (8.62), отримаємо перший розподіл Ерланга (8.25), справедливий для моделі $M/G/m$ і за яким можна розрахувати імовірності всіх станів системи. Однак у цьому випадку, неодмінно, математичне сподівання випадкового процесу надходження вимог з інтенсивністю Λ дорівнює її дисперсії σ^2 . Але, якщо при розрахунках (8.87) замість імовірностей $Q(H_i)$ і $Q(H_k)$ підставити імовірності, що розраховано за нормальним законом, то отримаємо формулу розрахунку станів системи саме в моделі $H/D/m$:

$$P_i = \frac{1}{\sqrt{2\pi\sigma}} e^{-(i-\Lambda)^2 / 2\sigma^2} = \frac{1}{\sum_{k=0}^m \frac{1}{\sqrt{2\pi\sigma}} e^{-(k-\Lambda)^2 / 2\sigma^2}} = \frac{1}{\sum_{k=0}^m \exp\left[\frac{-(k-2\Lambda+i)(k-i)}{2\sigma^2}\right]}. \quad (8.88)$$

При $\sigma^2 = \Lambda$ нормальний випадковий процес є *марковським* і при цьому відповідно до (8.24) при $\Lambda < m$ розрахункові значення, одержувані за формулами розподілу Ерланга (8.25) і (8.88) досить близькі – розбіжність не більше 1 %.

При $\sigma^2 > \Lambda$ розподіл інтервалів часу між вимогами вже не експонентний і потік вимог втрачає властивість відсутності післядії. Процес обслуговування стає складнішим і виникає залежність імовірностей станів системи від її початкового стану. Крім того, з'являється залежність і від виду розподілу тривалості обслуговування вимог, на відміну від розподілу Ерланга, правильному за будь-якого закону розподілу тривалості обслуговування з кінцевим середнім значенням. Через це імовірність заняття всіх серверів системи вже не збігається з імовірністю відмови в обслуговуванні, як це спостерігається за пуассонівського потоку.

8.7.2 Імовірність втрат системи $H/D/m$

В моделі $H/D/m$ імовірність відмови в обслуговуванні залежить від пікфактора інтенсивності навантаження S і за умови $\Lambda < m$ дорівнює імовірності зайняття всіх m серверів системи, помноженій на S :

$$P_B = \frac{1}{\sum_{k=0}^m \exp\left[\frac{-(k-2\Lambda+m)(k-m)}{2\sigma^2}\right]} \frac{\sigma^2}{\Lambda}. \quad (8.89)$$

$$P_2 = \frac{\lambda}{2\mu} P_1 = \frac{\lambda}{2\mu} \frac{\lambda}{\mu} P_0 = \left(\frac{\lambda}{\mu}\right)^2 \frac{1}{2 \cdot 1} P_0.$$

При $k = 2$ з (8.20) отримуємо $(\lambda + 2\mu)P_2 = \lambda P_1 + 3\mu P_3$ і тому

$$P_3 = \frac{\lambda}{3\mu} P_2 = \frac{\lambda}{3\mu} \frac{\lambda}{2\mu} \frac{\lambda}{\mu} P_0 = \left(\frac{\lambda}{\mu}\right)^3 \frac{1}{3 \cdot 2 \cdot 1} P_0.$$

Без продовження з наведених розрахунків (індукції) видно, що

$$P_k = \frac{\lambda}{\mu k} P_{k-1}, \quad (8.21)$$

та

$$P_k = \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 \quad (k = 1, 2, \dots, m). \quad (8.22)$$

З урахуванням умови нормування, внесеної до стаціонарного розподілу імовірностей станів системи (8.16) імовірність P_0 визначиться так:

$$P_0 = \left[\sum_{i=0}^m \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} \right]^{-1}. \quad (8.23)$$

Розподіл (8.22) є шуканим стаціонарним розподілом імовірностей станів повнодоступної системи з втратами, яка обслуговує пуассонівський потік вимог. Дана задача вперше вирішена А.К. Ерлангом в припущенні про експонентний розподіл тривалості обслуговування, але згодом доведено, що (8.22) вірно за будь-якого довільного закону, тобто й для моделі $M/G/m$.

Тут досліджується сталий режим системи, який буде на нескінченному відрізку часу, коли система обслуговування працює в стані статистичної рівноваги. Якщо даний процес є *ергодичним*, то отримані імовірності станів системи (кількості зайнятих серверів) не залежать від того, у якому стані система була у початковий момент. За теоремою Маркова будь-який *транзитивний* (з будь-якого стану можна перейти в будь-який інший) *однорідний* (імовірності переходу зі стану в стан не залежать від того, у який момент часу початок переходу) процес з кінцевою кількістю станів має ергодичну властивість [6, с. 149]. Для будь-якого марковського процесу, що має ергодичну властивість, після досить тривалого часу функціонування системи наступить стаціонарний режим, де імовірність того, що система буде в k -му стані, не залежить від того, у якому стані вона знаходилася в початковий момент часу. Відповідно до зазначеного в ергодичній теоремі [7] критерієві ергодичності процесу, процес буде ергодичним, якщо виконується умова

$$\frac{\lambda}{\mu} < m. \quad (8.24)$$

Це значить, що в середньому повинно надходити менше вимог, ніж їх можна обслужити або середня кількість вимог на одиницю часу обслуговування не повинна перевищувати кількості серверів.

З урахуванням (8.10) і (8.15) з'ясовується, що параметром стаціонарного розподілу імовірностей станів системи (8.22) і (8.23) є інтенсивність вхідного навантаження Λ , і таким чином можна записати:

$$P_k = \frac{\Lambda^k}{k!} \quad (k = 1, 2, \dots, m). \quad (8.25)$$

$$\sum_{i=0}^m \frac{\Lambda^i}{i!}$$

Розподіл (8.25) називається *першим розподілом Ерланга*, за яким можна розрахувати імовірність зайнятості k серверів повнодоступної m -серверної системи з втратами при обслуговуванні пуассонівського потоку вимог.

Для стану $k = m$ з (8.25) отримуємо *B-формулу Ерланга*, що позначається $E_m(\Lambda)$, і за якою можна розрахувати основну характеристику якості обслуговування потоку вимог – імовірність втрати вимоги P_B (8.11). Для спрощення обчислень значення $E_m(\Lambda)$ наводяться в таблицях Ерланга.

Фактично при $k = m$ з (8.25) отримуємо значення імовірності зайняття всіх серверів системи або імовірності втрат за часом (див. п.8.5). Чергова вимога, що надходить до системи, буде втрачена (заблокована) в тому випадку, коли всі сервери зайняті. Але тільки за умов пуассонівського потоку значення імовірності зайняття всіх серверів системи співпадає з імовірністю втрати вимоги незалежно від того, надходить в цей момент вимога або ні. Із-за властивості відсутності післядії інтенсивність надходження вимог не залежить від стану системи.

8.6.2 Система із необмеженою чергою M/M/m/∞

Повнодоступна схема СРІ з m серверами та необмеженою кількістю місць очікування в черзі обслуговує пуассонівський потік вимог з дисципліною обслуговування черги *FIFO*. Інтервал часу між вимогами та тривалість їх обслуговування мають експонентні закони розподілу. Параметр потоку вимог λ , а середнє значення тривалості обслуговування – \bar{x} . Визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0, \dots, \infty$) та характеристики QoS .

Дискретні стани системи S_k змінюються при кожній події надходження вимоги або закінченні її обслуговування. Ланцюг Маркова з кінцевою кількістю станів відображається у виді діаграми переходів (рис. 8.5).

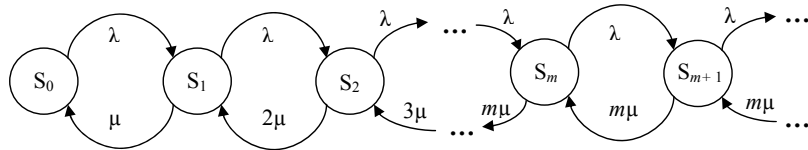


Рисунок 8.5 – Діаграма переходів m -серверної системи з чергою

Параметр потоку звільнення серверів системи μ розраховується за формулою (8.15). Інтенсивність потоку звільнення для станів системи

m). Простір P , безсумнівно, менше простору Q , оскільки стани $i > m$ для системи неможливі, а j може бути як завгодно великим.

Оскільки система обслуговування із втратами, то незалежно від її поточного стану для кожного з випадків надходження $j > m$ вимог відбувається подія «відмова в обслуговуванні». Це очевидно, оскільки за час \bar{x} (постійна тривалість зайняття серверів) надійде $j > m$ вимог і жоден зі знову зайнятих серверів за цей же час не звільниться. Якщо система перебуває в початковому стані $i = 0$ (усі сервери вільні), то в цьому випадку для кожного з варіантів надходження за той же відрізок часу $j \leq m$ вимог подія «відмова в обслуговуванні» не відбувається. Події, які полягають в надходженні за час \bar{x} точно j вимог, утворюють повну групу неспільних гіпотез H_0, H_1, \dots, H_j з апіорними імовірностями $Q(H_0), Q(H_1), \dots, Q(H_j)$ і тому $\sum_{j=0}^{\infty} Q(H_j) = 1$. Подія A ,

яка полягає у «відсутності відмови в обслуговуванні» може відбуватися тільки разом з однією із гіпотез групи H_0, H_1, \dots, H_m . Умовні імовірності гіпотез надходження $j \leq m$ вимог (апостеріорні), за умови здійснення події A (відсутність втрат вимог) обчислюються за формулою Байеса:

$$Q(H_j | A) = \frac{Q(H_j)Q(A|H_j)}{\sum_{k=0}^m Q(H_k)Q(A|H_k)}$$

Через те що подія A (відсутність відмови в обслуговуванні), яка з'являється з кожною із гіпотез групи H_0, H_1, \dots, H_m є достовірною, то умовні імовірності $Q(A|H_j) = 1$. Тому, якщо для здійснення події можлива тільки частина гіпотез, а інші неможливі, то для одержання апостеріорних імовірностей треба кожному з апіорних імовірностей цієї частини можливих гіпотез розділити на їхню суму. Отже:

$$Q(H_j | A) = \frac{Q(H_j)}{\sum_{k=0}^m Q(H_k)}$$

Для системи, що знаходиться в початковому стані (усі сервери вільні), умовні імовірності $Q(H_j|A)$ відповідають імовірностям зайняття $i = j$ серверів системи, тобто імовірностям P_i . Вони також утворюють повну групу неспільних подій підпростору P і тому $\sum_{i=0}^m P_i = 1$. Таким чином, імовірності зайняття серверів P_i представлено через імовірності надходження вимог $Q(H_j)$, тобто

$$P_i = \frac{Q(H_i)}{\sum_{k=0}^m Q(H_k)}. \quad (8.87)$$

Умовою стаціонарності отриманого розподілу є ергодичність процесу обслуговування вимог. Це значить, що отримані імовірності станів системи (кількості зайнятих серверів) не повинні залежати від того, в якому стані

того, до якої групи за інтенсивністю навантаження належить джерело і яке співвідношення чисельності цих джерел з іншими. Отже, більш адекватно описати такий потік або розподіл інтервалів часу між вимогами можна не експонентним розподілом, а їхньою сумішшю, називаною *гіперекспонентним розподілом (H)*. Йому відповідає перервний пуассонівський потік n -го порядку. Практичні виміри свідчать, що реальні потоки достатньо апроксимувати з $n = 2$. Щільність гіперекспонентного розподілу при цьому така:

$$p(z) = p_1 \lambda_1 e^{-\lambda_1 z} + p_2 \lambda_2 e^{-\lambda_2 z}.$$

Отже, з імовірністю p_1 інтервал часу між вимогами має експонентний розподіл з параметром λ_1 , а з імовірністю p_2 – з параметром λ_2 (природно, що $p_1 + p_2 = 1$). Цей розподіл описує більший розкид величини інтервалу часу між вимогами z і забезпечує значення коефіцієнтів $v_z \geq 1$, а це, у свою чергу, дозволяє описувати потоки з дисперсією σ^2 інтенсивності навантаження (кількості вимог в одиницю часу), що перевищує її математичне сподівання Λ в десятки разів. Співвідношення цих перших двох моментів інтенсивності навантаження визначає пікфактор (скупченість) інтенсивності навантаження S :

$$S = \frac{\sigma^2}{\Lambda}. \quad (8.86)$$

Для реальних потоків, як правило, $S > 1$. Крім того, гіперекспонентний розподіл інтервалу часу між вимогами призводить до нормального розподілу кількості вимог за середню тривалість обслуговування вимоги.

Отже для реальних потоків вимог мультисервісних мереж зв'язку більш прийнятною математичною моделлю є модель з нормальним розподілом інтенсивності навантаження Λ з пікфактором $S > 1$.

Визначення основної характеристики QoS – імовірності відмови в обслуговуванні через зайнятість всіх серверів системи – базується на визначенні імовірнісної функції розподілу станів системи P_i .

8.7.1 Функція розподілу станів системи із втратами $H/D/m$

Повнодоступна схема СРІ з m серверами функціонує за дисципліною із втратами. Тривалість обслуговування постійна й дорівнює \bar{x} . До системи надходить гіперекспонентний потік вимог з інтенсивністю Λ , пікфактором $S > 1$ та нормальним розподілом кількості вимог на одиницю часу \bar{x} . Визначити стаціонарний розподіл імовірностей станів системи P_k ($k = 0 \dots m$).

Стан системи визначається кількістю зайнятих серверів. Випадковий процес надходження вимог модифікує стан системи зі швидкістю, обумовленою інтенсивністю навантаження Λ . Інтенсивність навантаження Λ – це сумарна кількість вимог j , які надійшли за час, що дорівнює середній тривалості обслуговування \bar{x} . Таким чином, інтенсивність переходу системи з одного стану в інший залежить від властивостей потоку вимог, що описується відповідним розподілом імовірностей Q_j , де j – кількість вимог за час \bar{x} , яка може бути в діапазоні від 0 до ∞ . Усі події надходження вимог належать просторові станів Q . Події зайняття серверів утворюють новий дискретний підпростір P , який описано розподілом імовірностей P_i , де i – кількість зайнятих серверів ($i = 0, \dots,$

$k = 0, \dots, m$ залежить від k і визначається аналогічно до її визначення на діаграмі переходів рис. 8.3 – вона є змінною та дорівнює $k\mu$. Для станів системи $k = m, \dots, \infty$ ця інтенсивність незмінна і дорівнює $m\mu$, оскільки при цьому вже є черга і зміна її стану залежить від звільнення будь-якого зі всіх m серверів.

Тому визначення стаціонарних імовірностей станів системи P_k складається з двох частин:

- для станів системи $k = 0, \dots, m$ – кількість зайнятих серверів або вимог, що обслуговуються;
- для станів системи $k = m, \dots, \infty$ – кількість зайнятих місць очікування черги, або вимог, що очікують.

Розподіл кількості зайнятих серверів визначається за алгоритмом формул (8.16) – (8.22) і для станів $k = 0, \dots, m$ маємо

$$P_k = \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 \quad (k = 1, 2, \dots, m). \quad (8.26)$$

Для станів $k = m, \dots, \infty$ рівняння балансу системи подібне до (8.20), але за незмінної інтенсивності потоку звільнення серверів, що дорівнює $m\mu$, маємо:

$$(\lambda + m\mu)P_k = \lambda P_{k-1} + m\mu P_{k+1}. \quad (8.27)$$

Звідси, якщо для станів $k = 0, \dots, m$ аналогічно до (8.21)

$$P_k = \frac{1}{k} \frac{\lambda}{\mu} P_{k-1}, \quad (8.28)$$

то для станів $k = m, \dots, \infty$

$$P_k = \frac{1}{m} \frac{\lambda}{\mu} P_{k-1}. \quad (8.29)$$

Результат розв'язання рівняння (8.29) очевидний. Шляхом послідовної підстановки в (8.29) значень $k = m + 1$; $k = m + 2$; $k = m + 3$ і т.д. перевіряємо:

$$P_{k=m+1} = \frac{\lambda}{m\mu} P_m = \left(\frac{\lambda}{m\mu}\right)^{k-m} P_m;$$

$$P_{k=m+2} = \frac{\lambda}{m\mu} P_{m+1} = \frac{\lambda}{m\mu} \frac{\lambda}{m\mu} P_m = \left(\frac{\lambda}{m\mu}\right)^{k-m} P_m;$$

$$P_{k=m+3} = \frac{\lambda}{m\mu} P_{m+2} = \frac{\lambda}{m\mu} \frac{\lambda}{m\mu} \frac{\lambda}{m\mu} P_m = \left(\frac{\lambda}{m\mu}\right)^{k-m} P_m,$$

а це (індукція) підтверджує, що для будь-якого $k = m, \dots, \infty$ правильне

$$P_k = \left(\frac{\lambda}{m\mu}\right)^{k-m} P_m. \quad (8.30)$$

Оскільки з (8.26) для P_m маємо

$$P_m = \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} P_0, \quad (8.31)$$

то у підсумку з (8.30) і (8.31) для станів $k = m, \dots, \infty$ виходить, що

$$P_k = \left(\frac{\lambda}{m\mu}\right)^{k-m} \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} P_0, \quad (k = m, \dots, \infty). \quad (8.32)$$

Рівняння (8.26) і (8.32) остаточно визначають стаціонарний розподіл імовірностей станів системи P_k для всіх можливих значень k або станів серверів та місць очікування черги. Імовірність P_0 для цього випадку визначається з умови нормування $\sum_{k=0}^{\infty} P_k = 1$, звідки підстановкою всіх значень P_k формул (8.26) та (8.32) дістаємо (доданки, що не залежать від k винесено за знак другої суми):

$$P_0 \left[\sum_{k=0}^{m-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} \sum_{k=m}^{\infty} \left(\frac{\lambda}{m\mu}\right)^{k-m} \right] = 1. \quad (8.33)$$

В цій формулі друга сума від $k = m$ до ∞ існує тільки за умови $(\lambda / m\mu) < 1$ (збіжність ряду за ознакою Даламбера) і тому вона може бути розрахованою за відомою формулою суми членів нескінченно спадної геометричної прогресії:

$$\sum_{k=m}^{\infty} \left(\frac{\lambda}{m\mu}\right)^{k-m} = \sum_{k=0}^{\infty} \left(\frac{\lambda}{m\mu}\right)^k = \frac{1}{1 - \frac{\lambda}{m\mu}} = \frac{m\mu}{m\mu - \lambda}. \quad (8.34)$$

Підставивши цей вираз в (8.33) отримуємо

$$P_0 = \left[\sum_{k=0}^{m-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!} \frac{m\mu}{m\mu - \lambda} \right]^{-1}. \quad (8.35)$$

Використана в (8.33) умова збіжності ряду $(\lambda / m\mu) < 1$ співпадає з (8.24), яка є умовою ергодичності процесу. Це в свою чергу не дозволяє нескінченно зростати черзі, що є запорукою нормального функціонування всієї системи.

Оскільки $\Lambda = \lambda / \mu$ (тому з (8.24) обов'язково $\Lambda < m$), то рівняння (8.26), (8.32) і (8.35), що остаточно визначають стаціонарний розподіл імовірностей станів P_k , можна записати в одну систему рівнянь так:

$$P_k = \frac{\Lambda^k}{k!} P_0 \quad k = 1 \dots m \quad (8.36)$$

$$P_k = \left(\frac{\Lambda}{m}\right)^{k-m} \frac{\Lambda^m}{m!} P_0 \quad k = m \dots \infty \quad (8.37)$$

$$P_0 = \left[\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \frac{m}{m - \Lambda} \right]^{-1} \quad (8.38)$$

Це є *другий розподіл Ерланга*, за яким можна розрахувати імовірність станів (зайнятості серверів та місць очікування в черзі) повнодоступної системи з необмеженою чергою при обслуговуванні пуассонівського потоку вимог.

Основною характеристикою якості обслуговування цієї системи є імовірність очікування $P_{w>0}$ ($w > 0$ означає, що час очікування ϵ), яка може бути визначена з функції розподілу станів системи P_k . В умовах пуассонівського

Таблиця 8.1 – Характеристики якості обслуговування моделі $M/G/1/\infty$

Хар-ка QoS	Модель		
	$M/G/1/\infty$	$M/D/1/\infty$	$M/M/1/\infty$
$P_{w>0}$	ρ	ρ	ρ
Q	$\frac{\rho^2(1+v_x^2)}{2(1-\rho)}$	$\frac{\rho^2}{2(1-\rho)}$	$\frac{\rho^2}{1-\rho}$
W	$\frac{\rho(1+v_x^2)}{2(1-\rho)}$	$\frac{\rho}{2(1-\rho)}$	$\frac{\rho}{1-\rho}$
t_q	$\frac{1+v_x^2}{2(1-\rho)}$	$\frac{1}{2(1-\rho)}$	$\frac{1}{1-\rho}$
N	$\rho + \frac{\rho^2(1+v_x^2)}{2(1-\rho)}$	$\rho + \frac{\rho^2}{2(1-\rho)}$	$\rho + \frac{\rho^2}{1-\rho}$
T	$1 + \frac{\rho(1+v_x^2)}{2(1-\rho)}$	$1 + \frac{\rho}{2(1-\rho)}$	$1 + \frac{\rho}{1-\rho}$

До табл. 8.1, з урахуванням того, що для постійної та експонентної тривалості обслуговування коефіцієнт її варіації $v=0$ та $v=1$ відповідно, занесені й значення всіх характеристик QoS моделей $M/D/1/\infty$ та $M/M/1/\infty$. З таблиці видно, що за постійної тривалості обслуговування значення W , t_q та Q в два рази менше, ніж за експонентної. Наведені в табл. 8.1 значення характеристик QoS моделі $M/M/1/\infty$ відповідно збігаються з формулами (8.41), (8.46), (8.47) та (8.48).

8.7 Аналіз систем в умовах реального потоку вимог

В математичній моделі пуассонівського потоку вимог інтервал часу між вимогами z_i розподілений за експонентним законом. Ступінь відхилення інших потоків від моделі пуассонівського потоку можна оцінити за коефіцієнтами варіації v_z , асиметрії Sk і ексцесу Ex функції розподілу інтервалу z . Для експонентного розподілу $v_z \equiv 1$, $Sk \equiv 2$ і $Ex \equiv 6$. Модель пуассонівського потоку не завжди адекватно описує реальні потоки вимог телекомунікаційних мереж і тому необхідно вибирати інші розподіли для їхнього опису, що забезпечують кращу згоду з даними вимірів. Заміна експонентного розподілу будь-якими іншим функціями значно ускладнює математичну модель, а складні моделі не завжди піддаються аналітичному рішенню.

Реальні потоки вимог в мультисервісних мережах формуються безліччю джерел з різною питомою інтенсивністю навантаження. В процесі створення потоку вимог беруть участь джерела, що належать до тієї або іншої служби (групи) із близькими значеннями інтенсивності навантаження. Значення інтенсивності результуючого потоку вимог в кожний момент часу залежить від

2. Середній час очікування у системі W .

Нехай $W(t)$ – стаціонарна функція розподілу часу очікування і $W^*(s)$ – її перетворення Лапласа:

$$W^*(s) = \int_0^{\infty} e^{-st} dW(t).$$

Нехай V – функція розподілу часу перебування в стаціонарному стані; час перебування дорівнює часу чекання плюс час обслуговування. Оскільки час чекання і наступна тривалість обслуговування стохастично незалежні, то для перетворення Лапласа $V^*(s)$ часу перебування в стаціонарному стані на підставі теореми згортки (множення) маємо:

$$V^*(s) = W^*(s)B^*(s). \quad (8.81)$$

Кількість вимог, що перебувають в системі, коли її залишає деяка вимога, дорівнює кількості вимог, що надходять до системи за час перебування в системі вимоги, що йде з неї. Тому між функцією розподілу часу перебування $V(t)$ і розподілом P_j існує такий зв'язок:

$$P_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dV(t). \quad (8.82)$$

Для твірної функції $P^*(z)$ розподілу (P_i) із (8.82) з урахуванням (8.81) та (8.74) отримуємо:

$$P^*(z) = \sum_{j=0}^{\infty} z^j P_j = \int_0^{\infty} e^{-\lambda t(1-z)} dV(t) = V^*(\lambda - \lambda z) = W^*(\lambda - \lambda z)B^*(\lambda - \lambda z).$$

З цього та із (8.79) виходить, що

$$W^*(s) = \frac{P^*(1-s/\lambda)}{B^*(s)} = \frac{(1-\rho)s}{s - \lambda + \lambda B^*(s)}. \quad (8.83)$$

Оскільки математичне сподівання $W = -W^*(0)$, то для стаціонарного стану отримуємо

$$W = \frac{\lambda \mu^2}{2(1-\rho)} = \frac{\lambda(\bar{x}^2 + \sigma_x^2)}{2(1-\rho)}. \quad (8.84)$$

Для характеристики другого центрального моменту розподілу тривалості обслуговування σ_x^2 часто використовується коефіцієнт варіації v_x :

$$v_x = \frac{\sigma_x}{\bar{x}}. \quad (8.85)$$

Суму $\bar{x}^2 + \sigma_x^2$ з урахуванням (8.85), можна представити як $\bar{x}^2(1 + v_x^2)$. Підставивши це до (8.80) і (8.84) отримуємо остаточні формули розрахунку N та W . З урахуванням (8.13), (8.14), (8.58) та (8.59) отримуємо всі інші характеристики QoS , формули яких занесені до табл. 8.1. Значення W , t_q та T нормуються по \bar{x} .

потоків $P_{w>0}$ дорівнює імовірності зайнятості всіх m серверів системи з урахуванням всіх можливих станів черги (m серверів зайнято і 0 вимог в черзі, m серверів зайнято і 1 вимога в черзі, 2 вимоги в черзі і т.д., а тому не плутати із станом P_m):

$$P_{w>0} = \sum_{k=m}^{\infty} P_k. \quad (8.39)$$

де k – стан системи ($0 < k \leq m$ – сервери, $m < k \leq \infty$ – черга).

Підстановкою (8.37) в (8.39) з урахуванням властивості (8.34) отримуємо:

$$P_{w>0} = \sum_{k=m}^{\infty} P_k = \frac{\Lambda^m}{m!} P_0 \sum_{k=m}^{\infty} \left(\frac{\Lambda}{m}\right)^{k-m} = \frac{\Lambda^m}{m!} \frac{m}{m-\Lambda} P_0. \quad (8.40)$$

Підстановка (8.38) в (8.40) дає

$$P_{w>0} = \frac{\frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}}{\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}}. \quad (8.41)$$

Формула (8.41) називається *C-формулою Ерланга* і позначається $C_m(\Lambda)$. Для системи з необмеженою чергою, що обслуговує пуассонівський потік вимог, за нею можна розрахувати імовірність очікування $P_{w>0}$ (8.12).

Її можна переписати так:

$$P_{w>0} = \frac{\frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}}{\sum_{k=0}^m \frac{\Lambda^k}{k!} - \frac{\Lambda^m}{m!} + \frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}} = \frac{\frac{\Lambda^m}{m!} \frac{m}{m-\Lambda}}{\sum_{k=0}^m \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \left(\frac{m}{m-\Lambda} - 1\right)}.$$

Якщо поділити чисельник і знаменник на $\sum_{k=0}^m \frac{\Lambda^k}{k!}$, то з урахуванням формули (8.25) та наступним скороченням отримуємо:

$$C_m(\Lambda) = \frac{E_m(\Lambda) \frac{m}{m-\Lambda}}{1 + E_m(\Lambda) \left(\frac{m}{m-\Lambda} - 1\right)} = \frac{E_m(\Lambda)m}{m - \Lambda[1 - E_m(\Lambda)]}. \quad (8.42)$$

Формула (8.42) встановлює взаємозв'язок між B - та C -формулами Ерланга, а це дозволяє розраховувати значення імовірності $C_m(\Lambda)$ за значеннями $E_m(\Lambda)$ B -формули Ерланга (8.25), які наводяться в таблицях Ерланга.

Середня довжина черги Q (середня кількість вимог в черзі) може бути знайдена із розподілу імовірностей станів системи P_k при $k = m+1, m+2, \dots, \infty$ (в черзі 1, 2 і т.д. вимог) за правилом обчислення математичного сподівання:

$$Q = \sum_{k=m+1}^{\infty} (k-m)P_k. \quad (8.43)$$

Підставивши з (8.37) імовірності станів черги P_k масмо (доданки, що не залежать від k винесено за знак суми):

$$Q = \frac{\Lambda^m}{m!} P_0 \sum_{k=m+1}^{\infty} (k-m) \left(\frac{\Lambda}{m}\right)^{k-m}.$$

Відомо, що $\sum_{i=0}^{\infty} ix^i = \frac{x}{(1-x)^2}$, а тому

$$Q = \frac{\Lambda^m}{m!} P_0 \frac{\frac{\Lambda}{m}}{\left(1 - \frac{\Lambda}{m}\right)^2}. \quad (8.44)$$

З (8.31) і (8.40) виходить, що імовірність стану P_m дорівнює:

$$P_m = \left(1 - \frac{\Lambda}{m}\right) P_{w>0}. \quad (8.45)$$

Підставивши до (8.44) замість імовірності $P_0 = P_m \left(\frac{\Lambda^m}{m!}\right)^{-1}$, яку отримано з

формули (8.31), а замість P_m її значення з (8.45), після скорочень маємо

$$Q = \frac{\Lambda}{m - \Lambda} P_{w>0}. \quad (8.46)$$

Середню тривалість очікування для будь-якої вимоги W отримаємо із формули Літгла – за W одиниць часу очікування до черги надійде в середньому $Q = \Lambda W$ вимог, а отже:

$$W = \frac{Q}{\Lambda} = \frac{1}{m - \Lambda} P_{w>0}. \quad (8.47)$$

Середню тривалість очікування для затриманих вимог t_q отримаємо із тлумачення величини Q , як інтенсивності навантаження, яке створюється вимогами, що очікують в черзі. Величина $\Lambda P_{w>0}$ – це інтенсивність потоку вимог з черги, де кожна вимога в середньому очікує час t_q . Отже $Q = \Lambda P_{w>0} t_q$, звідки з урахуванням (8.46):

$$t_q = \frac{Q}{\Lambda P_{w>0}} = \frac{1}{m - \Lambda}. \quad (8.48)$$

Значення W та t_q подано в одиницях середньої тривалості обслуговування. Такий нормований вид показує, в скільки разів ці значення зростають або убавують в порівнянні із середнім часом обслуговування \bar{x} .

Таким чином всі основні характеристики QoS знайдено. Середня кількість вимог у системі N визначається за формулою (8.13), а середній час перебування вимоги в системі T – за формулою (8.14).

8.6.3 Система із обмеженою чергою $M/M/m/r$

Повнодоступна схема СРІ з m серверами та обмеженою кількістю місць очікування в черзі r обслуговує пуассонівський потік вимог з дисципліною

тому маємо

$$P^*(z) = P_0 \sum_{j=0}^{\infty} k_j z^j + \sum_{j=0}^{\infty} z^j \sum_{i=1}^{j+1} P_i k_{j-i+1} = P_0 K^*(z) + \sum_{i=1}^{\infty} P_i z^{i-1} K^*(z).$$

Після скорочень отримуємо

$$P^*(z) = P_0 K^*(z) + \frac{K^*(z)}{z} (P^*(z) - P_0) = \frac{P_0(1-z)K^*(z)}{K^*(z) - z}. \quad (8.77)$$

У (8.77) входить невідома поки константа P_0 . Її можна визначити за допомогою наступних обчислень, типових при роботі з твірними функціями.

Оскільки $\lim_{z \rightarrow 1} P^*(z) = \lim_{z \rightarrow 1} K^*(z) = 1$, а

$$\lim_{z \rightarrow 1} \frac{K^*(z) - z}{1 - z} = \lim_{z \rightarrow 1} \left(\frac{K^*(z) - 1}{1 - z} + \frac{1 - z}{1 - z} \right) = K^{*'}(1) + 1 = 1 - \rho,$$

то з (8.77) при $z \rightarrow 1$ випливає рівняння: $1 = P_0 / (1 - \rho)$, відкіля

$$P_0 = 1 - \rho \quad (8.78)$$

$$P^*(z) = \frac{(1 - \rho)(1 - z)K^*(z)}{K^*(z) - z} = \frac{(1 - \rho)(1 - z)}{1 - \frac{z}{B^*(\lambda - \lambda z)}}. \quad (8.79)$$

Формулу (8.79) називають співвідношенням Поллачека-Хінчина, яке задає твірну функцію стаціонарного розподілу вкладеного марковського ланцюга в точках, коли вимоги залишають систему $M/G/\infty$, як функцію від перетворення Лапласа функції розподілу тривалості обслуговування.

Крок 4. Перерахуємо отримані на кроці 3 результати у шукані величини системи: 1. Середня кількість вимог в системі 2. Середній час очікування в системі (в тому числі і для точок, відмінних від t_n).

1. Середня кількість вимог у системі N .

В розглянутій моделі $M/G/\infty$ імовірності того, що в стаціонарному стані надходячи до системи вимога застає в ній j інших вимог, дорівнюють стаціонарним імовірностям P_j наявності в системі j вимог безпосередньо після закінчення обслуговування деякої вимоги і тому вони задаються формулою (8.79). Через властивість пуассонівського потоку вимог ці імовірності дорівнюють стаціонарним імовірностям P_j у будь-який момент часу.

Математичне сподівання (середнє значення) кількості вимог N , що знаходяться в системі, дорівнює:

$$N = P^{*'}(1) = \rho + \frac{\lambda^2 \mu_x^2}{2(1 - \rho)}, \quad (8.80)$$

де

$$\mu_x^2 = \int_0^{\infty} t^2 dB(t) = \bar{x}^2 + \sigma_x^2.$$

Формулу (8.80) називають формулою Поллачека-Хінчина.

$$P_{i,j} = \begin{cases} k_j & \text{при } j=0,1,2,\dots, i=0, \\ k_{j-i+1} & \text{при } j>i-1, i=1,2,\dots, \\ 0 & \text{в інших випадках} \end{cases} \quad (8.71)$$

Оскільки вимоги надходять за пуассонівським законом з інтенсивністю λ , то для них імовірність того, що на інтервалі часу тривалістю t надійде точно j вимог, дорівнює (8.7), і тому

$$k_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t), \quad j=0,1,\dots \quad (8.72)$$

Твірна (рос. – производящая) функція $K^*(z)$ розподілу (k_j) за визначенням:

$$K^*(z) = \sum_{j=0}^{\infty} k_j z^j. \quad (8.73)$$

де z – комплексна змінна. Із (8.73) з урахуванням (8.72) і згідно з [5, с. 205]

$$K^*(z) = B^*(\lambda - \lambda z). \quad (8.74)$$

Оскільки $B(t)$ є функцією розподілу тривалості обслуговування, то отримане з (8.72) рівняння встановлює взаємозв'язок між твірною функцією K^* дискретного розподілу випадкової величини k_j (кількості вимог за час обслуговування) і перетворенням Лапласа B^* щільності неперервного розподілу випадкової величини x (тривалості обслуговування) в точці $(\lambda - \lambda z)$. Перша похідна твірної функції в точці $z=1$ та перетворення Лапласа в точці $z=0$ дозволяє обчислити перші моменти випадкової величини, що розглядається. Тому з цього для даного рівняння виходить, що

$$K^{*'}(1) = -\lambda B^{*'}(0) = \lambda \bar{x} = \rho. \quad (8.75)$$

В (8.75) останнє рівняння формується з (8.10) та (8.55).

Крок 3. Обчислимо стаціонарний розподіл для вкладеного ланцюга Маркова. Вкладений марковський ланцюг неприводимий, аперіодичний та у випадку $\rho < 1$ – ергодичний (доведено за допомогою теореми Фостера). Тому існує точно один стаціонарний початковий розподіл (P_j) ланцюга, де P_j – імовірність того, що в стаціонарному стані вимога, що покидає систему, після себе залишає j вимог. Тому імовірності P_j задовольняють системі рівнянь, аналогічній (8.16):

$$P_j = \sum_i P_i P_{i,j}, \quad (8.76)$$

тобто:

$$P_0 = k_0(P_0 + P_1); \quad P_j = P_0 k_j + \sum_{i=0}^{j-1} P_i k_{j-i-1}, \quad j=1,2,\dots$$

Аналогічно до (8.73) твірна функція $P^*(z)$ розподілу $P = (P_j)$

$$P^*(z) = \sum_{j=0}^{\infty} z^j P_j, \quad |z| < 1,$$

обслуговування черги *FIFO*. Інтервал часу між вимогами та тривалість їх обслуговування мають експонентні закони розподілу. Інтенсивність вхідного навантаження Λ . Необхідно визначити стаціонарний розподіл імовірностей станів системи P_k ($k=0, \dots, m+r$) та характеристики *QoS*.

У разі обмеженої черги, де кількість місць очікування $r=n$ ($n < \infty$), отримуємо комбіновану систему $M/M/m/r$ – з чергами та втратами. Якщо в момент надходження вимоги в системі є хоча б один вільний сервер, то вона негайно починає обслуговуватись. Якщо ж всі сервери системи зайняті, то вимога стає до черги за умови, що в ній очікує менше, як r вимог (є місця для очікування). Якщо ж в черзі є вже r вимог, що надійшли раніше, то поточна вимога втрачається. Таким чином, вимозі відмовляється в обслуговуванні у тому разі, коли в системі знаходяться $l=m+r$ вимог. З цих l вимог m обслуговуються, а r очікують своєї черги.

Очевидно, якщо $r=0$, то система описується моделлю $M/M/m$ і задача вирішується за допомогою першого розподілу Ерланга (8.25). Якщо $r=\infty$, то система описується моделлю $M/M/m/\infty$ і задача зводиться до попередньої.

Для комбінованої системи $M/M/m/r=n$, яка частково є аналогічною до системи $M/M/m/\infty$, рівняння (8.36) і (8.37) визначають стаціонарний розподіл імовірностей станів системи P_k для $k=1, \dots, m$ та $k=m, \dots, l$ відповідно. Тому імовірність P_0 визначається з умови нормування $\sum_{k=0}^l P_k = 1$, звідки підстановкою всіх значень P_k формул (8.36) та (8.37) дістаємо

$$P_0 \left[\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \frac{\Lambda^m}{m!} \sum_{k=m}^l \left(\frac{\Lambda}{m} \right)^{k-m} \right] = 1. \quad (8.49)$$

Оскільки сума від $k=m$ до l може бути розрахованою за формулою суми $r=l-m$ членів геометричної прогресії зі знаменником (Λ/m) , то

$$P_0 = \left[\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \left(\frac{\Lambda^m}{m!} \right) \frac{1 - \left(\frac{\Lambda}{m} \right)^{r+1}}{1 - \frac{\Lambda}{m}} \right]^{-1}. \quad (8.50)$$

Імовірність повної зайнятості системи P_{3H} (очікування або втрати вимоги) у цьому випадку визначається аналогічно (8.40), але сума має бути не до ∞ , а тільки до l . Замінивши цю суму так само, як у формулах (8.49) і (8.50), маємо

$$P_{3H} = \sum_{k=m}^l P_k = \frac{\Lambda^m}{m!} P_0 \sum_{k=m}^l \left(\frac{\Lambda}{m} \right)^{k-m} = \frac{\Lambda^m}{m!} P_0 \frac{1 - \left(\frac{\Lambda}{m} \right)^{r+1}}{1 - \frac{\Lambda}{m}}. \quad (8.51)$$

Підстановка P_0 (8.50) в (8.51) дає

$$P_{\text{зн}} = \frac{\left(\frac{\Lambda}{m}\right)^{r+1} \frac{1 - \left(\frac{\Lambda}{m}\right)^{r+1}}{1 - \frac{\Lambda}{m}}}{\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \left(\frac{\Lambda}{m}\right)^{r+1} \frac{1 - \left(\frac{\Lambda}{m}\right)^{r+1}}{1 - \frac{\Lambda}{m}}}. \quad (8.52)$$

Очевидно, якщо $r = \infty$, то формула (8.50) співпадає з (8.38), а формула (8.52) – з C -формулою Ерланга (8.41). Якщо $r = 0$, то (8.50) співпадає з (8.23), а формула (8.52) – з B -формулою Ерланга (8.25). Отже (8.52) є узагальнюючим рішенням для повнодоступних систем з втратами, з необмеженою та обмеженою чергами за умови обслуговування пуассонівського потоку вимог.

В системі з обмеженою до r чергою подія втрати вимоги відбудеться у тому випадку, коли при надходженні вимоги у черзі вже є r вимог, що надійшли раніше. Отже імовірність втрати вимоги P_B дорівнює імовірності того, що в системі вже є $l = m + r$ вимог або система в стані P_l .

Оскільки $r = l - m$, то відповідно до (8.37) та (8.50):

$$P_B = P_l = \left(\frac{\Lambda}{m}\right)^{l-m} \frac{\Lambda^m}{m!} P_0 = \frac{\left(\frac{\Lambda}{m}\right)^r \frac{\Lambda^m}{m!}}{\sum_{k=0}^{m-1} \frac{\Lambda^k}{k!} + \left(\frac{\Lambda}{m}\right)^{r+1} \frac{1 - \left(\frac{\Lambda}{m}\right)^{r+1}}{1 - \frac{\Lambda}{m}}}. \quad (8.53)$$

Окремо імовірність очікування за обмеженою до r черги $P_{w>0}(r)$ або частку вимог, які очікують, можна визначити так: $P_{w>0}(r) = P_{\text{зн}} - P_B$, що дасть формулу, подібну до (8.52), але в чисельнику буде не $r + 1$, а тільки r .

З урахуванням (8.51) можна записати:

$$P_B = \left(\frac{\Lambda}{m}\right)^{l-m} \frac{\Lambda^m}{m!} P_0 = P_{\text{зн}} \frac{\left(\frac{\Lambda}{m}\right)^r \left(1 - \frac{\Lambda}{m}\right)}{1 - \left(\frac{\Lambda}{m}\right)^{r+1}}. \quad (8.54)$$

Якщо позначити:

$$\rho = \frac{\Lambda}{m}, \quad (8.55)$$

де ρ – інтенсивність питомого навантаження (навантаження на один сервер), то (8.54) можна переписати:

$$P_B = P_{\text{зн}} \frac{\rho^r (1 - \rho)}{1 - \rho^{r+1}}. \quad (8.56)$$

випадкового процесу, що описує поведінку розглянутої моделі, конструюється зручний марковський ланцюг, досліджуваний аналітичними методами, звичайними для ланцюгів Маркова, і отримані результати (наприклад, стаціонарні імовірності станів) перераховуються для величин вихідної системи.

Нехай досліджувана модель приймає стани $0, 1, 2, \dots$, а g_i – кількість вимог, що перебувають в системі в момент часу t або g_i є стан системи в момент часу t ; (g_i) – не марковський ланцюг. Визначимо характеристики QoS – середні значення кількості вимог і часу очікування в системі.

Метод вкладених ланцюгів складається з наступних чотирьох кроків:

Крок 1. Визначимо послідовності моментів часу (t_n) , $0 < t_1 < t_2 < \dots < \infty$, що дозволяє конструювання вкладеного ланцюга Маркова. Зрозуміло, якщо всі розподіли експонентні, то (g_i) – однорідний марковський ланцюг і в якості (t_n) можна вибрати будь-яку зростаючу послідовність моментів часу $0 \leq t_n < \infty$. Якщо не всі розподіли, що характеризують поведінку елементів системи, експонентні, то (g_i) – не марковський ланцюг, і при цьому коли тільки один розподіл не експонентний, то стан системи в момент часу t можна описати за допомогою величини ξ^*_i :

$$\xi^*_i = (g_i, r_i), \quad (8.70)$$

де ξ^*_i – перетворення Лапласа неперервної функції розподілу ξ_i , за допомогою якого шляхом диференціювання можна визначити відповідні моменти цього розподілу.

Для додаткової змінної r_i покладається таке: r_i дорівнює залишковому часові неекспонентно розподіленої величини в момент часу t , якщо такий існує (наприклад залишковий час обслуговування, залишковий час паузи); $r_i = 0$, якщо відсутній такий залишковий час (наприклад, не обслуговується жодна вимога; немає паузи з неекспонентним розподілом).

Значення t_n визначається в такий спосіб. Залишковим часом r_i є залишкова тривалість обслуговування. Вона буде дорівнювати нулеві, якщо деяка вимога залишає систему. Тому покладається, щоб t_n дорівнювало моментів часу, коли n -а вимога залишає систему (момент виходу).

Сконструйований за допомогою додаткової змінної r_i двовимірний стохастичний процес (ξ^*_i) розглядається тільки в певних точках t_n , тобто в моменти виходу вимог із системи, а тому він є однорідним марковським і розподіл цього процесу можна визначити.

Крок 2. Обчислимо перехідні імовірності для вкладеного ланцюга Маркова. Нехай ξ_n – кількість вимог, що знаходяться в системі безпосередньо після моменту t_n , тобто $\xi_n = g_{t_n+0}$. При цьому (ξ_n) – однорідний марковський ланцюг. Нехай p_{ij} – незалежна від n імовірність переходу системи в момент виходу із неї вимоги:

$$p_{i,j} = \Psi(\xi_{n+1} = j | \xi_n = i), \quad i, j = 0, 1, 2, \dots$$

Позначимо через k_j імовірність того, що за час обслуговування деякої вимоги в систему надходить j нових вимог, тоді

$$t_{q(D)} = \frac{W_{(D)}}{P_{w>0}} = \frac{C_m(\Lambda) \cdot \frac{1}{m-\Lambda} \cdot 2^1 \left(\frac{m}{m+\Lambda}\right)^2}{C_m(\Lambda) \cdot 2^{-1} \left(\frac{m}{m+\Lambda}\right)^{-1}} = \frac{1}{m-\Lambda} 2^2 \left(\frac{m}{m+\Lambda}\right)^3. \quad (8.65)$$

Аналізуючи формули (8.60), (8.64) і (8.65) помічаємо, що однойменні характеристики QoS в моделях $M/M/m/\infty$ і $M/D/m/\infty$ при експонентній і постійній тривалості обслуговування пов'язані між собою наступною апроксимуючою функцією:

$$F(k) = 2^{k-1} \left(\frac{m}{m+\Lambda}\right)^k, \quad (8.66)$$

де $k = 1, 2$ і 3 для характеристик $P_{w>0}$, $W_{(D)}$ та $t_{q(D)}$ відповідно.

Таким чином, шляхом застосування апроксимуючої функції (8.66) можуть бути розраховані основні характеристики якості обслуговування в моделі $M/D/m/\infty$ через C -формулу Ерланга, справедливу для моделі $M/M/m/\infty$:

$$P_{w>0} = \frac{C_m(\Lambda)}{2 \cdot F(k)}, \quad W_{(D)} = \frac{C_m(\Lambda)}{m-\Lambda} \cdot F(k+1), \quad t_{q(D)} = \frac{1}{m-\Lambda} F(k+2) \quad (8.67)$$

де $k = 1$ для всіх наведених характеристик.

Імітаційним моделюванням встановлено, що при заміні в (8.67) коефіцієнта $k = 1$ на $k = 0,01\Lambda$ (може бути не цілим) точність оцінки основних характеристик якості обслуговування підвищується до $\pm 5\%$.

8.6.5 Система із необмеженою чергою $M/G/1/\infty$

Односерверна система із необмеженою кількістю місць очікування обслуговує пуассонівський потік вимог. Тривалість обслуговування вимог має довільний (будь-який) закон розподілу $B(t)$ з математичним сподіванням \bar{x} . Інтенсивність потоку вимог λ . Визначити характеристики QoS .

У разі не експонентного розподілу тривалості обслуговування є тільки один експонентний розподіл моделі – інтервал часу між вимогами в потоці вимог. Тому дана задача вирішується методом вкладених ланцюгів Маркова.

Поняття вкладеного ланцюга Маркова зводиться до наступного. Нехай (g_i) – випадковий процес з безперервним часом, що набуває тільки цілих числових значень з деякої безлічі G , і існує послідовність t_1, t_2, \dots (теж випадкова) моментів часу така, що процес (ξ_n)

$$\xi_n = g_{t_n}, \quad n = 1, 2, \dots, \quad (8.68)$$

є однорідним марковським ланцюгом з дискретним часом, тобто для $n = 2, 3 \dots$ має місце:

$$\begin{aligned} \Psi(\xi_n = j | \xi_{n-1} = i, \xi_{n-2} = i_{n-2}, \dots, \xi_1 = i_1) = \\ \Psi(\xi_n = j | \xi_{n-1} = i) = \Psi(\xi_2 = j | \xi_1 = i) \end{aligned} \quad (8.69)$$

де $j, i, i_1 \dots \in G$.

Процес (ξ_n) є марковським ланцюгом, вкладеним в (g_i) . Основний зміст методу вкладених ланцюгів Маркова полягає в тому, що для даного

3 (8.56) та урахуванням (8.30), де $k = l$ та $r = l - m$, виходить, що

$$P_{zn} = P_l \frac{1 - \rho^{r+1}}{\rho^r (1 - \rho)} = P_m \frac{1 - \rho^{r+1}}{1 - \rho}. \quad (8.57)$$

Таким чином всі основні характеристики якості обслуговування в системі з обмеженою чергою знайдено.

8.6.4 Система із необмеженою чергою $M/D/m/\infty$

В телекомунікаційних системах, що базуються на технології комутації пакетів, застосовується дисципліна обслуговування з чергами. Є системи, в яких час обслуговування постійний, наприклад, процес обслуговування вимог керуючими пристроями вузлів комутації або обслуговування пакетів однакової довжини. В моделі $M/M/m/\infty$ з експонентним часом обслуговування вимог (пакетів) імовірність очікування розраховується за допомогою C -формули Ерланга (8.41). Для моделі $M/D/m/\infty$ за дисципліни черги $FIFO$ рішення отримано К.Д. Кроммеліном [3], в якому використано рівняння станів Фрая. Система таких рівнянь розв'язується методом похідних функцій. Ці обчислення настільки складні, що на практиці для інженерних розрахунків замість точних аналітичних формул застосовуються відповідні діаграми (криві Кроммеліна).

В [8] запропонований більш простий метод розрахунку моделі $M/D/m/\infty$, в якому використовується C -формула Ерланга.

Необхідно визначити основні характеристики якості обслуговування QoS :

- середню тривалість очікування вимог у системі W ;
- середню довжину черги Q ;
- середню кількість вимог у системі N ;
- імовірність очікування $P_{w>0}$;
- середню тривалість очікування вимог у черзі t_q .

Основні характеристики якості обслуговування моделі $M/M/m/\infty$ перебувають між собою в певній функціональній залежності. Середня тривалість очікування для будь-якої вимоги W (що очікує і що не очікує) є середнім значенням часу очікування, віднесеним до всіх вимог. Якщо ж відома середня тривалість очікування тільки затриманих вимог t_q , то для визначення W необхідно помножити цю тривалість на імовірність $P_{w>0}$, яка показує середню частку затриманих вимог. Отже з (8.48) та (8.47) отримуємо:

$$W = t_q P_{w>0}. \quad (8.58)$$

За формулою Літтла за W одиниць часу очікування до черги надійде ΛW вимог, отже

$$Q = \Lambda W. \quad (8.59)$$

З наведених співвідношень характеристик QoS виходить, що для аналізу моделі $M/M/m/\infty$ з експонентним розподілом тривалості обслуговування необхідно лише розрахувати середню тривалість очікування вимог у черзі t_q (8.48) та імовірність очікування $P_{w>0}$ (8.41). При цьому середня тривалість очікування вимог $W_{(M)}$ (M – експонента) визначається з (8.47).

Для розрахунку $P_{w>0}$ в моделі $M/D/m/\infty$ з постійною тривалістю обслуговування викладений в п.8.6.2 алгоритм визначення станів статистичної рівноваги системи застосовувати неможливо. Тепер імовірність закінчення обслуговування вимоги, яка в разі експонентного розподілу \bar{x} пропорційна m , визначається не кількістю зайнятих серверів m (рис. 8.5), а моментами початку обслуговування, і тому буде залежати від розташування досліджуваного інтервалу на осі часу – процес не буде ергодичним (див. п. 8.6.1).

Коли точне теоретичне рішення виявляється складним, то нерідко досить одержати наближений вираз, на якому може ґрунтуватися весь метод розрахунку. Визнаним методом аналітичних наближень є імітаційне моделювання, яке дозволяє ефективно перевіряти якість використовуваних припущень. За результатами імітаційного моделювання з використанням алгоритму [9] встановлено, що час очікування W за постійної (D) та експонентної (M) тривалості обслуговування перебуває в такому співвідношенні:

$$W_{(D)} = W_{(M)} 2 \left(\frac{m}{m + \Lambda} \right)^2 = C_m(\Lambda) \frac{1}{m - \Lambda} 2 \left(\frac{m}{m + \Lambda} \right)^2. \quad (8.60)$$

Формула (8.59) визначає, що для багатоканальної системи при $m = \Lambda$ середня тривалість очікування за постійної та експонентної тривалості обслуговування відрізняються в 2 рази, що відповідає такому ж співвідношенню, встановленому формулою Поллачека-Хінчина для односерверної системи. Зі зростанням ємності системи m це співвідношення убуває і імітаційне моделювання свідчить, що в широкому діапазоні зміни m та Λ точність оцінки середньої тривалості очікування (8.60) не гірше $\pm 10\%$.

Імовірність очікування $P_{w>0}$ визначається з функції розподілу станів системи P_k і тут $P_{w>0}$ дорівнює імовірності того, що при надходженні вимоги вона застає всі m серверів зайнятими (8.39). Це можна записати так:

$$P_{w>0} = \sum_{k=m}^{\infty} P_k = 1 - \sum_{k=0}^{m-1} P_k. \quad (8.61)$$

В умовах необмеженої кількості серверів ($m = \infty$) вимоги обслуговуються без втрат. За постійної тривалості обслуговування, коли немає втрат, властивості потоку звільнень збігаються з властивостями потоку надходження вимог, тому що відбувається тільки зсув у часі на величину \bar{x} між моментом надходження вимоги й моментом закінчення її обслуговування. При цьому стан системи повністю визначається властивостями потоку вимог, а функції розподілу кількості вимог у системі P_k і кількості вимог P_i , що надійшли за час \bar{x} , цілком збігаються з розподілом Пуассона

$$P_k = P_i = \frac{\Lambda^i}{i!} \cdot e^{-\Lambda}. \quad (8.62)$$

При кінцевому значенні m і необмеженій кількості місць очікування вимоги також обслуговуються без явних втрат. Однак у цьому випадку вимоги, що надходять після заняття всіх серверів системи, попадають у чергу на очікування, і у випадку звільнення хоча б одного з m зайнятих серверів відразу

подаються із черги на обслуговування. Тепер на канали системи надходять вимоги з первинного потоку з інтенсивністю Λ та із черги з інтенсивністю ΛW .

Таким чином, з урахуванням (8.59) сумарна інтенсивність навантаження на канали системи збільшується до величини

$$\Lambda_{\Sigma} = \Lambda + Q. \quad (8.63)$$

Інтенсивність навантаження Λ_{Σ} дорівнює середній кількості вимог у системі N (8.13).

З наведених аргументів випливає простий метод розрахунку основних характеристик якості обслуговування системи $M/D/m/\infty$, який складається з шести кроків:

1. За (8.42) для заданих Λ й m розраховується $C_m(\Lambda)$ для моделі $M/M/m/\infty$.
2. За (8.60) для заданих Λ й m визначається $W_{(D)}$ (апроксимація).
3. За (8.59) для Λ і розрахованого значення $W_{(D)}$ визначається Q .
4. За (8.63) розраховується сумарна інтенсивність навантаження Λ_{Σ} .
5. За (8.61) і (8.62) для Λ_{Σ} розраховується імовірність очікування $P_{w>0}$.
6. За розрахованими $W_{(D)}$ й $P_{w>0}$ з формули (8.58) визначається t_q .

Слід зазначити, що при m близьких до Λ , тобто в діапазоні ємностей системи $\Lambda < m \leq (\Lambda + \sqrt{\Lambda}/2)$, додатковий потік вимог із черги суттєво збільшує не тільки сумарну інтенсивність навантаження Λ_{Σ} , але і її сумарну дисперсію σ_{Σ}^2 , що для пуассонівського потоку дорівнює $\sqrt{\Lambda}$. Тому в даному діапазоні ємностей системи для підвищення точності розрахунку $P_{w>0}$ та t_q рекомендується на кроці 5 замість розподілу Пуассона (8.62) використати розподіл нормального закону (Гаусса) з підставленням (8.63) Λ_{Σ} та $\sigma_{\Sigma}^2 = \sqrt{\Lambda} + Q/2$. Для виключення нескінченної черги обов'язково $m > \Lambda$.

Оцінку ступеня точності запропонованого методу розрахунку характеристик якості обслуговування моделі $M/D/m/\infty$ перевірено імітаційним моделюванням, результати якого [9] засвідчили, що в широкому діапазоні зміни m і Λ відносна помилка розрахунку всіх характеристик QoS не перевищує $\pm 10\%$.

Результатами імітаційного моделювання з такою ж точністю оцінки визначене й співвідношення імовірностей очікування при експонентній і постійній тривалості обслуговування в моделях $M/M/m/\infty$ і $M/D/m/\infty$:

$$P_{w>0} \approx C_m(\Lambda) \cdot \frac{1}{2 \cdot 2^0 \left(\frac{m}{m + \Lambda} \right)^1}, \quad (8.64)$$

де $P_{w>0}$ – імовірність очікування при постійній, а $C_m(\Lambda)$ – при експонентній тривалості обслуговування.

З (8.58) виходить, що імовірність очікування за будь-яких умов – це співвідношення середніх тривалостей очікування в системі W та черзі t_q . Тому з (8.60) і (8.64) маємо середню тривалість очікування вимог у черзі $t_{q(D)}$ при постійній тривалості обслуговування: