

**ОСНОВИ НАУКОВИХ
ДОСЛІДЖЕНЬ**

ПІДРУЧНИК

Харків 2012



УКРАЇНСЬКА ДЕРЖАВНА АКАДЕМІЯ
ЗАЛІЗНИЧНОГО ТРАНСПОРТУ

ОСНОВИ НАУКОВИХ ДОСЛІДЖЕНЬ

ПІДРУЧНИК

*Затверджено Міністерством освіти і науки, молоді
та спорту України як підручник для студентів
вищих навчальних закладів*

Харків 2012

ISBN 978-966-2033-82-3

© Українська державна академія
залізничного транспорту, 2012.

УДК 621.396.6
ББК 32.841:72.4(2)
О 753

*Затверджено Міністерством освіти і науки, молоді
та спорту України як підручник для студентів
вищих навчальних закладів (№ 1/11-4032 від 26.03.12 р.)*

Авторський колектив:

Г.В. Альошин, С.В. Лістровий,
С.В. Панченко, С.І. Приходько

Рецензенти:

Н.Н. Горобець, д-р физ.-мат. наук, професор
(ХНУ ім. В.М. Каразіна),
В.А. Краснобаєв, д-р техн. наук, професор (ХДТУСГ)

Основи наукових досліджень: Підручник. – Харків:
О 753 УкрДАЗТ, 2012. – 338 с.
ISBN 978-966-2033-82-3

Підручник призначений для студентів старших курсів, що вивчають дисципліни «Метрологія, стандартизація, сертифікація засобів зв'язку та акредитація», «Багатоканальні системи зв'язку», «Основи наукових досліджень», та «Контроль та діагностика комп'ютерних систем та мереж».

В підручнику викладені загальні принципи наукових досліджень, поняття про якість системи, об'єкта або процесу, типи критеріїв якості систем, етапи постановки й розв'язання задач оптимізації складних інформаційно-вимірвальних систем (ІВС).

Підручник може бути корисним для фахівців ІВС та ТС.

УДК 621.396.6
ББК 32.841:72.4(2)

Підручник

Альошин Геннадій Васильович,
Лістровий Сергій Володимирович,
Панченко Сергій Володимирович,
Приходько Сергій Іванович

**ОСНОВИ НАУКОВИХ
ДОСЛІДЖЕНЬ**

Відповідальний за випуск Альошин Г.В.

Редактор Еткало О.О.

Підписано до друку 20.03.12 р.

Формат паперу 60x84 1/16 . Папір писальний.

Умовн.-друк.арк. 10,25. Тираж 500. Замовлення №

Видавець та виготовлювач Українська державна академія залізничного транспорту
61050, Харків - 50, майдан Фейербаха, 7

Свідоцтво суб'єкта видавничої справи ДК № 2874 від 12.06.2007 р.

**УКРАЇНСЬКА ДЕРЖАВНА АКАДЕМІЯ
ЗАЛІЗНИЧНОГО ТРАНСПОРТУ**

Г.В.Альошин, С.В.Лістровий,
С.В. Панченко, С.І. Приходько

ОСНОВИ НАУКОВИХ ДОСЛІДЖЕНЬ

Підручник

Харків 2012

Удк 621.396.6

ББК 32.841

A57

ISBN 978-966-2033-13-7

Альошин Г.В., Лістровий С.В., Панченко С.В., Приходько С.І. Основи наукових досліджень:
Підручник.- Харків: УкрДАЗТ, 2012.- 229 с.

Підручник призначений для студентів старших курсів, що вивчають дисципліни «Метрологія, стандартизація, сертифікація засобів зв'язку та акредитація», «Багатоканальні системи зв'язку», «Основи наукових досліджень», та «Контроль та діагностика комп'ютерних систем та мереж».

В підручнику викладені загальні принципи наукових досліджень, поняття про якість системи, об'єкта або процесу, типи критеріїв якості систем, етапи постановки й розв'язання задач оптимізації складних інформаційно-вимірювальних систем (ІВС).

Висвітлені питання про корисність наукових результатів і про ступінь довіри до них, ухвалення рішення в умовах невизначеності.

Охарактеризований сучасний стан проблеми оптимізації ІВС за багатьма показниками якості і загальні питання їх оптимізації. Введені поняття системного аналізу показників якості ІВС і дослідження впливу технічних параметрів, збурень, нестабільностей та неідеальностей на оптимальний синтез параметрів систем. Наведені співвідношення та вирази для оптимального синтезу несуміщених та суміщених ІВС.

Суттєва увага приділена задачам цілочисельного лінійного програмування (ЦЛП) для задач оптимізації телекомунікаційних систем. Докладно викладені методи рішення задач ЦЛП та нелінійного булевого програмування, відомі підходи, принципи оптимізації ТС, моделі, алгоритми та правила. Наведені приклади експериментальних досліджень для задач різної складності та комбінаторних задач для різних критеріїв якості.

Підручник може бути корисним для фахівців ІВС та ТС.

Іл.104, табл.8, бібліогр.: 40 назв.

Рекомендовано Міністерством освіти і науки України як підручник для студентів вищих навчальних закладів.

(14/18.Г-73 від 10.01.2011р.)

Друкується за рішенням вченої ради УкрДАЗТ
(протокол №1 від 01.01. 2011 року)

ЗМІСТ

Список скорочень	4
Передмова	8
Вступ	12
1. Загальні принципи наукових досліджень	16
1.1. Відомі принципи наукових досліджень	16
1.2. Поняття про якість системи, об'єкта або процесу	20
1.3. Порівняльні характеристики критеріїв якості систем	24
1.4. Етапи постановки й розв'язання задач оптимізації складних систем	27
1.5. Про корисність наукових результатів і про ступінь довіри до них	30
1.6. Ухвалення рішення в умовах невизначеності	34
1.7. Принцип виключення в повній схемі подій	36
1.8. Метод аналогій у наукових дослідженнях	36
2. Загальні питання оптимізації ІВРТС	40
2.1 Сучасний стан проблеми оптимізації ІВРТС за багатьма показниками якості	40
2.2. Про подвійність багатокритерійних задач	46
2.3. Ефективність складних інформаційно-вимірювальних систем	50
3. Системний аналіз якості інформаційно-вимірювальних радіотехнічних систем	53
3.1. Вплив збурювань, нестабільностей і неідеальностей сигналів і схем на кореляційну й фільтраційну обробку сигналу	56
3.2. Вплив обмеженості смуги пропускання високочастотного тракту на кореляційну обробку ФМПВП сигналу	65
3.3. Вплив похибки кутового супроводу антени на потужність сигналу й оптимальне співвідношення між її спрямованістю і якістю кутової стабілізації променя	76
3.4. Вплив невизначеності в знанні частоти на якість кореляційної обробки сигналу, фазоманіпульованого за законом псевдовипадкової послідовності	81

3.5.	Вплив нестабільності тактової частоти на якість кореляційної обробки фазоманіпульованого сигналу	88
3.6.	Вплив довготривалої часової затримки на вихідний ефект приймача	90
3.7.	Одночасний вплив розстроювань, збурювань і неідеальностей на вихідний ефект приймача	92
4	Параметричний синтез інформаційно-вимірювальних радіотехнічних систем	102
4.1.	Показники якості радіотехнічних систем	102
4.2.	Вартість блоків системи як обмеження на відповідні технічні параметри	105
4.3.	Структура системи, що оптимізується, і позначення її технічних параметрів	111
4.4.	Похибка вимірювань радіальної швидкості як критерій для оптимізації системи.....	115
4.5.	Визначення мінімуму цільової функції	125
4.6.	Оптимальні параметри систем	132
4.7.	Приклад оптимізації системи за трьома параметрами, для яких відомі обмеження за вартістю	136
5.	Оптимізація технічних параметрів сумісних інформаційно-вимірювальних радіотехнічних систем	139
5.1.	Принципи суміщення інформаційно-вимірювальних систем	140
5.2.	Постановка задачі	141
5.3.	Розв'язання задачі параметричного синтезу суміщених радіотехнічних систем	149
5.4.	Обговорення результатів	157
5.5.	Оптимізація сумісних РТС за основними ТТВ методом геометричного програмування при позиноміальних залежностях вартості від параметрів	160
6.	Задачі лінійного булевого програмування в телекомунікаційних системах і мережах	171
6.1.	Методи розв'язання задач дискретної оптимізації і їх класифікація	171
6.1.1.	Задачі дискретної оптимізації і їх постановка	176
6.2.	Класифікація методів розв'язання задач дискретної оптимізації	176

6.3 Ранговий підхід до розв'язання задачі 0, 1-рюкзак	184
6.3.1 Формальна модель n-вимірного одиничного куба в ранговому підході	184
6.3.2. Принцип оптимізації по напрямку в n-мірному одиничному кубі й узагальненій процедурі розв'язання ЦЛП із БП на основі ідеї рангового підходу	189
6.3.3. Правила відсікання безперспективних варіантів рішень на основі узагальненої процедури рангового підходу	192
6.4. Алгоритми розв'язання задачі 0, 1-рюкзак на основі рангового підходу	222
6.4.1. Наближені алгоритми розв'язання задачі ЦЛП із БП	223
6.4.2. Точні алгоритми розв'язання задачі ЦЛП із БП	229
6.4.3. Багатоетапні алгоритми розв'язку задачі ЦЛП із БП	233
6.4.4. Особливості розв'язання ЦЛП із БП у випадку рівності коефіцієнтів у функціоналі	243
6.5. Ранговий підхід до розв'язання задачі про найменше покриття й розбивку (ЗНП і ЗНР)	247
6.5.1. Формальна модель ЗНП	247
6.5.2. Ранговий підхід до розв'язання ЗНП	248
6.5.3. Правила відсікання безперспективних варіантів при розв'язання ЗНП на основі рангового підходу	252
6.5.4. Наближені й точні алгоритми розв'язання ЗНП ...	258
6.5.5. Особливості рангового підходу при розв'язанні ЗНР	262
6.5.6. Ранговий підхід до розв'язання діофантових рівнянь із БП	264
6.6. Експериментальне дослідження алгоритмів розв'язання задачі ЦЛП із БП на основі рангового підходу	266
6.6.1. Показники ефективності алгоритмів	266
6.6.2. Результати експериментального дослідження алгоритмів	269
6.6.3. Порівняльний аналіз рангових алгоритмів з відомими алгоритмами	279
6.6.4. Аналіз впливу сортувань коефіцієнтів при функціоналі й обмеженнях на величину похибки розв'язання	282

6.6.5. Гарантовані прогнози і їх вплив на тимчасову складність і похибки рангових алгоритмів розв'язку задачі 0, 1-рюкзак	292
7. Загальна схема розв'язання задач комбінаторної оптимізації й задач нелінійного булевого програмування	300
7.1. Загальний підхід до розв'язання задач комбінаторної оптимізації й теорії графів	300
7.2. Загальний метод розв'язання довільних задач булевого програмування	313
7.3. Метод розв'язання задачі визначення найкоротших гамільтонових шляхів	325
7.4 Експериментальне дослідження алгоритмів розв'язання задач булевого програмування на основі теорії графів	328
Бібліографічний список	335

СПИСОК СКОРОЧЕНЬ

- ОПР – особа, що приймає рішення
УКП – умовний критерій переваги
САПР – система автоматизованого проектування
ІВС – інформаційно-вимірвальна система
ТС – телекомунікаційна система
ТТВ – тактико-технічні вимоги
ФМ ШШС – фазоманіпульований широкосмуговий шумоподібний сигнал
ПВП – псевдовипадкова послідовність
ФМ ПВП – фазоманіпульована псевдовипадкова послідовність
ШШС – широкосмуговий шумоподібний сигнал
ПРЧ – підсилювач радіочастоти
КСД – коефіцієнт спрямованої дії
ГТІ – генератор тактових імпульсів
ФАПЧ – фазова автопідстройка частоти
ППЧ – підсилювач проміжної частоти
ФД – фазовий детектор
ФНЧ – фільтр нижньої частоти
ВЧТ – високочастотний тракт
ПГ – перестроюваний генератор
МНК – метод найменших квадратів
ПРД – передавач
ККД – коефіцієнт корисної дії
ШСС – широкосмуговий сигнал
ЦЛП – цілочисельне лінійне програмування
ФАК - функція автокореляції

Не швидкому дістається успішний біг,
Не сильному – перемога й не розумному багатство.
Але час і випадок всьому.
Екклезіаст

ПЕРЕДМОВА

Статистика показує, що найчастіше успіху досягають майстри своєї справи. Як стати майстром своєї справи?

На роботі ми проводимо кращий час свого життя – двадцять і більше молодих років і кожен робочий день з 8-ї до 18-ї години. Виходить, щоб наше життя було легким і цікавим, потрібно, щоб робота була легкою й цікавою. Що означає – «легкою й цікавою»? Іноді й навантаження-розвантаження не здається важкою роботою, якщо придумав для цього спосіб, пристосування або щось корисне й ефективне.

Досвід людства говорить про те, що робота повинна бути творчою й висококваліфікованою, чим би ви не займалися. Тому що творчість означає творення, вигадку, кваліфікація означає, що щось робиться вами вперше, хоча б у вашому регіоні. Навіть для найпростішої роботи потрібна відповідна майстерність. Тоді є й моральне задоволення, самоствердження й стабільність у цьому житті, у вашому колективі й у суспільстві. А легкість роботи приходить тоді, коли ще на етапі навчання ви **намагалися** все головне (для вас і спеціальності) **знати й розуміти**. Що означає слово «зрозуміти»?

Зрозуміти – це означає установити логічний, причинно-наслідковий зв'язок між тим, що ви вже знаєте, і тими знаннями, які одержали.

Тобто в ієрархії знань не повинно бути пропусків проміжних знань. **Поступовість, багатоетапність навчання – це його основний принцип.** Більш шкідливо пропустити перші шаблі знань і вмінь, тобто пропустити початкові знання. Зубріння не тільки некорисне для одержання знань, а й шкідливе. Змушувати людину вчити те, що вона не розуміє, – це знущання над людиною. Послідовний, логічний виклад матеріалу

викладачем і послідовне розуміння студентами сутності матеріалу укладається в пам'яті й запам'ятовується легко, як би саме собою. І особливої пам'яті не потрібно. Тому необхідно просто уважно, не відволікаючись, послухати викладача. І розуміння, і гарна оцінка вам уже забезпечені. Однак здатності зосередитися, щоб чути, а не тільки слухати, також потрібно навчитися.

Таким чином, розуміння – це встановлення логічного або причинно-наслідкового зв'язку між наявною й отриманою інформацією.

Розуміння забезпечує гарне запам'ятовування саме за рахунок додаткового зв'язку між старою й новою інформаціями. Краще розуміння буває тоді, коли нова інформація незначно відрізняється від старої. Звідси виходить, що складна інформація повинна подаватися поступово, багатоетапно. Пропуск одного із шаблів, або етапів, приводить до нерозуміння кінцевої інформації. Виходить, що нераціонально пропускати які-небудь етапи, особливо перший етап.

У цих шаблях, що підбираються викладачем, також є свій оптимум: викладати занадто докладно чи просто докладно – це іноді стомлює, тому що давно все ясно, тобто «пережоване – не смачне», а якщо занадто революційно – типу: «здогадайтеся самі», – це може бути цікавим, як ребус, або незрозумілим.

Ефект багатоетапного розуміння пов'язаний або з асоціативною пам'яттю – запам'ятовування супутніх деталей, або із плюралізмом – всебічним розглядом об'єкта, явища, інформації. Плюралізм дає надійні знання за рахунок додаткового погляду з іншого боку, з інших позицій, при інших умовах та ін. Звідси приказка при розгляді сутності: «з одного боку..., з другого боку ...».

Наше завдання підказати вам, як стати успішним у будь-яких справах, тобто як стати творчим фахівцем або майстром своєї справи найбільш коротким і легким, як кажуть, «оптимальним» шляхом. І хоча поширена думка, що вам уже все пізно, і ви себе в цьому переконали, ця думка помилкова.

З історії людства відомо, наприклад, що видатний французький математик А. Пуанкаре тільки почав навчатися, займатися математикою із 45 років. У вас не все втрачено. До того ж важка спочатку для вас математика вам потрібна найчастіше лише в межах спеціальності. Однак без математики неможливо підготувати висококваліфікованого фахівця, тому що кожна, у тому числі прикладна, «наука досягає досконалості, якщо їй вдається користуватися математикою». Потрібні в техніці й у суспільстві математичні відношення – це, власне кажучи, абстрактні логічні закони природи. Для успіху в будь-якій складній справі не завжди можлива й потрібна математика. Однак принципи, викладені далі, допоможуть вам бути успішним скрізь.

Помічено, що неважливо, чим людина зайнята, що вона робить, важливо лише скільки часу й уваги вона цій справі приділяє, що нового вона засвоїла. Такий з людини й вийде майстер своєї справи. А майстрові буде легко й цікаво працювати. А у вас багато часу витрачається на навчання?. Так чи не краще його витратити продуктивно, тобто слухати уважно, щоб стати хорошим фахівцем?

Як стати висококваліфікованим фахівцем?

Студентів не буває нездатних. Просто студент часто відстає у засвоєнні матеріалу, що викладається. Тому його так і називають: «невстигаючий студент». Можливі такі причини неуспішності студента: пропустив щось у школі, на лекціях у вузі, був неуважним, розмовляв з товаришами, не зрозумів ідеї, які викладалися до цього, й ін. Як бути успішним у навчанні? Для цього в нас такі поради.

1. Намагатися уважно слухати лекції з метою **зрозуміти** ті ідеї, які викладаються. Як уже відзначено, зрозуміти ідею – це означає **встановити зв'язок нових відомостей з тими ідеями, які вже були викладені**. Найчастіше це буває **причинно-наслідковий зв'язок**. Такий логічний зв'язок легше запам'ятовується. Тому пропускати, особливо перші лекції, нерационально.

2. Щоб легше стало розуміти, ви в себе розвиваєте творче мислення – креатив (creative). Частіше собі й іншим задавайте питання: чому ? – і намагайтеся відповісти, хоча б на рівні гіпотез.

3. Умійте виділяти головне й знаходити зв'язок з іншою інформацією.

4. Користуйтеся викладеними далі принципами, виробленими людством.

Технічну літературу й дану книгу варто читати багатоетапно. Спочатку ви переглядаєте зміст, заголовки й намагаєтеся знайти ключові слова й теми, що вас цікавлять. Потім відшукуєте головні ідеї, закладені в цікавих для вас місцях книги, пропускаючи подробиці й аналіз формул. Потім головні ідеї класифікуєте за стандартним принципом, відповідаючи для себе на питання: навіщо це вам потрібно й корисно? Що було відомо? Які нові результати? Як вони отримані? Чи всі коректні? Чи можна застосувати нові результати у ваших цілях? І так далі, до цікавих подробиць, адже це часто і є фундамент ідей.

ВСТУП

При розробленні телекомунікаційних систем можна помітити деякі загальні відомі й маловідомі принципи й методи їх оптимальної побудови й використання. Причому вони виявляються настільки загальними, що або неможливі для використання, або в нагоді тільки для загальних задач чи для локальних задач оптимізації, для оцінки оптимальності, для оцінки адекватності будь-яких систем, об'єктів, суб'єктів і процесів.

Слід зазначити споконвічне прагнення людства до поліпшення всього, що йому потрібно. Склалась навіть народна приказка: «риба шукає, де глибше, а людина – де ліпше». Однак навіть саме, загалом кажучи, не просте поняття, що таке «краще», ще має потребу в поясненні. Інакше кажучи, людство тільки й займається оптимізацією чогось: об'єктів або процесів, хоча іноді про це не завжди знає. І вся жива природа при адаптації до нових умов за рахунок природного або будь-якого відбору цим і займається. Вищих організмів, і особливо людину, відрізняє від іншого світу те, що в них більше розвинене мислення, що дозволяє ще перед своєю діяльністю одержати (уявити), якщо не оптимум результату, то хоча б прийнятний його варіант. Це дозволяє заощадити матеріальний ресурс, час і зусилля.

Уже з даного розгляду виходить, що існують такі шляхи досягнення оптимуму: у деякому відношенні – це «сліпий» **метод «проб і помилок»** і **метод абстрактного мислення**, де використовується, якщо можна, вся накопичена інформація, продумані певні етапи ухвалення рішення, перш ніж приступити до реалізації потрібної дії. Можливий також **комбінований** шлях або типу **імітаційного моделювання** й т. д.

До теперішнього часу вже накопичений великий обсяг інформації про всі сфери діяльності людства. Це не тільки змушує якимось упорядкувати, «зачесати» відому, нажиту інформацію, тобто використати загальноновизнану класифікацію знань з відповідних галузей, але й змушує людство шукати

найкоротші й ефективні шляхи вдосконалювання різних природних і штучних систем, об'єктів і процесів. Тобто шукати нові й удосконалювати відомі відповідні оптимуми. **Цим і займається наука з метою поліпшення соціального, екологічного й іншого статусу людства або його частини.**

Через багатозначність і відносність поняття «кращого» можливі й шкідливі для суспільства напрямки діяльності.

У будь-якому разі наука – це пізнавальна діяльність людства і його представників, спрямована на виявлення стійких, повторюваних властивостей, ознак, ефектів, явищ, принципів, методів, тобто на виявлення маленьких або великих, загальних або не дуже загальних, але в якійсь мірі стабільних законів природи.

Закони природи відшуковуються як у реальному світі, у тому числі в техніці, у промисловості, у суспільстві, в економіці, так і в абстрактному світі накопиченої інформації. Нагромадження й відображення інформації відбувається в спеціалізованих інститутах, у бібліотеках, вузах на будь-яких носіях. Нагромадження інформації відбувається також цілеспрямовано, із застосуванням відповідної класифікації знань. У бібліотеках використовується Універсальна десяткова класифікація (УДК), а також алфавітний і систематичний каталоги. Наприклад, радіотехнічні системи мають класифікацію УДК 621.396. У патентній бібліотеці використовують європейську класифікацію - клас, підклас, група, підгрупа. Наприклад, вимірювальні прилади мають клас – Н01 S 28/12. Таким чином, у гуманітарних і технічних напрямках, в абстрактних і прикладних науках, у математиці та в інформації, по галузях науки й знань і у всіх сферах діяльності людини – усюди відбувається нагромадження знань та їх класифікація. Тому й будь-яка наука, і розумна діяльність починається з класифікації.

Сутність класифікації – відбір об'єктів і явищ за певними загальними ознаками. Об'єкти і явища в одному класі є відносно один до одного аналогом. При цьому найбільш значущими є аналоги стійких, повторюваних властивостей, ознак, ефектів, явищ, принципів, методів, загалом кажучи, законів природи.

Однак формальної їх класифікації практично не достатньо. Тому у математиці у вигляді численних абстрактних моделей відображені абстрактні логічні закони природи, що можуть бути якимсь чином використані у будь-яких галузях.

Пізнавальну діяльність людства варто відрізнити від пізнавальної діяльності окремої особистості, спрямованої на навчання. Перша діяльність звичайно відрізняється від другої насамперед цілями й засобами, хоча можливе суміщення цілей.

Таким чином, **пізнавальна діяльність людства опирається** на базу знань, отриманих у результаті досвіду, логічних міркувань і фактів. При цьому вона використовує **причинно-наслідкові зв'язки** для виявлення стійких, повторюваних явищ, законів і закономірностей у матеріальному або в абстрактному світі.

Таким чином, існують такі загальні методи наукових досліджень:

- експериментальний;
- теоретичний;
- імітаційного моделювання;
- метод висування гіпотез із перевіркою фактами;
- комбінований метод наукових досліджень.

Перевагою методу «проб і помилок» є простота думки й дій. Однак при великій апріорній невизначеності цей метод малоефективний, хоча в зовсім новій справі або в дії він може бути неминучий. Платою за простоту думки й дій є помилки з їх небажаними, іноді незворотними наслідками.

Теоретичний метод абстрактного мислення заснований на використанні комбінації узагальнень і причинно-наслідкових зв'язків для одержання уявних або відбитих на будь-яких носіях інформації образів об'єктів, процесів. Інакше кажучи, використовується абстрагування будь-якої конкретики з метою визначення чогось стійкого, повторюваного, у тому числі законів природи, явищ, ефектів, принципів, методів і т.д. Дія ще не починалася, а вже продуманий її прийнятний план. Об'єкт ще не побудований, а вже є його проект. Це дуже ефективно.

Усі складні й важливі об'єкти і дії нерационально здійснювати без попереднього плану.

У протилежному випадку дії з реалізації об'єкта або процесу будуть здійснюватися, власне кажучи, методом «проб і помилок», тобто з витратами часу й матеріальних засобів. Метод абстрактного мислення настільки ефективний, що навіть експериментальні роботи розумно виконувати після застосування так званого «методу оптимального планування експерименту».

Уже існує багатовіковий і сучасний досвід людства в питаннях, якщо образно, раціональної тактики й стратегії наукових досліджень. Існує також теорія оптимального планування експериментів. Цей досвід людства являє собою як би технологію наукового виробництва. Він дозволив одержати для наукових досліджень ряд стійких принципів, методів, прийомів і алгоритмів, які можна назвати відомим терміном – «**know how**», або «знаю як».

Далі все це буде викладено як із загальних позицій, так і для застосувань у галузі зв'язку й радіоелектроніки. Відомий досвід розв'язання багатьох задач оптимізації радіоелектронних систем [1 – 4 та ін.], як і наш власний [5, 6], підтверджує відомий досвід людства в наукових дослідженнях. Основні питання такого досвіду для тих, хто починає наукові дослідження, викладені далі.

1. ЗАГАЛЬНІ ПРИНЦИПИ НАУКОВИХ ДОСЛІДЖЕНЬ

1.1. Відомі принципи наукових досліджень

Якщо у вас виникла необхідність вирішення якого-небудь складного завдання або проблеми, причому однаково, наукової або життєвої, то рано чи пізно мають бути з'ясовані такі питання: із чого почати, що робити, як робити, чого можна досягти, як краще досягти, виходячи з усіх варіантів рішень. У кожному разі саме ви є особою, що приймає рішення (ОПР), незалежно від того, чи залучали ви до цього знавців, фахівців, експертів, чи ні. Тому що саме ви особисто найбільш зацікавлені у вирішенні завдань і проблем. Про початок будь-якої складної справи не випадково з'явилося висловлення відомих авторитетів про те, що почати яку-небудь справу – це вже ледве не пів справи. Щодо відсотків від усієї справи можна ще посперечатися, але ясно одне: якимось чином уже пройдений один з важливих етапів, прийнято з якихось міркувань перше, може бути, поки що передчасне рішення про подальші дії. Перший етап вирішення завдання або проблеми звичайно буває інтуїтивний (евристичний). При цьому ми намагаємося розкласти наше складне завдання на більш прості етапи й потім його вирішувати. Крім того, намагаємося використати відомі рішення аналогічних завдань, у випадку суперечливих даних знаходити «золоту середину» – оптимальне значення, у випадку великої невизначеності знаходити найбільш імовірне рішення і т. д. Такі закономірності відомі вже давно.

Відомий, наприклад, так званий «наріжний камінь» науки під назвою **«Бритва Оккама»**. Уільям Оккам – францисканець (XIII століття) – виклав свій філософський погляд на наукові дослідження у вигляді висловлення: **«Не слід розмножувати число сутностей понад необхідність»**.

Під «сутністю» розуміються всі існуючі об'єкти, суб'єкти й процеси. Інакше кажучи, якщо існує можливість не ускладнювати (не збільшувати) розгляд дій або завдань, тобто не брати до уваги непотрібні або малозначні в даному завданні сутності, без яких можна обійтися, хоча б спочатку, то це раціонально робити.

Таким чином, для вирішення своїх завдань і проблем треба по можливості спрощувати склад їх даних, зберігаючи головне. Інакше, найімовірніше, буде обрізаний результат вирішення завдань і проблем – «Бритва Оккама». Загрузнути в складностях вирішення завдань і проблем – це найімовірніша плата за зневажання принципом Оккама.

Принцип Оккама, власне кажучи, пропонує залишати для вирішення завдань і проблем найістотніші, найбільш значущі для розглянутого завдання або проблеми дані або відомості, але по можливості найпростіші.

Таким чином, пропонується перед рішенням завдань і проблем використати відомий у науці прийом – аналіз вихідних і, можливо, прогнозованих даних або відомостей з метою спрощення завдань і проблем на першому етапі. Природно, що при цьому залишають найбільш істотні дані або відомості, зневажаючи іншими. Логіка проста – якщо через складність або через обмежений час не вийде врахувати інші дані, то втрати не настільки великі, зате результат у першому наближенні можна одержати.

Більш повні рекомендації для ефективних наукових досліджень виклав **Декарт** у вигляді п'яти принципів:

- 1) не поспішати з висновками;
- 2) позбавлятися упередженості;
- 3) робити, якщо можна, більш повні огляди того, що зроблено попередниками;
- 4) кожне питання розкладати на більш прості питання або завдання (це завдання аналізу);
- 5) починати рішення від простого до складного (це завдання поступового поетапного синтезу моделі).

Ці теоретико-пізнавальні (гносеологічні) принципи є більш повною основою для ефективного досягнення мети наукових досліджень.

Розглянемо сутність кожного принципу Декарта.

Перший принцип означає таке. Необхідно, якщо можна, збільшувати або час спостереження чого-небудь, або набирати необхідну статистику, щоб випадковості менше діяли на результат, або розглядати це щось із різних боків або за

допомогою декількох спостерігачів (плюралізм). Це дозволяє підвищити вірогідність результату спостережень або висновків, а також знайти інші, більш ефективні шляхи вирішення завдань.

Другий принцип радить бути більш гнучким і не зациклюватись на одній ідеї, якою б чудовою вона не здавалася, і глянути на предмет досліджень з іншого боку. Тобто принцип радить здійснити плюралізм самим дослідником, не вдаючись до сторонньої кваліфікованої допомоги, яку важко отримати. Це також повинно підвищити вірогідність результатів, до того ж побачити інші, можливо більш ефективні, шляхи вирішення завдань.

Третій принцип означає нарощування бази даних з метою знайти якщо не те, що потрібно, то хоча б близький аналог. Цей принцип і дозволяє оцінити новизну й корисність результатів вирішення завдань шляхом виключення повторного їх вирішення. Крім того, він включає плюралізм поглядів попередників, тобто може дозволити побачити ті ж завдання з іншого боку або підказати інші шляхи їх вирішення.

Четвертий принцип – це фактично аналіз складних вихідних даних і використання, звичайно спрощених, постановок, рішень завдань і результатів.

П'ятий принцип – це порада використання поступового ускладнення постановок і рішень завдань, тобто порада послідовного (поетапного) синтезу систем або процесів за даними, що ускладнюють послідовно.

Перший **спрощений варіант постановки й рішення завдань, що називають моделлю**, ще не в повній мірі, тобто не цілком точно, відбиває властивості системи, об'єкта або процесу. Це означає, що **модель ще неадекватна** системі, об'єкту або процесу. Додатково й поступово з огляду на найбільш значущі фактори при постановці завдань часто буває можливим використати попереднє рішення також в ускладненій постановці й у рішенні завдань. Принаймні видно також усі нові властивості об'єктів і процесів в ускладненій постановці й у рішенні завдань.

Принципи У. Оккама й Р. Декарта носять настільки загальний характер, що вони застосовні як для точної постановки й рішення завдань за допомогою математики, так і для прийняття рішень у чисто інтуїтивних (евристичних) складних ситуаціях.

Ці принципи обмежують лише одну грань розгляду завдань – обмежують складність зверху. Дійсно, необхідно якомога більше обмежувати зверху складність завдань. Оскільки при більшій складності постановки завдань результат буде «обрізаний», тобто не буде отриманий результат (через зайвий ступінь складності).

Однак, як правило, існує й інша грань, протилежна зайвій складності, – переспрощення завдань. Про це свідчить більш повне висловлення **Р. Беллмана** [1]: **«Дослідник подібний до мандрівника, що йде по вузькій стежині між пасткою переспрощення і болотом переускладнення».**

Переспрощення завдань – це таке зайве спрощення, при якому модель (або проект) стає неадекватною системам, об'єктам або процесам. Тобто властивості спрощеної моделі неприпустимо сильно відрізняються від реальної системи, об'єкта або процесу. Наприклад, якщо якість об'єкта або процесу описується двома суперечливими показниками якості й забрати один з показників, то оптимум може зникнути. Тому що модель якості об'єкта стала простішою й перестала в значній мірі відповідати більш складній якості самого об'єкта, тобто модель стала неадекватною.

Уже у висловленні Р. Беллмана спостерігаються дві грані обираної складності для вирішення завдань на першому й далі на кожному етапі поступового ускладнення: з одного боку, переускладнення, а з іншого – переспрощення завдань. Евристично відчувається, що вже на першому етапідесь є оптимум обираної складності постановок і рішень завдань.

Далі буде показано, що **якість навіть не дуже «складної» сутності практично завжди описується декількома суперечливими показниками якості.** Як мінімум один з показників якості являє собою корисну властивість, а інший – витратний показник. Ця обставина частіше свідчить про існування оптимуму. Для складних сутностей (систем, об'єктів або процесів) оптимум лежить частіше між двома гранями – крайностями й рідко на самій грані. І цей оптимум може бути знайдений, якщо не строго математично, то принаймні інтуїтивно (евристично). Цей оптимум звичайно називають усім відомим терміном, а саме **«почуттям міри».**

1.2. Поняття про якість системи, об'єкта або процесу

Під якістю системи, об'єкта або процесу будемо розуміти одну або сукупність їх корисних властивостей, що визначають і відрізняють систему, об'єкт або процес від аналогічних інших.

Кожна властивість системи, об'єкта або процесу є, власне кажучи, однією з їхніх відмінних ознак. Тому, іншими словами, під терміном «якість» системи, об'єкта або процесу будемо розуміти сукупність відмінних **корисних** ознак системи, об'єкта або процесу, що визначають ступінь їхньої корисності для застосування. Будь-яка корисна властивість або ознака, або одна з якостей системи, об'єкта або процесу, може бути кількісною або якісною, коли властивості між собою істотно відрізняються.

Для технічних додатків зручніше або бажано мати кількісний опис властивостей об'єкта або процесу. Кількісна властивість може бути такою, що чим більша (або менша) дана властивість, тим краща система, об'єкт, процес або їхні функціональні елементи. Наприклад, таку якість, як завадостійкість системи зв'язку, можна й зручно оцінювати кількісно, визначаючи мінімум імовірності помилки дискретного повідомлення, а от дизайн або ергономічні властивості можна оцінити тільки в балах.

Таким чином, найістотнішим при оцінці якості системи, об'єкта або процесу є визначення їхнього ступеня корисності для застосування. Під застосуванням варто розуміти використання системи, об'єкта або процесу за необхідним призначенням.

Оскільки застосування може бути й не за їх звичайним призначенням, то **під якістю системи, об'єкта або процесу варто розуміти сукупність їх корисних властивостей за їх необхідними конкретними призначеннями у даному практичному або теоретичному завданні.**

Дизайн або споглядальні властивості систем також впливають на загальну якість системи, якщо вони визначають, хоча б у якійсь мірі, цінність або корисність системи. Таким чином, очевидно, що ступінь корисності певної якості систем, об'єктів або процесів може бути більш-менш істотною в даному

завданні й визначається в значній мірі інтуїтивно особою, що приймає рішення (ОПР) [2, 3].

Кожна якість із їх сукупності в різному ступені визначає загальну якість, а отже, ступінь корисності системи для даного призначення.

Числове значення певної властивості (якості) систем називається показником її якості.

Кожен показник якості системи іноді називають критерієм якості. У цьому випадку опис системи багатьма показниками якості часто називають **багатокритерійним описом якості системи** [2, 3], а **вирішувані завдання – багатокритерійними завданнями**. Якщо сукупність таких критеріїв якості розташувати в ряд (ранжувати) за ознаками їх значущості або корисності для системи, або для застосування системи, то такий вектор критеріїв якості називають **лексикографічним критерієм якості** [2, 3]. Очевидно, що для застосування в завданнях або неможливо врахувати всі численні показники якості системи (або «критерії», відповідно до термінології [2, 3]), або недоцільно, відповідно до викладених принципів У. Оккама, Р. Декарта й Р. Беллмана. Тому принаймні спочатку варто обмежувати склад критеріїв якості, з огляду на їх важливість, корисність і визначаючи достатню простоту для одержання можливості постановки й вирішення завдань.

Точність опису системи залежить від повноти врахування складу й від значень вектора показників якості системи. Тобто точність опису визначає її **адекватність**. **Адекватність** – це точність відповідності спрощеної у абстрактному вигляді або у проекті якості віртуальної системи відносно якості реальної системи.

Використання проектів систем, об'єктів і процесів настільки ефективно, що практично завжди ці проекти використовуються при їх описі й абстрактному поданні у вигляді ідей, креслень і формул, особливо для вдосконалювання важливих або дорогих систем, об'єктів або процесів.

Спрощений опис системи й спрощений опис її якості називають відповідно **моделлю системи або моделлю опису її властивостей**. Такі поняття з'явилися завдяки зазначеним принципам наукових та інших типів досліджень. **Поетапне**

рішення завдань оптимізації систем відповідно до принципів Декарта вимагає поетапного доповнення складу показників якості системи для визначення більш адекватного опису оптимальної якості системи.

Опис застосування самої системи й опис її якості у значній мірі корельовані, тобто вони зв'язані між собою випадковими й не випадковими залежностями. У системах зв'язку, крім морфології (структури) системи, на показники якості впливають також технічні параметри й структури сигналів.

Завдання визначення оптимальної радіоелектронної системи полягає в тому, щоб знайти кращу систему на множинах технічних параметрів, сигналів і структур.

Пошук кращої системи, або оптимізація, відбувається шляхом варіювання значень технічних параметрів, сигналів і структур в обмежувальних умовах їх технічної реалізації або при використанні вже відомих їх оптимальних значень.

Узагальненим показником якості системи будемо називати його багатовимірну залежність від сукупності показників якості.

Узагальнений показник якості, не завжди прямо, залежить через технічні параметри від показників якості. Технічні параметри при цьому описують властивості функціональних елементів системи. Тому вони у свою чергу є показниками якості функціональних елементів.

Простежується така ієрархія радіоелектронних виробів: мережі, комплекси, системи, функціональні елементи, блоки, вузли, радіоелектронні деталі. Як би не хотілося вирішувати завдання пошуку кращих систем (їх оптимізації) відразу на всіх рівнях, але частіше спрацьовує «Бритва У. Оккама», що «обрізає» результат.

Правило або умови, які потрібно застосувати до узагальненого показника якості або до сукупності показників якості, щоб одержати кращу систему, називається критерієм якості.

Стосовно інших видів задач критерієм називається також правило або умови, які потрібно застосувати до множини результатів вимірювань, до зібраної статистики й т.д., щоб одержати певні відповіді на питання про узгодження

(відповідності) отриманої статистики з даним законом розподілу ймовірності, про найменшу середньоквадратичну похибку і т.д.

У загальному випадку **критерій являє собою правило або умови виконання, чи якості виконання якого-небудь завдання, якщо відомі відповідні дані та обмеження.**

Як зазначалося раніше, якщо показники якості системи, об'єкта, процесу будь-якої сутності нерівноцінні, то цей випадок називають **лексикографічним критерієм якості** [2].

Ранжування показників якості за значущістю здійснює особа, що приймає рішення.

При використанні лексикографічних критеріїв якості порівняння систем, об'єктів і процесів може здійснюватися також і при неоднаковій розмірності відповідних векторів показників якості \bar{D}_1 і \bar{D}_2 , якщо нехтувати малозначущими показниками якості без істотних втрат адекватності моделі системи. Але якщо ними нехтувати, то залишаються лише значущі показники якості, без яких не можна обійтися без істотних втрат адекватності моделі. А раз не можна без них обійтися, то це означає, що ми маємо справу з рівноправними за своєю значущістю показниками якості.

Звідси впливає наступний важливий принцип для наукових досліджень. **Спочатку необхідно залишати найважливіші для завдання вихідні дані й вагомні показники якості. І якщо завдання незоре, тільки тоді скористатися принципами У. Оккама, Р. Декарта, Р. Беллмана.**

Принцип логічно зрозумілий: необхідно прагнути, якщо можна, досліджувати головне. Тому що тоді буде отриманий головний, найефективніший результат. І якщо при цьому не встигли врахувати другорядні питання, то це не так важливо.

Звідси впливає, що для повноти опису систем, об'єктів і процесів в основному необхідно залишати в завданні значущі функціональні й витратні показники якості, які варто вважати рівноправними, а не лексикографічні критерії або показники.

У відомій літературі [2, 3] вводиться поняття «умовний критерій переваги» (УКП). Умовному критерію переваги відповідає один з таких методів використання вектора показників якості в завданнях пошуку кращої системи:

- 1) метод, заснований на введенні результуючого показника якості;
- 2) мінімакний метод;
- 3) метод переведення всіх показників якості, за винятком одного, який вважають **цільовою функцією (або функцією мети, корисності)**, у розряд обмежень (або **функцій зв'язку**);
- 4) метод послідовних поступок;
- 5) комбіновані методи.

Про застосування згаданих методів будемо судити відповідно за методом виключення з даного списку.

Метод послідовних поступок належить до лексикографічних критеріїв якості, що для технічних систем, об'єктів і процесів не завжди прийнятно. Тому що показники рівноправні. До того ж можливості методу обмежені й дозволяють лише порівнювати системи, а не знаходити оптимальні системи.

Мінімакний метод може використовуватися для побудови систем з невизначеними умовами функціонування. Основна проблема в таких завданнях – одержання платіжних матриць. І вирішення такого завдання, тобто одержання кращого рішення, ефективно тільки в найгіршому випадку, що досить марнотратно й не завжди прийнятно.

Залишаються інші методи використання вектора показників якості систем. Найчастіше використовують методи, засновані на введенні результуючого показника якості й на переведенні всіх показників якості, за винятком одного, у розряд обмежень. Ці методи зводяться звичайно до **математичного програмування**. Розглянемо їх порівняльні характеристики за ознакою об'єктивності, оскільки суб'єктивність несе у собі суб'єктивні помилки.

1.3. Порівняльні характеристики критеріїв якості систем

Під **оптимальним значенням критерію оптимальності** [2, 4, 5] будемо розуміти один із трьох видів:

- 1) значення критерію, заснованого **на одержанні екстремального значення** шляхом визначення похідних по кожній змінній, прирівнювання їх до нуля й розв'язання отриманої системи рівнянь;

2) значення критерію, заснованого на одержанні кращого значення на відрізьку монотонних кривих;

3) значення критерію, заснованого на принципі достатності, тобто для кривих типу насичення не слід вибирати занадто великі значення параметрів, які не приводять до істотного поліпшення критерію, хоча при цьому більші значення цього параметра приводять до більших витратних показників. Тому варто обмежитися лише розумними їхніми значеннями. Наприклад, для ймовірності правильного виявлення не слід вибирати занадто велике значення енергетичного потенціалу, виходячи з інших міркувань - для зменшення витрат.

При знаходженні одного з видів значення оптимальності за яким-небудь умовним критерієм переваги (УКП) $f[\vec{D}(\vec{X})]$ важливо, щоб існувала область визначення критерію $\{\vec{X}\}$ й область його значень. Під символом $f[\vec{D}(\vec{X})]$ тут розуміється функція (або функціонал) від вектора показників якості $\vec{D}(\vec{X})$, що залежить від вектора технічних параметрів \vec{X} . Для цього показники якості повинні бути взаємозалежними [2 – 4].

Ідеально було б одержати критерій переваги у вигляді функціонала мети функціонування системи $p_m = p_m(\vec{D}(\vec{X}))$. Однак для складних систем це звичайно не вдається.

У такому випадку використовуються нижчеперераховані критерії оптимальності.

1. **Зважений критерій** $f_1 = \min_{\{\vec{X}\}} \sum_{i=1}^n \xi_i D_i(\vec{X})$,

де $\min_{\{\vec{X}\}}$ означає мінімум або інфімум (менше значення на множинах або в області припустимих значень параметрів).

Основна перевага такого критерію – **простота**.

Основний **недолік** – **довільність** у виборі **вагових коефіцієнтів** ξ_i .

2. **Відносний критерій** f_2 , у чисельнику якого один показник або зважена сума значущих показників $D_j(\vec{X})$, таких, що чим вони більші, тим краще, а в знаменнику – сума витратних показників $D_k(\vec{X})$.

$$f_2 = \max_{\{\vec{X}\}} \frac{\sum_{j=1}^n \xi_j D_j(\vec{X})}{\sum_{k=1}^n \xi_k D_k(\vec{X})}.$$

Перевага відносного критерію – порівняна простота.

Недоліки – довільність у виборі вагових коефіцієнтів і вплив гіперболічної залежності знаменника, що **приводить до обмеженості використововуваного діапазону змінних**.

3. **Векторний критерій** якості Парето [2, 3]: якщо всі компоненти (показники якості) одного вектора менші відповідних компонентів іншого вектора, то перша система краща (або гірша).

$$\vec{D}(\vec{X}) = \left\{ \frac{D_l(\vec{X})}{D_k(\vec{X})} \right\}.$$

Перевага – простота застосування. Залежність від технічних параметрів не завжди потрібна. Цей критерій найпростіший для порівняння якості систем.

Недоліки критерію – незрівнянність векторів, якщо один компонент одного вектора кращий, ніж в іншого, а інший – гірший.

Для цього найпоширенішого випадку є, теж непростий, метод послідовних поступок.

4. **Умовний критерій якості**, або умовний екстремум f_3

$$f_3 = \min_{\{\vec{X}\}} D_1(\vec{X})$$

$$D_2(\vec{X}) \leq D_{2 \text{ доп}}$$

при

...

$$D_n(\vec{X}) \leq D_{n \text{ доп}}$$

де $D_{i \text{ доп}}$ – допустиме значення $D_i(\vec{X})$.

Перевага критерію – немає вагових коефіцієнтів і тому немає суб'єктивізму, довільності при їх виборі.

Цей критерій якості найбільш об'єктивний.

Звичайно, при виборі складу показників якості однаково є деяка довільність або інтуїція дослідника. Однак це стосується усіх критеріїв якості рівною мірою.

Недоліки умовного критерію якості – **складне розв'язання задачі** із застосуванням методу множників Лагранжа або методу підстановки параметрів з рівнянь зв'язку в цільову функцію, або методу математичної індукції для однотипних функцій. Крім того, можуть бути неприємності, пов'язані з неоднаковістю рішення, із подвійним **рішенням**, якщо за цільову функцію узяти інший показник.

1.4. Етапи постановки й розв'язання задач оптимізації складних систем

Під складними розуміються такі системи, які описуються багатьма показниками якості. Поняття «складна система» до деякої міри умовне, оскільки ступінь корисності будь-якої системи з достатньою точністю описується не тільки одним яким-небудь показником якості функціонування, але й відповідним витратним показником, що неминуче також має пряме відношення до якості функціонування системи. Тобто якість функціонування системи як би «купується» ціною яких-небудь витрат або ціною погіршення якості іншого суперечливого показника. Тому під «складною» системою можна розуміти також систему, що описується числом показників якості, не меншим двох.

Етапи постановки й розв'язання задач оптимізації складних систем з урахуванням відомих принципів наукових досліджень:

- 1) формування списку, складу або вектора показників якості системи;
- 2) ранжування показників якості за принципом корисності їх для системи;

- 3) виявлення «бойових» показників якості систем або істотних показників за корисністю й за умовами їх застосування;
- 4) виявлення істотних витратних показників системи;
- 5) виявлення істотних обмежень для функціонування системи;
- 6) виявлення рівноправних за значущістю показників за п. 3-5;
- 7) спрощення завдання за рахунок зменшення складу показників якості, тобто залишається за можливістю якомога менше число «бойових» показників, але обов'язково залишається витратний або суперечливий показник;
- 8) якісно, можливо спрощеними графіками або на ПЕОМ, визначають характер показників якості, їхню незалежність, суперечливість, монотонність, одномодальність (один екстремум) і т.д.;
- 9) вибирається узагальнений показник якості й критерій якості;
- 10) інтуїтивно оцінюється наявність оптимуму, взявши для оцінки загального показника якості системи випадки, якби окремі суперечливі показники мали найбільш крайні значення; як відомо, оптимум буває звичайно компромісним рішенням між двома й більше суперечливими показниками якості;
- 11) приблизно у квадратурах або точно в числовому вигляді вирішується спрощене завдання оптимізації системи;
- 12) аналізується оптимальне рішення спрощеного завдання, для чого варіюються вихідні дані в реальних межах;
- 13) складаються «криві обміну» [2, 3, 5] у вигляді графіків або таблиць;
- 14) виявляються закономірності оптимальних показників та оптимального розв'язання, які для аналізу їх властивостей краще мати у квадратурах (у вигляді формул);
- 15) додається черговий за значущістю показник якості й робляться спроби застосування розширеної постановки й вирішення завдання оптимізації;
- 16) дії за можливістю повторюються до прийнятної адекватності опису системи;
- 17) виробляється за можливістю універсальна, уніфікована **система автоматизованого проектування (САПР)** системи, об'єкта, процесу;

18) вихідні дані до завдань періодично уточнюються аналогічно стандартам з появою нових технологій, елементів, принципів побудови, нових законів природи й т.д.;

19) тому рішення завдань також необхідно уніфікувати й стандартизувати.

Якщо визначити показники якості і їхньої залежності від технічних параметрів не вдається, то завдання ставляться й вирішуються інтуїтивно, але також за даним, але спрощеним сценарієм. Саме цим звичайно й займається кожна людина щодня, приймаючи які-небудь рішення. Інтуїтивні рішення також гарні, якщо немає об'єктивних даних і рішень.

Спрощений сценарій розгляду складних проблем може виглядати таким чином.

1. Відповідно до викладених принципів розглядаєте свою проблему.

2. Вводите свої цінності, тобто показники якості, і залишаєте з них лише важливі й суперечливі показники.

3. Якщо більша складність, за можливістю спрощуєте.

4. Відчувши інтуїтивно суперечливість впливу факторів на показник, берете дві крайності розмірів факторів: 1) коли якогось фактора багато, то з однієї причини показник гірше, 2) потім берете випадок, коли цього ж фактора мало, тоді показник також гірше, але з іншої причини. Відчувається інтуїтивний оптимум при виборі розміру фактора, що впливає.

Визначення оптимуму інтуїтивним шляхом називається «почуттям міри».

5. Поступово ускладнюєте ваші міркування.

Кожен громадянин повсякденно приймає порівняно прості рішення. Наприклад, споживати велику кількість їжі – погано з однієї причини (ожиріння), а мало – з іншої причини (надмірне схуднення й неповноцінне харчування). У ролі показників якості життя тут виступають причини. Аналогічні інтуїтивні завдання з визначення: кількості води, що випиває людина за добу, кількості солі, що з'їдає, тривалості сну; кількості руху, перерозподілу часу, інтуїтивного розкладу, якості й кількості закупівель, оптимальних маршрутів, розмірів відпочинку й т.д., практично про все. У всіх питаннях є почуття міри й цю свою міру кожному потрібно знайти.

Є складніші проблеми – ви хочете зберегти свої не занадто великі заощадження на певному інтервалі часу. Для відкриття своєї справи суми недостатньо й ви вирішуєте питання про те, вкласти гроші в банк чи ні. Берете дві крайності. Якщо зберігаєте їх вдома, то навіть при стійкій економіці ви однаково втратите частину суми за рахунок інфляції. Плюс є ризик від квартирної крадіжки. Це погано. Друга крайність – викладаєте все в банк. Втрати менші, але є ризик банкрутства банку, ризик кризи, невиплати в потрібний момент і т.д. Оцінюючи ризики інтуїтивно, але в цифрах, можна знайти компромісне рішення в умовах впливу випадків, наприклад, розділивши суму між домом і банками з різною репутацією й ставками.

Таким чином, **викладені принципи можуть бути корисними не тільки в наукових дослідженнях.**

Але перш ніж братися за будь-яку справу, необхідно застосувати **принцип «Робінзона Крузо»**. Згадайте, що Робінзон Крузо цілий рік із твердого дуба довбав човен. Видовбав човен, але не зміг дотягнути його до океану. Тому **принцип «Робінзона Крузо» – це насамперед оцінка того, чи буде взагалі отриманий очікуваний результат і чи варто братися за дану справу.**

1.5. Про корисність наукових результатів і про ступінь довіри до них

Цінність наукових досліджень визначається їх корисністю, новизною та своєчасністю для суспільства або її частини й залежить також від довіри суспільства й споживачів цієї продукції, які беруть або візьмуть її на реалізацію або до подальшого розроблення.

Довіра до результатів наукових досліджень (до вихідних даних, до постановки, до рішення завдання й до результатів рішення) відіграє значну роль у визначенні корисності й реалізованості її результатів.

Але ще до початку наукових досліджень на підставі відомостей у відповідній галузі знань і на підставі довіри до них **оцінюється (прогнозується) ефективність наукових**

досліджень за такими показниками їх якості, як новизна, корисність і своєчасність.

Якщо напрямок або тема наукових досліджень малокорисні або несвоєчасні – ще наука й техніка не готові до реалізації або запізнилися – вже все відомо або є інша, альтернативна, більш корисна пропозиція, то вважається, що напрямок або тема неактуальні. Наприклад, гвинтолiт Леонардо да Вінчі в його часи був несвоєчасний, неефективний, хоча й занадто новий. А в наш час він (ця ідея) вже несвоєчасний, неефективний і вже не новий. Тобто зараз він неактуальний, поки не знайдеться попит на нього в новій якості, як на раритет.

Актуальністю напрямку або теми називається такий критерій для оцінки їх якості, що визначається наявністю й значущістю таких показників, як **новизна, корисність і своєчасність**. Ці показники не завжди є кількісними величинами. Тоді ступінь їх корисності можна оцінити лише суб'єктивно або, наприклад, у балах.

Це лише попередні оцінки корисності напрямку або теми наукових досліджень. А для оцінки якості наукових досліджень цього вже недостатньо. Потрібно також оцінювати, теж практично суб'єктивно, довіру до всіх етапів роботи: до вихідних даних, до постановки завдання, до її рішення й до результатів її рішення.

Якщо до будь-якого етапу, до постановки й рішення завдання у вас немає довіри, то й до результату його рішення теж не буде довіри. Невипадково у всіх розвинених країнах є свій суспільно-організаційний механізм випробувань будь-якої продукції, у тому числі наукової, на певний ступінь довіри до якості.

Цей механізм є, власне кажучи, **інститутом довіри**. Він містить систему рецензентів, опонентів, публікацій, апробацій, колегіальних органів, організацій для ліцензування, кваліметрії, метрології, стандартизації, інститути патентних досліджень, систему фірмових брендів, школи наукових і навчальних інститутів або школи авторитетів і т.д.

Але перш ніж завантажувати своїми результатами суспільні інститути випробувань результатів наукових досліджень на довіру до своєї роботи, дослідник повинен бути сам зацікавлений

у тому, щоб оцінити довіру до того, чим він займається. Якщо він цього не зробить, все це йому й тому, хто в нього повірив, дорожче обійдеться. Платою за таку помилку буде втрачений час, витрачена праця, втрата авторитету, недовіра до нього й інші втрати.

Від ступеня довіри до наукових досліджень і до всіх їх етапів істотно залежить довіра до їх результатів. Для точних наук це настільки важливо, що вони не можуть обійтися без чітких числових показників довіри.

Наприклад, у метрології й у статистиці [6] мало знати результат вимірювань – точкову оцінку, яка є, за правилом, математичним сподіванням. Треба ще знати як ступінь довіри – довірчу ймовірність або довірчий інтервал (інтервальну оцінку), або квантиль, середньоквадратичну похибку, або клас точності. Для метрології довірча ймовірність стандартна і дорівнює 0,997. Якщо, наприклад, у метрології результат вимірювань напруги в електричній мережі (точкова оцінка) відомий і дорівнює 220 В, а інтервальна оцінка, що пов'язана з довірчою ймовірністю або з максимальною похибкою, або невідома, або неточна, наприклад, дорівнює 220 ± 21 В, то можливе псування електроприладів.

Якщо для ряду напрямків у точних науках знайдені числові показники оцінки або критерії довіри, то для оцінки довіри до вихідних даних досліджень і взагалі до всіх етапів досліджень ще не вироблені числові критерії довіри. Є поки тільки експертні показники довіри. Оцінка наукової, педагогічної, творчої й гуманітарної праці може виконуватись у вигляді експертних оцінок, порівняльних характеристик результатів за всіма показниками, у вигляді зважених балів і т.д. Аналогічно можна оцінювати довіру до фактів, наукових результатів, положень, теорій, відкритих законів і закономірностей природи.

У якій мірі можна довіряти якому-небудь факту або відомостям?

Для оцінки такої довіри нижче пропонується **порівняльний критерій довіри або ймовірності** (рис. 1.1).

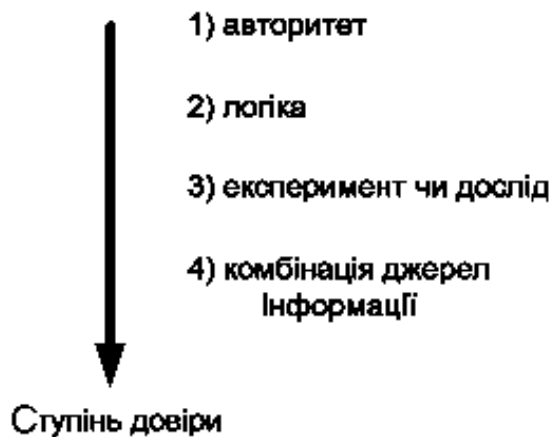


Рис. 1.1. Порівняльний критерій довіри

Логічне обґрунтування цього порівняльного критерію довіри полягає ось у чому. Відомості, що заявляються одним авторитетом, залежать від ступеня його авторитетності, але мають більш низьку довіру в порівнянні навіть із простою логікою та з достовірним і незалежним фактом. Тому що під логікою розуміється логічний, причинно-наслідковий зв'язок або логічне відношення декларованих відомостей і незалежного факту. При наявності логічного зв'язку істинність декларованих відомостей підтверджується двома джерелами: ймовірністю джерела інформації та ймовірністю незалежної підтверджуючої інформації. Підтверджуючою інформацією при цьому може бути як причина, так і наслідок або їх комбінація.

Коректно поставлений експеримент або інформація з досвіду мають найбільшу ймовірність, якщо це визначений факт або інформація, яка багаторазово перевірена незалежними спостерігачами, або доведена досвідом.

Комбінація зазначених джерел має відповідну вірогідність. Якщо джерел незалежної інформації для всіх її різновидів декілька, то ймовірність відомостей відповідно підвищується. Ймовірність відомостей джерел інформації може бути оцінена в балах або у вигляді ймовірностей істинності, ймовірності відомостей.

Далі ймовірність вірогідності відомостей будемо називати просто: «ймовірність». Вона може оцінюватися експертами, якщо

немає інших способів. Наприклад, якщо ймовірність декларованих тих самих відомостей від одного авторитету p_{a1} , а від іншого незалежного авторитету – p_{a2} , то результуюча ймовірність цих відомостей збільшується й дорівнює

$$p_a = 1 - (1 - p_{a1})(1 - p_{a2}). \quad (1.1)$$

Оскільки ймовірність загальної помилки дорівнює ймовірності суміщення незалежних подій, то вона значно менше ймовірності помилок відомостей від кожного авторитету.

Аналогічні міркування й для інших джерел інформації. Формулою (1.1) можна скористатися й для оцінки ймовірності другого типу джерела інформації, коли є логічний або причинно-наслідковий зв'язок з іншим незалежним джерелом інформації. Якщо ймовірність другого джерела, позначувана також p_{a2} , – достовірна подія, то й результуюча ймовірність – також достовірна подія.

Можливі й інші схеми й співвідношення для оцінки ймовірності відомостей. Найбільш ненадійні – це відомості, для яких ймовірність дорівнює 0,5. Ймовірність, рівна, наприклад, рівню 0,1, більш виразно свідчить про невірогідність (про хибність) відомостей.

Сформулювати черговий принцип для оцінки ймовірності відомостей у наукових дослідженнях можна також у вигляді нижченаведеного положення.

Більш просте пояснення складної події або іншої сутності, більш імовірно, оскільки ймовірність помилки для простого пояснення звичайно менша.

1.6. Ухвалення рішення в умовах невизначеності

Невизначеність у завданнях буває таких типів:

- 1) відомий закон розподілу випадкової величини;
- 2) невідомий закон розподілу, але є статистика;

3) відомий закон розподілу випадкової величини, але невідомі параметри розподілу (параметрична невизначеність);

4) нема або невідомий закон розподілу – **нечіткі множини** [4], оскільки **випадковою такою величиною назвати не можна**, тому що **немає ніякої стійкої закономірності** в її появі.

Стосовно пп. 1, 2, якщо хоча б один технічний параметр або показник якості – випадкова величина, то загальноприйнятий метод вирішення таких завдань – **замість випадкової величини використати її математичне сподівання**. Це природно, оскільки математичне сподівання або середньоарифметичне значення випадкової величини – це точкова, не випадкова, найстійкіша або найімовірніша оцінка. Далі необхідно оцінювати також довіру до результату досліджень.

Стосовно п. 3 існують складні непараметричні методи оптимізації систем.

Якщо величина належить нечіткій множині (п. 4), то для цього випадку є пропозиції дослідника Заде [4], що використовують евристичні функції приналежності нечіткій множині. Це теж інформація, якщо немає більш чітких подань. Найпоширеніший **приклад нечітких множин** – це **маркетингова вартість або ціна систем** та їх елементів. Метод обробки вартісної статистики для **перетворення нечіткої величини вартості у випадкову величину** й використання її в завданнях оптимізації систем викладений у роботі [4].

Крім викладеного, відомий також принцип урахування в задачах оптимізації випадкових і нечітких величин з повною невизначеністю. Це **принцип урахування невизначеності для гіршого або для фіксованого випадку**. Наприклад, енергетичний потенціал розраховують для гіршого випадку – для максимальної дальності радіолінії, надійність системи – для гірших умов застосування й т.д.

Відповідно до викладених принципів можна сформулювати **принцип одержання нових знань або відомостей в умовах значної невизначеності**. На основі відомих фактів, відомостей, законів природи висувається, спочатку як попередня, **робоча гіпотеза або припущення**. Оцінюється її ймовірність. Якщо ймовірність не досягає необхідного значення, висувається інша гіпотеза й т.д. У результаті залишається найбільш достовірною

гіпотеза. Наприклад, прямо експериментально перевірити ефекти теорії відносності практично складно. Не всі дослідження й відомості підтверджують висновки теорії відносності, наприклад, суперечить квантова теорія. Однак поки більш правдоподібною гіпотези немає. Тому гіпотеза приймається, але з відповідною ймовірністю, що визначається з урахуванням експериментів, ustalених думок кваліфікованих авторитетів і можливих природних і логічних зв'язків з відповідними розділами фізики. З кожним відкриттям або з кожною новою ідеєю гіпотеза переглядається або уточнюється.

1.7. Принцип виключення в повній схемі подій

Під повною схемою подій будемо розуміти n -альтернативну ситуацію, коли можливі всі n взаємовиключних подій. Принцип виключення подій полягає в тому, що якщо відомо, що відбулися (або не відбулися) всі $n-1$ події, то можна вважати, що не відбулась (або відбулась) подія, що залишилась. Щоб правильно ухвалити рішення щодо події, про яку невідомо, відбулася вона, чи ні, потрібна надійна інформація про $n-1$ події. Наприклад, у двохальтернативному виявленні можливі дві події: є сигнал або його немає. Якщо відома одна подія, то відома й інша, взаємовиключна подія.

1.8. Метод аналогій у наукових дослідженнях

Метод використання аналогій у наукових дослідженнях досить розповсюджений і часто використовується при створенні нових теорій, ідей і винаходів. Розглянемо його сутність.

Об'єкти, системи й процеси мають яку-небудь сукупність властивостей або ознак. Як відзначалося раніше, прийнятний за якістю і кількістю склад ранжованих і рівноцінних властивостей або ознак визначає **модель системи або процесу** й точність подання цієї моделі або її адекватність.

Модель – це спрощений опис дійсності. У свою чергу складна властивість може бути представлена своєю моделлю своїх ознак (за принципом матрьошки). Тут також простежується ієрархічність властивостей або ознак. Модель може бути реальним об'єктом, процесом, математичним описом, імітаційною й комбінованою.

Класифікація об'єктів, систем, процесів або їх властивостей за якими-небудь ознаками (властивостями) або за рівнями є, власне кажучи, відбір аналогічних об'єктів, систем, процесів або їх властивостей. Із класифікації звичайно починається будь-який розділ науки. Більш повна класифікація містить класи, підкласи, групи, підгрупи. Вона потрібна тому, що спрощує визначення й використання властивостей знову отриманих відомостей. Як приклади можна навести періодичну систему Д.І. Менделєєва, бібліографію, класифікацію винаходів у патентних організаціях, класифікацію флори й фауни, класифікацію галузей знань, науки й техніки, радіоелектронних систем та їх властивостей і т. д.

Класифікація є, власне кажучи, формалізований пошук аналогів за якими-небудь властивостями будь-якої сутності. Формалізація допомагає науковим дослідженням і використовує **принцип: «краще поганий порядок, ніж гарне безладдя».** Сама жива й нежива природа підтверджує тенденцію й прагнення до певного порядку, тобто до виконання своїх стійких, відкритих і не відкритих людиною, законів природи. Серед живої природи – це відбір кращого, як би прагнення видів до вдосконалювання, оптимізації тварин за всіма важливими для них напрямками, їх адаптація до мінливих умов перебування. Для неживої природи – це закони термодинаміки, найкоротші шляхи поширення енергії полів будь-якої природи, прагнення будь-якої сутності до стійкості, хоча б і тимчасової. Відомо, наприклад, що тривалі впливи полів тяжіння приводять до того, що форма небесних тіл прагне до кулястої форми, відома також стратифікація (розшарування) матеріалів Землі за щільністю і т. д. Не випадкові також по-своєму добрі канонічні результати математичних теорій – абстрактні закони природи. Геніальне – просте.

Але цим не обмежується роль класифікації й роль самого прагнення до пошуку аналогів. **Пошук аналогів – це істотний елемент будь-якої творчої роботи.** Від будь-якої «найрозумнішої» машини людина відрізняється за своїми здатностями тим, що вона здатна до творчості, а машина – ні. Людина може придумати нове, а машина – ні. Розібратися в цих питаннях – це створити, власне кажучи, передумови для розроблення штучного інтелекту.

Дотепер під штучним інтелектом розуміється лише здатність ЕОМ або програм до їх адаптації до яких-небудь складних умов. Будь-яка адаптація є, власне кажучи, автоматичне регулювання певних впливів або умов. Тут немає творчості, а є виконання логічних відносин по типу імплікацій: «якщо ..., то..., інакше...». У програмах це реалізується звичайно у вигляді умовних операторів: «if ..., do ..., else do...».

У людини мислення й творчість відбувається з використанням багатьох рівнів абстракції. Під терміном «абстракція» можна розуміти – узагальнення в якомусь сенсі. При узагальненні втрачаються деякі ознаки конкретних аналогів, але зате залишаються більш загальні ознаки. Тобто аналоги бувають не тільки за складом будь-яких ознак (за принципом дії, створення, за умовами й т.д.), але й за рівнями абстракції. Тому нижній рівень образів, знання щодо верхнього рівня абстракції є конкретикою, а верхній рівень щодо нижнього є узагальненням – абстракцією. Абстракцією може бути не тільки класифікація або узагальнення сутностей та ідей, у тому числі при переході від матеріальної сфери до сфери ідеальних образів. А також це принципи, загальні властивості, закони природи в матеріальній, суспільній і соціальній сферах, спостереження, елементи мудрості й т.п.

При класифікації рівень абстракції (підгрупи, групи, підкласи, класи) збільшується з узагальненням понять об'єктів, предметів, процесів, а також зі зменшенням числа відмітних ознак, властивостей. І в абстрактній області, такій як математика, також існують свої рівні абстракції. Наприклад, це насамперед арифметика, алгебра, функціональний аналіз, диференціальне й інтегральне обчислення й т.д. Арифметика – це абстракція типу: неважливо «чого», а важливо – «скільки». Абстрагуємося від

фізичної сутності об'єктів, процесів. В алгебрі абстрагуємося від фізичної сутності й від кількості: неважливо «чого» й «скільки», а важливо, що ми з ними робимо, і т.д. Тому теореми й стійкі властивості логічних побудов – це теж, але тільки абстрактні, закони природи. А жити в природі й не знати найголовніші її закони – це не тільки соромно в культурному суспільстві, але й невігідно для кожного.

Рівні абстракції звичайно взаємообумовлені й взаємозалежні між собою. Вони можуть розгалужуватися, бути незалежними або утворювати замкнуті в кільце зв'язки або утворювати інші конфігурації зв'язків. Узагальнення звичайно називають індукцією, а перехід до частки (до конкретики) – дедукцією.

Шерлок Холмс при роботі застосовував метод дедукції, тому що використовував загальні, відомі йому принципи, знання й закономірності для конкретних, часткових застосувань. А от свої специфічні знання він одержував, як і в науці, методом індукції. Оскільки у своїх спостереженнях він збирав факти й учився їх узагальнювати, тобто знаходити стійкі ознаки, принципи, відомості й т.п. Звідси випливає, що у своїй діяльності людина використовує які-небудь комбінації цих методів. І хто правильно це вміє робити, той буде успішнішим у своїх справах.

Сукупність знань будь-яких рівнів абстракції настільки багата й об'ємна, що з великою ймовірністю в ній можна знайти, якщо не те, що ви шукаєте, то хоча б аналог якої-небудь абстракції. Здатність до такого пошуку і є творчість. Тому не даремно говорять, що нове – це добре забуте старе. Однак для розвитку творчості це прислів'я слід істотно уточнити: **нове – це добре забуте** (частіше аналогічне !) **старе**. А що таке добре забуте старе? Це, власне кажучи, – старий аналог уже затребуваного нового, можливо, у нових умовах, з новою якістю, з новою конкретикою.

2. ЗАГАЛЬНІ ПИТАННЯ ОПТИМІЗАЦІЇ ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ РАДІОТЕХНІЧНИХ СИСТЕМ

У розділі викладений сучасний стан проблеми оптимізації інформаційно-вимірювальних радіотехнічних систем (ІВРТС) за багатьма показниками якості, поняття про подвійність розв'язання задач оптимізації й поняття про узагальнений показник якості, про критерії якості й про ефективність систем.

При подальшому викладі використовується попередня в розділах анотація й завдання на прочитання, які дозволяють «за деревами побачити ліс», тобто за подробицями викладу побачити головні ідеї.

Завдання на прочитання

Відповідно до викладених принципів У. Оккама, Р. Декарта, Р. Беллмана й інших принципів необхідне таке:

- 1) виявити ідею актуальності оптимізації й методи оптимізації РТС;
- 2) порівняти недоліки, відзначені в розділі, з відомими до цього недоліками відомих систем того ж класу;
- 3) виділити головне з того, що пропонується для поліпшення систем;
- 4) зрозуміти, що таке «подвійність» розв'язання задачі;
- 5) зрозуміти, що таке «ефективність» системи;
- 6) зрозуміти, що таке показник і критерій якості.

2.1. Сучасний стан проблеми оптимізації ІВРТС за багатьма показниками якості

Зміст даного розділу розраховано на студентів старших курсів, ознайомлених з радіотехнічними системами, або тих, хто прочитав цю книгу в першому читанні. Іншим повинні бути зрозумілими основні ідеї й особливості проблем оптимізації ІВРТС.

Про головні ідеї. Радіотехнічні системи завжди описуються значним числом показників якості. При цьому вони, як правило, суперечливі. Це означає, що якщо при проектуванні РТС ми використовували один показник якості й ухвалили рішення щодо технічних параметрів, сигналів і структури системи, то ввівши додатково необхідні показники якості, ми вводимо, власне кажучи, додаткові обмеження на перший показник. Отже, перший показник може значно погіршитися, якщо, звичайно, його необхідне значення ще можна реалізувати.

Тому, щоб хоч якось зменшити збиток від погіршення першого показника якості, потрібно його оптимізувати. Слід зазначити, що вже зараз замовник системи не допустить, щоб система була неоптимальною. Що таке «оптимальна або краща» система описано в підрозділі «Ефективність ІВРТС».

Задачі такого (оптимального) вибору технічних параметрів, сигналів і структур, при яких досягається краще значення основного показника якості при заданих обмеженнях, називаються задачами оптимізації, або задачами оптимального синтезу зазначених параметрів, сигналів і структур.

Саме постановка й розв'язання таких задач для ескізного проектування ІВРТС і є поки що найбільшою складністю. Ще нещодавно це вважалося взагалі неможливим, якщо враховувати всі показники якості системи. Відомі такі досягнення в галузі синтезу РТС і складних систем за сукупністю показників якості [2-5, 7-10] та ін. У роботах [2, 3 й ін.] викладені, з одного боку, досить загальні методи оптимізації систем, з іншого боку, дані часткові приклади оптимізації за двома-трьома показниками якості, що не дозволяють із єдиних позицій, одним математичним апаратом робити оптимізацію РТС за значним числом змінних і показників якості. Ці й інші відомі роботи із синтезу сигналів, структури вимірювачів і технічних параметрів є частковими й обмеженими або занадто абстрактними, щоб їх використовувати, і не дають відповідей на питання про можливість загального оптимального синтезу РТС, про взаємозв'язки часткових видів синтезу, про єдину ідеологію вимірювань та ін.

Питання про склад показників якості й параметрів системи виникати не повинно. По-перше, кращого обґрунтування складу показників якості ІВРТС, ніж це зроблено в тактико-технічних

вимогах (ТТВ) до реальної системи, навряд чи можна чекати. Запропонований метод синтезу ІВРТС дозволяє уникнути ранжування показників якості, оскільки використовує рівноправні, а не лексикографічні [2, 3] критерії якості. По-друге, методи рішення в принципі дозволяють урахувати майже всі основні технічні параметри системи, тому їх склад також не має потреби в спеціальному обґрунтуванні. При виборі класу сигналів, якщо врахувати сучасні вимоги до них та зміст робіт [11, 12], варто віддати перевагу широкосмуговим сигналам, які при їхній кореляційній обробці успішно використовуються для підвищення завадостійкості стосовно вузькосмугової завади і які допускають сигнальну й апаратурну сумісність і створення значних ансамблів сигналів з хорошими взаємкореляційними властивостями при передачі інформації. Оптимальний вибір структур вимірювальних каналів за заданими ТТВ вже не являє собою складну проблему, оскільки в даній частині книги зазначено, як його робити. У книзі викладено, як оптимально вибирати технічні параметри з урахуванням показників, узятих із ТТВ. У будь-якому разі, для складних систем зазначено, як це можна зробити.

Загального синтезу або оптимізації ІВРТС за списком основних показників із ТТВ шляхом одночасного знаходження відповідних технічних та сигнальних параметрів і структур ще не було. Однак у даній книзі такий метод запропонований. Отже, загального синтезу не було через складність проблеми й з таких причин:

- 1) існуючі теорії радіовимірювань не завжди чітко використовують фізичне трактування вимірювань;
- 2) синтезу параметрів систем за заданим ТТВ практично не було;
- 3) не було повної класифікації реальних вимірювачів і методів оцінювання параметрів сигналу, тому що пошуковий метод в ІВРТС є допоміжним і не призначається для оцінювання параметрів сигналу, не було поняття багатоетапного (комбінованого) методу оцінювання параметра сигналу, не було загальної системної оцінки ефективності багатоетапних і багатошкальних вимірювачів;

4) існуючий синтез сигналів носить у значній мірі евристичний характер.

Матеріали підручника дозволяють оптимально вибрати вимірювальний сигнал. Не випадково сигнал рекомендується шукати в класі широкосмугових сигналів. При цьому він задовольняє вимоги суміщення каналів, високої точності й завадозахищеності. Крім того, він повинен задовольнити також вимоги відповідності заданим діапазнам вимірюваних параметрів і розв'язки приймального і передавального трактів, особливо, якщо вони працюють з об'єктами при малих кутах місця, умови узгодження шкал або етапів та інші вимоги, наприклад такі, як розмірність ансамблю сигналів із заданими взаємкореляційними властивостями й т.п.

Необхідність ухвалення рішення конструкторами РТС про технічні параметри була завжди при створенні нових систем. У цей час розробників систем не можуть влаштувати звичайні, інтуїтивні, рішення на припустимій множині технічних параметрів, що призводять до істотного зниження ефективності систем. Звідси впливає актуальність проблеми оптимального синтезу параметрів і спроби її рішення [4, 5], викладені також далі. Елементи параметричного синтезу існували у вигляді часткових задач прийняття оптимального рішення про технічні параметри за критеріями завадостійкості [5], максимуму енергетичного потенціалу [4] при обмежених асигнуваннях.

До основних причин відсутності дотепер рішення проблеми оптимального синтезу технічних параметрів РТС можна віднести таке:

1) складність проблеми, різнорідність і маловивченість зв'язків технічних параметрів з показниками якості РТС;

2) нерозробленість єдиних досить простих методик постановок і розв'язання задач синтезу технічних параметрів РТС за багатьма критеріями якості для великої розмірності задач, тобто для великої кількості параметрів;

3) неповнота, нечіткість (відсутність закону розподілу величини) і нестабільність техніко-економічних, масогабаритних та інших даних про показники якості РТС.

Характеристика зазначених причин.

Складність вирішення проблеми полягає в багатофункціональності систем і багатокритерійності їх опису. Зв'язок показників якості з технічними параметрами достатньо вивчений для систем зв'язку [7, 9] й у меншій мірі – для ІВРТС, і особливо – для сумісних систем, їх опису. Незважаючи на добре розвинені й відомі [13, 14] методи аналізу різних факторів, що впливають на якість передачі сигналів у радіолінії, ще недостатньо вивченими є багатofакторні процеси обробки сигналів, особливо складних. Припущення про синтез високоякісних систем у даній книзі, як не здається це дивним, істотно спрощує розв'язання задачі оптимізації. Тоді система працює з більшим відношенням сигнал-шум, що, по-перше, дозволяє порівняно просто врахувати вихідний ефект у вигляді співмножників та оцінити точність його представлення в показниках якості й, по-друге, дозволяє вважати досить стійкими числові характеристики (моменти) розподілів випадкових процесів, які в ряді випадків описують міру відповідної якості ІВРТС.

Таким чином, а також завдяки використанню сучасних відомих представлень і готових результатів [4, 5] й ін., вдається спростити постановку відповідної задачі загального синтезу, що є задачею стохастичного програмування [15], тобто задачею математичного програмування, де є хоча б один випадковий показник або параметр. Упевненість у стійкості розв'язання цієї задачі впливає більше з фізичної її постановки, ніж з розв'язання її класичними методами.

Складність рішення проблеми безпосередньо відбувається також від числа й багатоваріантності представлення показників якості і їх розмірності («прокляття багатовимірності»). Це є підставою й другою причиною відсутності рішення проблеми загального синтезу ІВРТС. Оскільки не існує універсальних методів пошуку оптимального рішення в математичному програмуванні, то при класичному підході до вирішення проблеми не може бути єдиної методики оптимізації ІВРТС за сукупністю показників якості, яка б лягла в основу теорії такої оптимізації ІВРТС. Виникає дилема: чи точно розв'язувати щораз конкретну часткову багаторозмірну задачу математичного програмування на умовний екстремум у числовому вигляді за

великий інтервал часу й ще з не цілком точними вихідними даними, чи застосувати спрощення й універсалізацію постановки задачі за рахунок використання властивостей показників якості. Останнє дозволяє істотно скоротити час розв'язання, автоматизувати проектування систем і для простих типів обмежень одержати рішення в аналітичному вигляді.

Такий шлях, що викладений у цій книзі, припускає:

1) приведення показників якості з обмежень до єдиної форми перетворенням координат, що для багатьох прикладних задач не є проблемою;

2) лінеаризацію цільової функції (при освоєнні методу надалі можна використовувати квадратичні форми, гесіан тощо);

3) аналітичне розв'язання отриманої в такий спосіб задачі виродженого динамічного програмування або застосування розв'язання з парку відомих задач [4, 5];

4) використання при необхідності розв'язання такої спрощеної задачі в якості ітераційної формули до необхідної близькості до попереднього рішення.

Третя причина відсутності рішення проблеми синтезу ІВРТС обумовлена специфікою розвитку радіотехніки, спрямованої в основному на поліпшення так званих головних показників якості (завадозахищеності, точності й ін.), і відстаючої від інших галузей виробництва в питаннях обліку техніко-економічних і масо-технічних даних. Запропонований тут шлях вирішення проблеми синтезу ІВРТС органічно враховує цю особливість статистичних даних.

Характеристику причин відсутності рішення проблеми загального, тобто всеосяжного, синтезу ІВРТС варто доповнити також проблемами гносеологічного типу. Якою повинна бути задача загального синтезу? Лексикографічною з елементами евристики при ранжуванні показників якості, при послідовних поступках тощо? За зваженим критерієм з евристичними коефіцієнтами? За відносним критерієм з обмеженою областю визначення функціонала? Де знайти для ІВРТС узагальнений критерій якості або функцію корисності? Яка порівняльна ефективність згаданих постановок задач. І взагалі, що таке «ефективність». Строгих вказівок на цей рахунок поки немає, однак книга містить ряд таких відповідей. Як буде викладено

далі, оптимізація за умовним критерієм якості містить найменше інтуїтивних припущень (які вносяться, як і скрізь, тільки при відборі показників якості), а складність проблеми буде переборена вищевказаним шляхом, що буде видно також і далі.

2.2. Про подвійність багатокритерійних задач

Як зазначалося в попередньому підрозділі, умовний критерій якості має недоліки: складність розв'язання задач і подвійність розв'язання. Складність розв'язання обумовлена, крім форми виразів для цільової функції й обмежень, необхідністю знаходження множників Лагранжа або розв'язання задачі методом підстановки. При методі підстановки кожне обмеження розв'язується щодо однієї зі змінних, котра підставляється в цільову функцію. Труднощі полягають у тому, що далеко не завжди це вдається зробити. Не менш складним може виявитися алгоритм визначення множників Лагранжа. Ці обставини змушують звернутися до проблеми подвійності задач. Що ж таке подвійність розв'язання задач. Логіка така. Показники якості системи ми вважаємо рівноправними, або рівноцінними. Виходить, однаково, який показник якості системи ми приймемо за цільову функцію. З фізичної сутності задач відчувається, що при однакових значеннях показників якості оптимальне розв'язання повинне бути те саме. Однак це враження оманливе: можливо одержати й різні розв'язання. Це і є один з типів подвійності. Звичайно, якби ми були впевнені в тому, що розв'язання для всіх постановок задач те саме, то потрібно було б знайти лише одне яке-небудь розв'язання. Це істотно спростило б розв'язання задач. Розглянемо умови, при яких розв'язання задач єдине. Для простоти розглянемо дві задачі зі зворотною постановкою: у другій задачі обмеження є цільовою функцією, а цільова функція - обмеженням з тим числовим значенням показника якості, що було кращим з першої задачі.

$$\begin{aligned} & \text{Задача 1} \\ \hline \min D_1(\bar{X}) &= D_{1\partial}; \\ D_2(\bar{X}) &= D_{2\partial}; \end{aligned}$$

$$\begin{aligned} & \text{Задача 2} \\ \hline \min D_2(\bar{X}) &= D_{2\partial}; \\ D_1(\bar{X}) &= D_{1\partial}. \end{aligned}$$

Причому $\min D_1(\bar{X})$ першої задачі дорівнює значенню $D_{1\partial}$, що є рівнем обмеження другої задачі. Нехай вирази $D_1(\bar{X})$ і $D_2(\bar{X})$ – унімодальні функції, тобто функції з одним екстремумом. Ця умова для технічних задач необтяжлива, оскільки для систем вони навіть монотонні: тільки зростають (або спадають), залежно від змісту показників якості. Зобразимо їх на рис. 2.1 лініями однакових рівнів у двовимірному просторі. А і В – мінімуми відповідно функцій $D_1(\bar{X})$ і $D_2(\bar{X})$. Якщо обмеження $D_2(\bar{X}) = D_{2\partial}$ – досить жорстке, що відповідає малому значенню $D_{2\partial}$, тобто вимозі до високої якості системи, то лінія рівня $D_2(\bar{X}) = D_{2\partial}$, швидше за все, буде розташовуватися ближче до мінімуму В і проходити між мінімумами А і В. Знайдемо розв’язання задачі 1 на рис. 2.1, тобто мінімум функції $D_1(\bar{X})$ при $D_2(\bar{X}) = D_{2\partial}$ (пунктирна лінія). Очевидно, що на рівні $D_{2\partial}$ мінімум функції $D_1(\bar{X})$ досягається в точці С, що розташовується на лінії рівня $D_{1\partial}$. Виходить, С – розв’язання \bar{X}_1 задачі 1. Якщо візьмемо лінію рівня $D_1(\bar{X}) = D_{1\partial}$ як обмеження другої задачі й будемо знаходити на ній мінімум $D_2(\bar{X})$, то виявимо, що розв’язання \bar{X}_2 задачі 2 також належить точці С, тобто $\bar{X}_2 = \bar{X}_1$. Очевидно, що значення показників якості $(D_{1\partial}, D_{2\partial})$ однакові для двох задач. Аналогічно можна припустити й для багатовимірних і багатокритерійних задач, хоча останнє ще не доведено. Неважко помітити, що сказане справедливе також для обмежень типу нерівності, тобто $D_2(\bar{X}) < D_{2\partial}$ в задачі 1 і $D_1(\bar{X}) \leq D_{1\partial}$ у задачі 2. Зазначимо, що властивість опуклості не завжди є обов’язковою. Обов’язковою є монотонність спаду, що виконується практично завжди. Таким чином, при виконанні зазначених умов однозначність розв’язання двох задач виконується для високоякісних систем. Саме їх і доцільно розглядати. Однак переконаємося в тому, що властивість незмінності розв’язання, тобто інваріантності його щодо постановки задач не виконується для систем порівняно більш низької якості. Це відповідає великому $D_{2\partial}$ (рис. 2.2,

пунктирна лінія), коли мінімум A функції $D_1(\bar{X})$ знаходиться в середині лінії рівня $D_2(\bar{X}) = D_{2\partial}$ (рис. 2.2). Тоді при обмеженні $D_2(\bar{X}) = D_{2\partial}$, тобто на відповідній лінії рівня, мінімум $D_1(\bar{X})$ знаходиться в точці C (рис. 2.2). Це розв'язання \bar{X}_1 задачі 1. Якщо шукати на лінії рівня $D_1(\bar{X}) = D_{1\partial}$ розв'язання \bar{X}_2 другої задачі, тобто $\min D_2(\bar{X})$ при $D_1(\bar{X}) = D_{1\partial}$, то виявиться, що цей мінімум знаходиться в точці D , а не C . Звідси $\bar{X}_1 \neq \bar{X}_2$ і показники якості $(D_{1\partial}, D_{2\partial})$ і $(D'_{1\partial}, D_{2\partial})$ також неоднакові. У випадку обмежень типу нерівностей, розв'язання задачі 1 знаходиться в точці A , а розв'язання задачі 2 – у точці D . Помітивши, що система, для якої вектор $(D'_{1\partial}, D_{2\partial})$ істотно краще $(D_{1\partial}, D_{2\partial})$, оптимальна й для неї розв'язання зворотної задачі інваріантне. Тому будемо прагнути досягти саме цього розв'язання $\bar{X}_2 = D$.

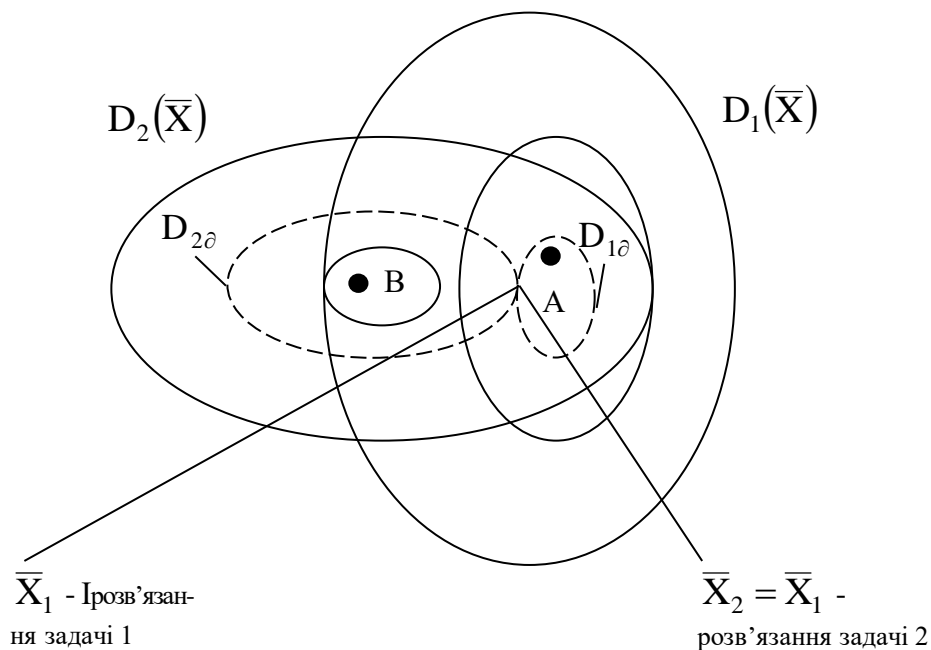


Рис. 2.1

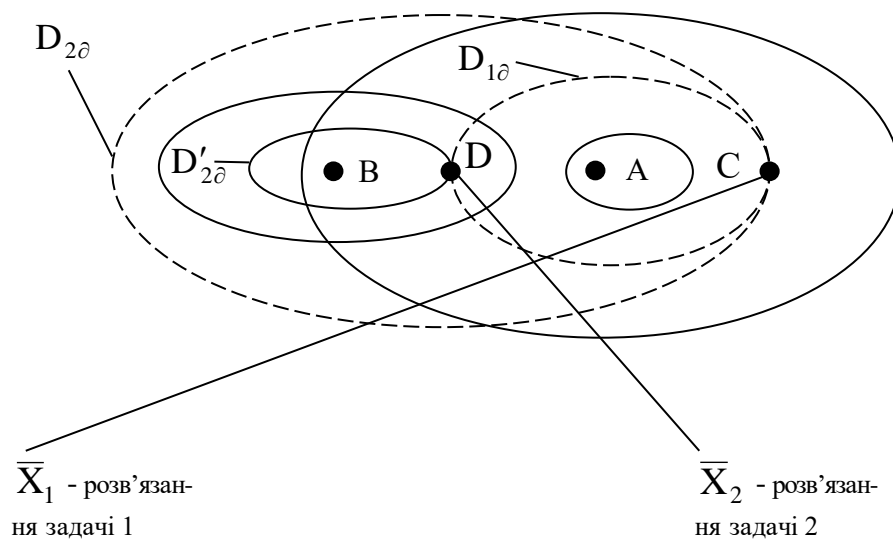


Рис. 2.2

Для того, щоб визначити, чи наявна властивість інваріантності задач на умовний екстремум, можна запропонувати таку методику. Необхідно лише перевірити зазначений факт про те, що мінімуми А і В (рис. 2.1) відповідно до функцій $D_1(\bar{X}), D_2(\bar{X})$ знаходяться з різних сторін від кривої $D_2(\bar{X}) = D_{2\partial}$. Якщо функції унімодальні або монотонні, то інваріантність існує при виконанні умови

$$D_2(\bar{X}) > D_{2\partial}.$$

Для монотонної цільової функції й обмеження щодо своїх змінних властивість інваріантності розв'язання умовних задач витримується автоматично, оскільки в цьому випадку $\bar{X} = 0$, а $\bar{X} \rightarrow \infty$ і будь-якій припустимій кривій $D_2(\bar{X}) = D_{2\partial}$ лежить між цими точками мінімуму.

І ще одна важлива властивість: якщо умова інваріантності розв'язання задач витримується, то для монотонних показників якості вона справедлива й для обмежень типу нерівностей, що очевидно з рис. 2.1.

Контрольні питання

1. У чому полягають труднощі теорії РТС?
2. Чи покращуються результати оптимізації ІВРТС при збільшенні числа обмежень?
3. Який критерій якості ІВРТС найбільш об'єктивний і необмежений в області визначення?
4. Які недоліки умовного критерію якості?
5. Що таке подвійність задач оптимізації за умовним критерієм якості?
6. Які умови інваріантності розв'язання задач оптимізації стосовно різних постановок задач?
7. Які умови вибору найкращого розв'язання задач за умовним критерієм якості?

2.3. Ефективність складних інформаційно-вимірювальних систем

Під ефективністю систем у даний час розуміють ступінь поліпшення узагальненого показника якості.

Оскільки всі радіотехнічні системи описуються сукупністю показників якості й обмеженнями на ресурси, за умовами роботи й т.п., виникає питання про те, як у цих умовах визначити найкращу ефективність систем і відповідне оптимальне рішення про її параметри, структуру й сигнали. Через те, що велика кількість параметрів і деякі показники якості – числові характеристики випадкових процесів, оптимальне рішення (якщо сформулювати задачу) буде носити евристичний характер. Тому закономірне питання про те, чи набагато погіршиться ймовірність розв'язання, якщо евристика існує й при постановці задачі. Проте, будемо дотримуватися принципу: чим менше евристики, тим якісніше розв'язання при інших рівних даних. Це означає, що вектору рівноцінних показників якості системи віддається перевага перед лексикографічними критеріями якості \vec{D} .

Інформаційні й вимірювальні системи описуються такими показниками якості:

1) точністю, що визначається середньоквадратичною σ_λ похибкою вимірювань параметра λ ;

2) апріорним діапазоном, що визначається дисперсією $\sigma_{\alpha\lambda}^2$ ймовірності апріорного розподілу вимірюваного параметра;

3) вартістю системи С;

4) завадостійкістю або надійністю вимірювань; що визначаються як імовірністю безпомилкової або безвідмовної роботи $P_{бр}$, так і довірчою ймовірністю $P_{дов}$, оцінювання параметра λ ;

5) часом спостереження параметра сигналу.

Примітно, що вартість, як й інші показники РТС, цілком можна вважати рівноправним за значущістю показником якості.

У результаті системного аналізу можуть бути визначені залежності показників якості $\vec{D}(\vec{X})$ від вектора технічних параметрів.

Якщо відома функція (функціонал) мети $f(\vec{D}(\vec{X}))$, то найкращою буде система з такими оптимальними технічними параметрами \vec{X}_0 , при яких досягається $\text{extr}_{\{\vec{X}\}} f(\vec{D}(\vec{X}))$ на множині $\{\vec{X}\}$:

$$\text{extr}_{\{\vec{X}\}} f(\vec{D}(\vec{X})) = f(\vec{D}(\vec{X}_0)).$$

Функціонал $f(\vec{D}(\vec{X}))$ звичайно виступає в ролі ймовірності досягнення мети $P_y = P_y(\vec{D}(\vec{X}))$. Однак найчастіше в складних системах функціонал $f(\vec{D}(\vec{X}))$ визначити не вдається.

В цих умовах використовується один із зазначених у розд. 1 критеріїв оптимальності.

Задачу на умовний екстремум при наявності обмежень на параметри при будь-яких додаткових умовах називають також **математичним програмуванням**.

Якщо розв'язання задач на умовний екстремум не залежить від того, яку функцію візьмемо як функцію мети, тобто при будь-якій постановці задач розв'язання збігається, цю властивість будемо називати інваріантністю розв'язання до постановки задачі на умовний екстремум.

Умовний критерій позбавлений зазначених раніше недоліків, однак йому властива складність розв'язання задач і можливість відсутності інваріантності розв'язання в постановці задачі. Складність розв'язання спеціальних задач можна обійти [4, 5]. І для більшості реальних задач, де потрібен синтез високоякісних систем і де залежності $D_i(\bar{X})$ є монотонними, інваріантність розв'язання переважно зберігається.

Якщо властивість інваріантності розв'язання до постановки задачі на умовний екстремум виконується, то для порівняння ефективності систем можна вибрати як цільову функцію будь-який показник якості.

Задачі на умовний екстремум мають такі методи розв'язання: 1) метод множників Лагранжа; 2) метод підстановки параметрів з функцій зв'язку (з обмежень), що перетворює задачу в задачу на безумовний екстремум; 3) метод математичної індукції; 4) метод геометричного програмування для ряду сепарабельних задач.

Однак такі задачі синтезу параметрів систем варто розв'язувати на всій дискретній множині структур.

Властивість інваріантності розв'язання до постановки задач справедлива принаймні для однопараметричних вимірювальних систем. Навіть для таких систем порівняльні оцінки ефективності одержати непросто: якщо склад показників якості значний, якщо техніко-економічні залежності задані з обмеженою точністю, якщо система працює з об'єктами, що кооперуються, і якщо енергетичний потенціал радіолінії залежить від великої кількості параметрів системи. Проте такі задачі порівняння вимірювальних структур також можливо розв'язувати.

3. СИСТЕМНИЙ АНАЛІЗ ЯКОСТІ ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ РАДІОТЕХНІЧНИХ СИСТЕМ

У даному розділі наведені відомі й невідомі в літературі **результати системного аналізу** якості ІВРТС, що завжди передує задачам оптимізації або синтезу РТС за узагальненим критерієм якості. **Задачею системного аналізу є дослідження впливу на показники якості РТС технічних параметрів системи і параметрів паразитних факторів: розстроювань, збурювань і неідеальностей сигналу й схем.**

Основними труднощами таких досліджень є вивчення спільного впливу всіх паразитних факторів, які можуть бути випадковими й нечіткими. Тому **центральною ідеєю** в розділі є припущення про те, що вплив усіх паразитних факторів на вихідний сигнал досить малий. Це дає можливість спростити задачу оптимізації. При цьому не потрібно враховувати одночасний вплив усіх параметрів, що вважається неможливим. Необхідно тільки знати одновимірні залежності показника від параметрів. Показано, що при малих впливах **результуюче ослаблення енергетичного потенціалу виражається як добуток усіх одновимірних ослаблень.** Зовсім не враховувати спільний вплив всіх факторів не можна, оскільки цих факторів багато.

Завдання на прочитання

1. Зрозуміти принцип урахування спільного впливу технічних параметрів і параметрів паразитних факторів.
2. Звернути увагу, що одночасний вплив паразитних факторів на енергетичний потенціал істотно ускладнює задачу дослідження. РТС при цьому навряд чи буде близька до оптимальної.
3. Звернути увагу на новий ефект впливу похибки стабілізації діаграми спрямованості антени за напрямком. Він впливає при супроводі рухомих об'єктів, у радіорелейних лініях передачі й т.д. При малих відхиленнях цей ефект із великою точністю спостерігається також у синхронному детекторі, у виділювачі тактової частоти (джитер) і т.д. Обговорити питання

про те, чи можна використати ці ж результати для випадку синхронного детектування.

4. Звернути увагу на вплив вибігу тактової частоти на вихідний потенціал радіолінії. Цей же ефект, який називається «вандером», впливає також у цифрових лініях передачі. При цьому оцінка припустимої довжини фазоманіпульованого сигналу справедлива також для оцінки розмірів кадру, фрейма.

Оптимізація радіотехнічних систем за основними показниками якості повинна складатись з нижченаведених етапів.

1. Вибір, обґрунтування й розроблення складу показників якості РТС, які впливають або з результатів розв'язання задач, прийняття рішень для задовільного досягнення поставленої мети (дослідження операцій) на більш високому ієрархічному рівні структури системи, або, при обмеженій інформації про оптимальне рішення, з результатів оптимізації відомих систем приблизно того ж класу, або методом експертних оцінок.

2. Системний аналіз задачі оптимізації РТС, що являє собою специфічні дослідження взаємозв'язку й взаємозумовленості показників якості систем та їх технічних параметрів. При цьому, як відзначав Беллман [1], дослідник іде по вузькій стежині між пасткою переспрошення та болотом переускладнення. Це означає, що навіть підготовка аналітичного математичного апарата теж повинна бути оптимальна, хоча б евристично. З одного боку, потрібно, щоб отримана залежність із достатньою точністю відбивала сутність досліджуваного ефекту, а з іншого боку, щоб вона не була занадто складною й громіздкою, через що задача може бути нерозв'язною.

Досвід розв'язання задач оптимізації систем показує, що звичайно досить мати лише перше наближення ефекту їхнього функціонування, оскільки в задачах оптимізації РТС точність задання хоча б деяких показників якості й знання окремих технічних, найчастіше випадкових, параметрів є обмеженою. Тому при врахуванні ефектів або результатів функціонування систем у першому наближенні необхідно враховувати також усі прийняті обмеження й припущення, які необхідні при складанні остаточної програми оптимізації РТС за параметрами.

3. Формалізація задачі, яка полягає в тому, що з'ясовані зв'язки між показниками або критеріями якості й технічними

параметрами подані у формі математичної задачі знаходження кращих значень показників якості систем із заданого списку тактико-технічних вимог або у формі знаходження умовного критерію якості, що багато в чому кращий.

4. Розв'язання задачі визначення найкращого в прийнятому сенсі вектора (списку) значень показників якості й вектора значень оптимальних технічних параметрів системи. При цьому розв'язання задач в аналітичному вигляді має ряд переваг перед числовими методами, які тільки й зустрічаються в математичному програмуванні. Переваги в тому, що аналітичне розв'язання дає більше інформації про характер і природу рішення, про його стабільність, про вплив на нього показників систем і що розв'язання придатне для широкого класу систем, умов, обмежень і діапазонів змін показників і технічних параметрів.

Перший етап залежить від призначення РТС і задач, розв'язуваних нею самостійно або в складі комплексу систем. Третій і четвертий етапи істотно залежать від першого й другого етапів оптимізації. Щодо цього другий етап є найменш залежним від інших. По-перше, залежності, що відбивають вплив якого-небудь фактора на показник якості, можуть бути досить загальними, використовуваними в багатьох показниках. По-друге, задачу оптимізації завжди можна спростити, зафіксувавши значення якого-небудь фактора, тобто виключивши його із числа оптимізованих у цей момент параметрів задачі, що дозволяє, маючи максимально можливий набір залежностей оптимальних показників якості від оптимального розв'язання, здійснювати подальшу оптимізацію і за фіксованими параметрами або показниками. Це може бути подане у вигляді блокового програмування, у вигляді динамічного програмування і т. д. Виходячи із зазначених міркувань, а також з того факту, що значне число показників якості систем, що описують результати функціонування систем, залежить від відношення сигнал/шум, задачею оптимізації РТС подамо системний аналіз впливів різних факторів і їх сукупності на показники якості систем.

3.1. Вплив збурювань, нестабільностей і неідеальностей сигналів і схем на кореляційну й фільтраційну обробку сигналу

При обробці сигналу з метою його виявлення або вимірювання його параметрів велику роль відіграє формування вихідного сигналу, що будемо розглядати на вході кінцевого, або вирішального пристрою. У процесі системного аналізу оцінюється вплив різних збурювань, неідеальностей схем або апаратури і нестабільностей параметрів її елементів на показники якості систем. Оцінювання найчастіше здійснюється в порівнянні з ідеальною системою з так званими потенційними можливостями. У ролі такого (опорного) рішення, або в ролі моделі, приймається система з потенційними можливостями відносно виконання вимог до таких показників якості, як завадостійкість, точність, що уточнюються функціями впливу неідеальностей. Припущення про те, що реальна система високоточна й завадостійка, дозволяє нам спростити постановку задач оптимізації, оскільки простіше утворяться такі залежності показників від параметрів неідеальностей. Це буде видно з подальших прикладів. Але найголовніше – це те, що саме такі найкращі системи нам і потрібні. Зазвичай кількість факторів збурювань, нестабільностей і неідеальностей X_i у системі велика.

Їх спільний вплив на показник якості або хоча б на вихідний сигнал ще не вивчено в достатній мірі, оскільки ці фактори не завжди незалежні. Однак якщо припустити досить типову ситуацію, коли вплив кожного фактора досить малий, то результат їхніх впливів може бути поданий у сепарабельному вигляді, тобто у вигляді розділених функцій: у вигляді суми функцій або у вигляді добутку. Доведемо це положення.

Показник якості виявлення сигналу або вимірювання його параметра залежить від вихідного сигналу або від сигнальної функції

$$\Psi(\alpha, X_1, \dots, X_n) = \operatorname{Re} \int_0^T S(t, \alpha) S(t, \alpha_1, \bar{X}) dt ,$$

де $S(t, \alpha)$ і $S(t - \tau, X_1, \dots, X_n)$ – відповідно опорний і прийнятий сигнали.

Розкладемо прийнятий сигнал у ряд Маклорена щодо нормованих параметрів неідеальностей схем, збурювань і нестабільностей

$$S(t, \alpha, \bar{X}) = S(t, \alpha, 0) + \sum S_i^1 X_i + \sum \sum S_i^1 S_j^1 X_i X_j + \dots$$

Тоді

$$\Psi(\alpha, \bar{X}) = \Psi(\alpha, 0) + \sum \Psi^1 X_i + \sum \sum \Psi_i^1 \Psi_j^1 X_i X_j + \dots$$

або

$$\Psi(\alpha, \bar{X}) = \overline{\Psi(\alpha, 0) \prod (1 + \Psi^1 X_i / \Psi(\alpha, 0))}, \quad (3.1)$$

що доводить незалежність впливу неідеальностей при малих, приблизно рівновеликих X . Висновок про незалежність впливу неідеальностей не є несподіваним, оскільки виконання умови малості неідеальностей призводить до того, що похідні, які старші першої, виявляються малими вищого порядку. Геометрично це означає заміну взагалі нелінійної гіперповерхні гіперплощиною поблизу 0 , де наближення справедливе. Практично досить, щоб $\Psi(X) \leq 1/2$. Таким чином, піклуючись про створення на виході системи обробки сигналу максимально можливого відношення сигнал/шум, ми звертаємось до погодженої фільтрації сигналу або до кореляційної його обробки. При цьому на виході цієї системи ми одержуємо вихідний сигнал, пропорційний кореляційній функції вхідного сигналу. Тому доцільно розглянути вплив збурювань, неідеальностей схеми, відхилень і нестабільностей параметрів її елементів на вихідне відношення сигнал/шум, оскільки воно наявне практично у всіх показниках якості, що характеризують ефект.

Відомі роботи [13, 16], у яких описано вплив відхилень характеристик фільтра й сигналу на відношення сигнал/шум. Наведемо тут деякі з них.

У роботі [13] описаний вплив відхилення форми імпульсу, який має вигляд трапеції, від прямокутної форми на відношення сигнал/шум q . При цьому відношення сигнал/шум зменшується в G разів:

$$G = \frac{q_{opt}}{q} = \frac{16(2 + X_{\varrho})}{3(3 + X_{\varrho})}, \quad (3.2)$$

де X_{ϱ} – відносна тривалість вершини імпульсу, $X_{\varrho} = \frac{\tau_1}{\tau}$;
 q_{opt} – відношення сигнал/шум оптимального фільтра;
 q – відношення сигнал/шум неоптимального фільтра.

За рахунок зміни тривалості τ прямокутного відеоімпульсу до τ_1

$$G = \begin{cases} \frac{\tau}{\tau_1} & \tau_1 \leq \tau; \\ \frac{\tau_1}{\tau} & \tau_1 \geq \tau. \end{cases} \quad (3.3)$$

Нехай амплітудно-частотний спектр сигналу зазнає «косинусні» спотворення $K(\omega)$ виду:

$$K(\omega) = a_0 + a_1 \cos \omega c, \quad (3.4)$$

де $c \geq 2\tau_1$,

τ_1 – тривалість сигналу;
 a_1 – рівень спотворювань;
 a_0 – загасання;
 ω – колова частота.

Тоді

$$G = 1 + \frac{a_1}{2a_0}. \quad (3.5)$$

За рахунок фазочастотних спотворювань сигналу виду:

$$\arg \dot{A}(\omega) = \exp(jb \sin \omega c),$$

де b – крутість спотворень, програш у відношенні сигнал/шум складатиме

$$G=[J_0(b)]^2, \quad (3.6)$$

де $J_0(b)$ – функція Бесселя першого роду нульового порядку.

Якщо фазова спотворювальна функція лінійна протягом тривалості сигналу, то

$$G=\frac{\Psi_m^2}{2(1-\cos\Psi_m)}, \quad (3.7)$$

де Ψ_m – набіг фази за тривалість імпульсу.

Припустимо розстроювання F за частотою пропорційне

$$F=\frac{\Psi_m}{2\pi\tau}.$$

Якщо

$$\psi(t)=\begin{cases} -\Psi_m/2, & (0 < t < \tau) \\ \Psi_m/2, & (\tau/2 < t < \tau) \end{cases},$$

то

$$G=2/(1+\cos\Psi_m). \quad (3.8)$$

Якщо зміна фази вхідного сигналу є стаціонарним, нормальним процесом із нульовим середнім і дисперсією σ^2 , то

$$G=\exp\sigma_y^2. \quad (3.9)$$

При спотворенні комплексної обвідної сигналу в $1+\gamma(t)$ разів, де $\gamma(t)$ – стаціонарний випадковий процес із нульовим середнім і функцією кореляції $R(t)=\sigma_y^2\rho(t)$, то програш за потужністю або у відношенні сигнал/шум, при незмінному середньому значенні вихідної напруги, визначається таким чином:

$$G=1+\sigma_y^2. \quad (3.10)$$

Розстроювання F за частотою лінійно-частотно модульованого сигналу відносно фільтра призводить до зсуву відгуку на час $\Delta t\xi$

$$\Delta t = \frac{F}{\Delta f_{fm}} \tau_1 \quad (3.11)$$

і ослаблення його максимуму

$$\sqrt{G} = \left(1 - \frac{|F|}{\Delta f_{fm}} \right) \cdot \quad (3.12)$$

Розширення тривалості вихідного сигналу при невеликих розстроюваннях визначається як

$$\tau_2 \cong \frac{1}{\Delta f_{fm} - |F|} \cdot \quad (3.13)$$

Застосування рециркулятора [13] замість оптимального фільтра для послідовності прямокутних імпульсів дозволяє в найкращому випадку при оптимальному коефіцієнті зворотного зв'язку $m=1-0,27/N$ одержати програш у відношенні сигнал/шум

$$G=0,8N, \quad (3.14)$$

де N – число імпульсів.

При квазікогерентному, майже когерентному, прийманні фазоманіпульованих (ФМ) сигналів енергетичний програш при малих систематичних зрушеннях оцінюється таким чином [13]:

$$G = \frac{1}{\cos^2 \varphi} \cdot \quad (3.15)$$

Згідно з [13], втрати енергії імпульсу дзвонової форми при проходженні ним підсилювача із дзвоною амплітудно-частотною характеристикою дорівнюють

$$G = \frac{4a}{1 + 4a^2}, \quad (3.16)$$

де $a = \Pi\tau$ – добуток смуги пропускання підсилювача на тривалість імпульсу $a_{opt} = 0,5$.

При прийманні тим же радіопідсилювачем прямокутних радіоімпульсів, або навпаки,

$$G = \frac{[\Phi(\sqrt{\pi/2})]^2}{a} \quad \text{при } a_{opt} = 0,8. \quad (3.17)$$

Неузгоджена обробка прямокутних імпульсів радіопідсилювачем із прямокутною формою амплітудно-частотної характеристики дає втрати енергії

$$G = \frac{1}{a} \left[\text{Si}\left(\frac{\pi}{2} a\right) \right]^2, \quad (3.18)$$

де $\text{Si}(X)$ – інтегральний синус.

Наявність складних випадкових амплітудних і фазових спотворень із спектром флуктуацій, близьким до рівномірного, впливає на роздільну здатність за часом T_r і за частотою F_φ . У роботі [13] даються відповідні співвідношення

$$T_r / T_{rmin} = 1 / [1 - 0,5m_\varphi^2]^2; \quad (3.19)$$

$$F_\varphi / F_{\varphi min} = 1 / [1 - 0,5m_a^2]^2, \quad (3.20)$$

де m_φ – максимум фазочастотних спотворень, $m_\varphi = \max \varphi(2\pi f)$,
 m_a – максимум амплітудно-частотних спотворень,
 $m_a = \max a(2\pi f)$.

Оцінимо вплив точності виготовлення високочастотних ліній затримки з багатьма виводами, призначених для фільтраційної обробки фазоманіпульованих широкосмугових шумоподібних сигналів (ФМ ШШС), на вихідний ефект U_p

$$U_{вих} = U_0(t) \sum \cos(\omega t + \delta\varphi_i + a_i\pi), \quad (3.21)$$

де $U_0(t)$ – амплітуда вхідного ФМ ШШС;

a_i – закон фазової маніпуляції;

$\delta\varphi_i$ – випадкова апаратурна нормальна похибка із дисперсією кожного відводу $\sigma_{узг}^2$.

Середнє значення амплітуди (при усередненні за фазою) $U_p^{\varphi_i}$ головного піка вихідного ефекту визначається як

$$U_p^{\varphi_i} = U \sqrt{(\sum \cos \varphi_i)^2 + (\sum \sin \varphi_i)^2} \cong NU_0 (1 - \frac{1}{2N} \sum \varphi_i^2)^{\varphi_i}. \quad (3.22)$$

Через нормальність щільності розподілу ймовірності φ_i дисперсія фази

$$\overline{\varphi_i^2} = \sigma_{ст}^2.$$

Якщо затримки відрізків лінії незалежні випадкові величини з однаковими дисперсіями, то

$$\frac{\overline{U_p^{\varphi_i}}}{NU_0} \cong 1 - \frac{1}{2N} \sum_{i=1}^N \sigma_i^2 = 1 - 0,5\sigma^2. \quad (3.23)$$

Отже,

$$G = 1/(1 - \sigma^2). \quad (3.24)$$

Якщо затримка є вінерівським процесом першого порядку, коли затримка чергового відліку залежить лише від попередньої

(марковська властивість), то дисперсія установлення затримки залежить від номера відводу i

$$\sigma_i^2 = i\sigma_{cm}^2. \quad (3.25)$$

У цьому випадку з урахуванням (3.23)

$$\overline{G^{\varphi}} = 1 - N \frac{\sigma_{cm}^2}{4}. \quad (3.26)$$

З формул (3.23), (3.25) видно, що вимоги до точності установлення відводів у високочастотній лінії затримки досить високі, тому що дисперсія затримки

$$\sigma_{\tau}^2 = \sigma^2 T_0^2 / (2\pi)^2, \quad (3.27)$$

де $T_0 = \frac{2\pi}{\omega}$.

Визначимо вплив на вихідний ефект точності установлення відводів ліній затримки у випадку обробки бінарного відеосигналу, маніпульованого за законом псевдовипадкової послідовності. На виході оптимального фільтра, погодженого із сигналом, рівень головної пелюстки вихідного ефекту після обробки в лінії затримки визначається зі співвідношення

$$U_p(t) = U_0 \int \sum_{\tau e i=1}^{\tau e N} I_{\tau e}(t - \delta t) I_{\tau e}(t_1 - t) dt, \quad (3.28)$$

де $I_{\tau e}(t_1)$ – зрізаюча прямокутна функція тривалістю τ_e .

Оскільки кожний із суми інтегралів дорівнює $1 - \frac{|\tau_1|}{\tau_e}$, то максимум ефекту при деякому t_1 визначається як

$$\max U_p = U_0 \sum \left(1 - \frac{|\tau_i|}{\tau_e} \right) = U_0 N \left(1 - \frac{1}{N\tau_e} \sum_{i=1}^N |\tau_i| \right). \quad (3.29)$$

У випадку нормального розподілу щільності ймовірності затримки втрати G вихідної енергії сигналу визначаються зі співвідношення

$$G=1/(1-\sqrt{2/\pi} \sigma_{\tau} / \tau_e). \quad (3.30)$$

Зменшення відношення сигнал/шум в інтегрувальному пристрої характеризується коефіцієнтом втрат [45]. При підсумовуванні в інтеграторі N сигналів з амплітудою U_0 вихідний ефект буде дорівнювати

$$U_p=U_0(1-\exp[-(N-1)\frac{\sigma_{\tau}}{\tau_u}]/(1-\exp(-\frac{\sigma_{\tau}}{\tau_u}))), \quad (3.31)$$

де τ_u – стала часу інтегратора.

Якщо інтегратор виконаний на RC-ланцюгах, то втрати інтегрування можна задати таким чином [13]:

$$U_p=U_0T^2/(2(k+1)R^2C^2). \quad (3.32)$$

Беручи до уваги непогодженість поляризації φ приймальної і передавальної антен з лінійною поляризацією, втрати потужності сигналу складуть

$$G=\cos^2 \varphi . \quad (3.33)$$

Втрати у фідері згідно з [13] визначаються із співвідношення

$$G^{-1}=1+(k_b+1/k_b)\beta . \quad (3.34)$$

У загальному випадку відповідно до формули (3.1), а також з [13] необхідне відношення сигнал/шум зменшується в G разів

$$G=\prod_{i=1}^n G_i . \quad (3.35)$$

Відомі також методики врахування зменшення відношення сигнал/шум при візуальному спостереженні оцінок цілі на екрані електронно-променевого індикатора радіолокаційної станції й впливу зміни температури на зміну частоти генераторів

$$\Delta\omega_i = TKЧ(\omega)\Delta t^0, \quad (3.36)$$

де $TKЧ$ – температурний коефіцієнт частоти;

ω – номінальна колова частота;

Δt^0 – відхилення температури від вихідної.

Розглянутий далеко не повний перелік ефектів, властивих фільтраційній і кореляційній обробці сигналів, все-таки дає уявлення про моделі реальної системи й процеси, що відбуваються в них.

Аналогічно визначаються результати досліджень ефектів, що супроводжують будь-яку обробку сигналів. Головний результат тут у тому, що у випадку малості збурювань, неідеальностей і розстроювань у системі вихідний ефект приймача факторизується, тобто може бути поданий у вигляді добутку втрат.

3.2. Вплив обмеженості смуги пропускання високочастотного тракту на кореляційну обробку фазоманіпульованого за законом псевдовипадкової послідовності сигналу

Радіотехнічна система може використовувати фазоманіпульований сигнал. Розглянемо, як на нього впливає кореляційна система обробки, щоб використовувати результат такого системного аналізу в показнику якості РТС.

Нехай сигнал буде маніпульований по фазі псевдовипадковою послідовністю (ПВП). Цей сигнал, проходячи коливальну систему передавача й антени, вільний простір, антенно-фідерний тракт приймача, підсилювач радіочастоти й каскади радіоприймача, які розташовані до помножувача прийнятого сигналу з опорним, зазнає змін, що обумовлені

частотними характеристиками перерахованих функціональних елементів. Якщо вільний простір для простоти вважати недиспергуючим середовищем з рівномірною амплітудно-частотною й лінійною фазочастотною характеристиками, то для всього тракту передачі – приймання від передавача до помножувача необхідно знайти передавальну функцію, яка досить просто і правильно описує реальну частотну характеристику.

Нехай виконуються умови Ван-дер-Поля [17], тобто умови повільно мінливих амплітуд. Це реально, оскільки тривалість елемента τ , псевдовипадкової фазоманіпульованої послідовності (ФМ ПВП) містить не менше десятка періодів несучої частоти f_0 , а смугу пропускання всього високочастотного тракту до помножувача не можна вважати необмеженою.

Тоді згідно із [17] для ФМ ПВП можна визначити комплексну обвідну криву $H(t)$ перехідного процесу при інверсії від 0 до π фази θ

$$\dot{H}(t) = \sqrt{\cos^2 \theta + [1 - 2B(t)]^2 \sin^2 \theta} \exp[-j[1 - 2B(t)]tg\theta], \quad (3.37)$$

де $B(t)$ – речовинна складова обвідної кривої перехідної характеристики при вмиканні синусоїдального сигналу. Звичайно здійснюється маніпуляція $0-\pi$.

Тоді

$$H(t) = |1 - 2B(t)|, \quad (3.38)$$

а фаза при перехідному процесі змінюється стрибком від 0 до π при $H(t)=0$, де t_1 знаходиться з рівняння $B(t_1)=1/2$.

Напишемо частотну характеристику тракту до помножувача таким чином:

$$K(\omega) = \frac{1}{1 + j \frac{2(\omega - \omega_0)}{\Pi_{\text{вч}}}}, \quad (3.39)$$

де $\Pi_{\text{вч}}$ – смуга пропускання тракту до помножувача, яка відраховується на рівні $0,5K(0)$;

ω – поточна, а ω_0 – власна колова частота.

Тоді напруга на виході (при синусоїдальному одиничному стрибку) із частотою ω на вході визначається як

$$U(t) = K(\Delta\omega)[\cos(\omega t + \theta + \varphi_0) - e^{-\Pi_B t} \cos(\omega_0 t + \theta + \varphi_0)], \quad (3.40)$$

де φ – початкова фаза, а

$$\varphi_0 = -\operatorname{arctg} \frac{2(\omega - \omega_0)}{\Pi_B}.$$

При відсутності розстроювання $\omega = \omega_0$,

$$B_1(t) = 1 - e^{-\Pi_B t} \text{ – при вмиканні} \quad (3.41)$$

і $B_2(t) = e^{-\Pi_B t}$ – при вимиканні.

Тоді

$$H(t) = |1 - 2B_2(t)| = |1 - 2e^{-\Pi_B t}|. \quad (3.42)$$

Звідси

$$t_1 = \frac{\lg 2}{\Pi_B}. \quad (3.43)$$

Як впливає з (3.43) і рис. 3.1, обмеженість смуги пропускання призводить до розтягування перехідного процесу радіосигналу і, в кінцевому рахунку, до втрат енергії сигналу в основному за рахунок високочастотної частини спектра при відносно малих $\Pi_{вч} \tau_e = \Pi_{вч} / \Delta f_c$.

Крім того, система синхронізації за затримкою затримує опорну практично неспотворену послідовність щодо сигналу на час t_1 (3.43), якщо енергія перехідного процесу в послідовності буде більше енергії вихідного шуму при достатній реальній чутливості часового дискримінатора.

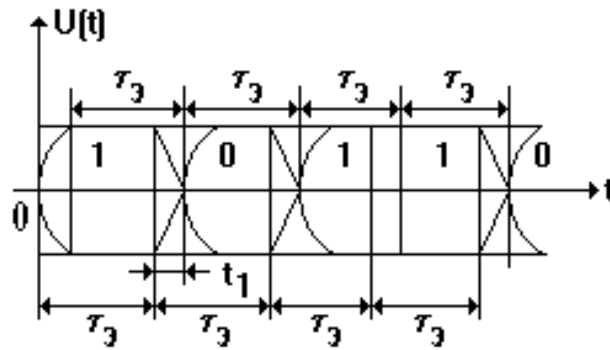


Рис.3.1. Вплив смуги пропускання на форму ФМ сигналу

Використовуючи вираз (3.42), можна одержати залежність енергії втрат E_B від смуги тракту до помножувача.

Енергію втрат E_B сигналу при одній інверсії фази можна визначити таким чином:

$$E_B = \int_0^{\tau_e} \cos^2 \Phi dt - [\int_0^{\tau_e} (1 - 2e^{-\Pi_B t})^2 \cos^2 \Phi dt + \int_0^{\tau_e} e^{-2\Pi_B t} \cos^2 \Phi dt], \quad (3.44)$$

де $\Phi = \omega t + \theta$.

Оскільки кількість серій для М-последовності – 2^{n-1} [29], кількість інверсій фази дорівнює $2^{n-1}-1$ і енергія втрат за період T визначається як

$$E_{BT} = \frac{N}{2} E_B = \frac{E_0}{2a} |1 - 2e^{-a} + e^{-2a}|, \quad (3.45)$$

де $a = \Pi_{вч} \tau_e = \Pi_{вч} / \Delta f_c$;

Δf_c – ширина спектра сигналу.

Аналогічні затягування фронтів сигналу існують і до погодженого фільтра при фільтраційному прийманні. У будь-якому випадку при так званій узгодженості ширини спектра

сигналу й смуги пропускання тракту ($a=1$) втрати енергії сигналу досягнуть

$$E_{ПТ}/E_0=0,16.$$

Потужність шумів на виході корелятора істотно не зміниться, оскільки в будь-якому випадку основне посилення приймача здійснюється в смузі $П_{ПРМ}=1/T=П_{ВЧ}/a$.

Тому відношення сигнал/шум на виході приймача погіршиться відповідно до (3.45).

Для задачі синтезу системи за критерієм мінімуму похибки вимірювань при обмеженнях на вартість становить інтерес область задання параметра $a \in [1, \infty)$, причому при $a \rightarrow \infty$ зменшується селективність високочастотного тракту, збільшується його вартість і зменшується доцільність його застосування.

З іншого боку, використання параметра $a \leq 1$ веде до різкого збільшення втрат енергії $E_{ПТ}$.

Для оцінки впливу обмеженості смуги пропускання високочастотного тракту на вихідний ефект приймача при кореляційній обробці сигналу фазочастотні спотворення сигналу в тракці будемо припускати малими. Фазові діаграми спрямованості антен також роблять свій внесок у набіг фази сигналу, однак час кореляції тілесних флуктуацій значно більше періоду сигналу T . Тому можна вважати, що набіг фази сигналу незмінний протягом одного періоду.

У радіоприймальному пристрої радіосигнал спотворюється за рахунок обмеженої смуги пропускання високочастотного тракту, а опорний відеосигнал формується практично без спотворень. Тому вихідний ефект являє собою, строго кажучи, спотворену кореляційну функцію сигналу в кінці періоду. Саме вихідний ефект і потрібно вивчити. Тоді в припущенні ідеальності інтегратора при розрахунку вихідного ефекту Z при кореляційному прийманні й при розстроюванні синхронізації $|\delta t| \leq \tau_e$ буде спостерігатися чотири можливих випадки (рис. 3.2).

$$\begin{aligned} 1) -\tau_e + t_1 \leq \delta t \leq 0; \quad 2) 0 \leq \delta t \leq t_1; \quad 3) t_1 \leq \delta t \leq \tau_e; \\ 4) \tau_e \leq \delta t \leq \tau_e + t_1. \end{aligned}$$

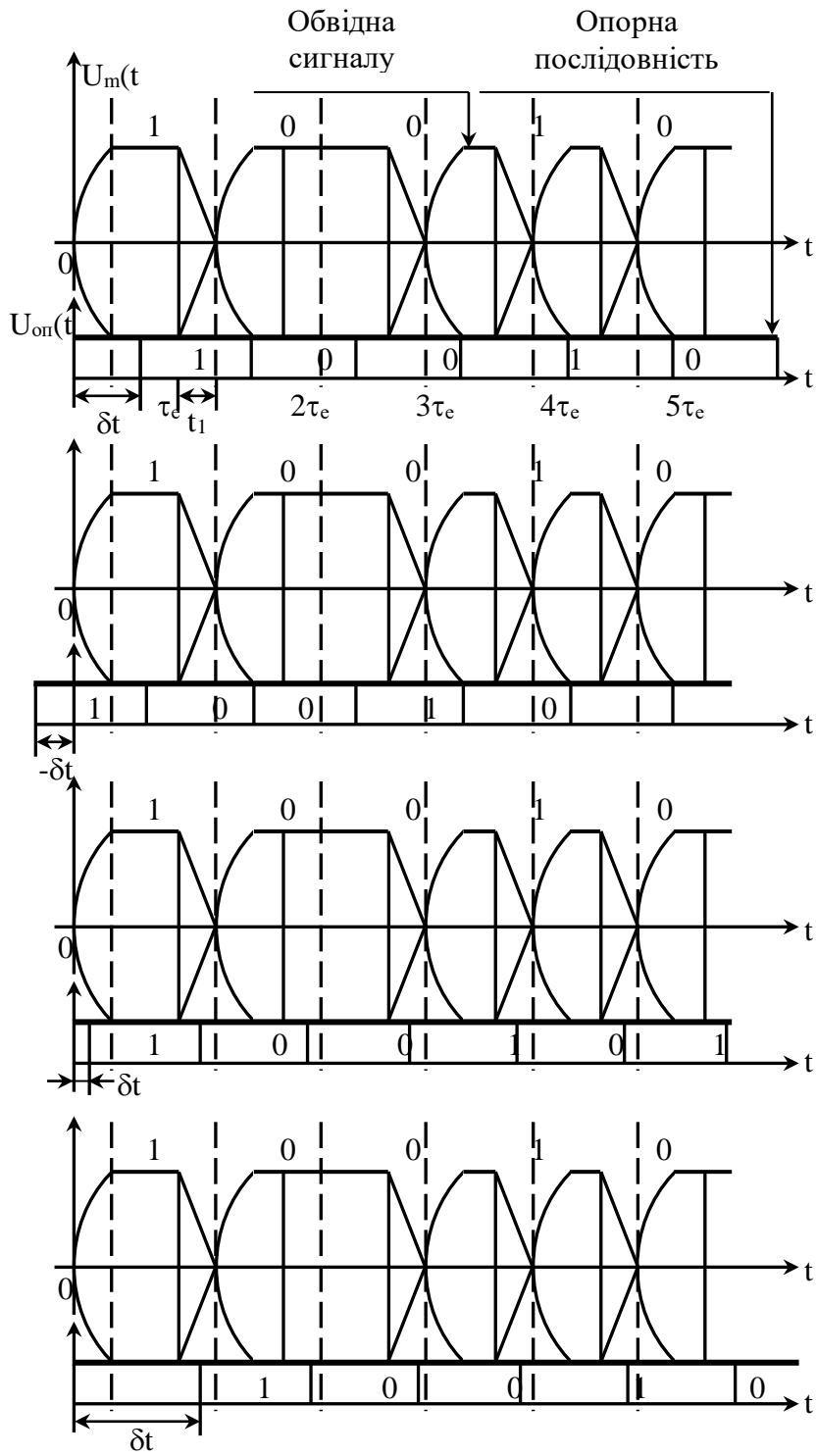


Рис. 3.2. Діапазони зсуву опорного сигналу

Розрахуємо вихідний ефект у першому випадку.

$$\begin{aligned}
 Z_1 = & \sum_{k=1}^N \int_{(k-1)\tau_e}^{k\tau_e - \delta t} [1 - 2e^{-\Pi_B(t - (k-1)\tau_e)}]^{a_m} \cos \Psi dt + \\
 & + \sum_{k=1}^N e^{ja_m \pi} \int_{(k-1)\tau_e}^{(k-1)\tau_e + t_1} (2e^{-\Pi_B(t - k\tau_e)} - 1)^{a_m + 1} \cos \Psi dt + \\
 & + \sum_{k=1}^N e^{ja_m \pi} \int_{k\tau_e - \delta t}^{k\tau_e} [1 - 2e^{-\Pi_B(t - k\tau_e)}]^{a_m} \cos \Psi dt, \quad (3.46)
 \end{aligned}$$

де $\Psi = (\omega - \omega_0)(t - \delta t) + a_m \pi + \Psi_c$;

ω_0 – розстроювання частоти;

N – число елементів у псевдовипадковій послідовності або кількість елементів послідовного складеного сигналу;

$\Pi_{BЧ}$ – смуга пропускання високочастотного тракту до помножувача;

$t_1 = \ln 2 / \Pi_{BЧ}$;

$a_m = a_k \cap a_{k-1}$ – послідовність, отримана перемноженням послідовності a_k на a_{k-1} ;

a_{m+1} – аналогічна послідовність. Елемент a_m набуває значення 0 або 1 залежно від k .

Оцінимо вихідний ефект при відсутності розстроювань. Тоді перший доданок I_1 (3.46) має вигляд

$$\begin{aligned}
 J_1 = & \sum_{k=1}^N \int_{t_1}^{\tau_e} (1 - 2e^{-\Pi_B t})^{a_m} dt = \sum_{k=1}^N \int_{t_1}^{\tau_e - \delta t} \left\{ \frac{1 - e^{-\Pi_B t} / \text{при } p_1(a_m = 1)}{1 / \text{при } p_0(a_m = 0)} \right\} dt = \\
 = & N \tau_e \left\{ \frac{1 - \frac{|\delta t|}{\tau_e} - \frac{t_1}{\tau_e} + \frac{2}{a} \left(e^{-a(1 - \frac{|\delta t|}{\tau_e})} - \frac{1}{2} \right) \text{при } p_1(a_m = 1)}{1 - \frac{|\delta t|}{\tau_e} - \frac{t_1}{\tau_e} / \text{при } p_0(a_m = 0)} \right\}.
 \end{aligned}$$

Для псевдовипадкової послідовності події $a_m=1$ і $a_m=0$ зустрічаються із частотами p_1 і p_0 . Звичайно для ПВП $p_0=p_1=1/2$. Тоді математичне сподівання I_1 при усередненні по a_m буде дорівнювати

$$J_1^{a_m} = N\tau_e \left[1 + \frac{\delta t}{\tau_e} - \frac{t_1}{\tau_e} - \frac{1}{a} \left(e^{-a\left(1+\frac{\delta t}{\tau_e}\right)} - \frac{1}{2} \right) \right]. \quad (3.47)$$

Таким же способом можна знайти вираз для математичного сподівання другого й третього доданків (3.46),

$$\bar{I}_2^{a_m} = N\tau_e \frac{\rho_1}{2a}; \quad (3.48)$$

$$\bar{I}_3^{a_m} = N\tau_e \left[-\frac{\delta t}{\tau_e} + \frac{e^{-a} \left(e^{-a\frac{\delta t}{\tau_e}} - 1 \right)}{a} \right] \rho_1, \quad (3.49)$$

де $\rho_1 = \frac{1}{N} \sum_{k=1}^N e^{j(a_k \oplus a_{k-1})\pi}$ – знакова нормована кореляційна функція при одиничному розстроюванні на один елемент τ_e .

Вихідний ефект для випадку $-\tau_e + t_1 \leq \delta t \leq 0$ визначається як сума (3.48), (3.49), (3.50):

$$\begin{aligned} \bar{Z}_1^{a_m} = N\tau_e \left[1 + \frac{\delta t}{\tau_e} B_1 - \frac{\ln 2}{a} + \frac{\rho_1}{2a} - \frac{\rho_1}{a} e^{-a} \left(1 - e^{-a\frac{\delta t}{\tau_e}} \right) \right] - \\ - \frac{1}{2a} \left(1 - 2e^{-a\left(1-\frac{\delta t}{\tau_e}\right)} \right), \end{aligned} \quad (3.50)$$

де $B_1 = 1 - \rho_1$.

Для випадків $0 \leq \delta t \leq -t$, $t \leq \delta t \leq \tau_e$, $\tau_e \leq \delta t \leq \tau_e + t$ вихідний ефект визначається аналогічно:

$$\bar{Z}_2^{a_m} = N\tau_e \left[1 + \frac{t_1}{\tau_e} + \frac{1}{a} \left(\frac{1}{2} + e^{-a} \right) - \frac{1}{a} e^{-a \frac{\delta t}{\tau_e}} - \frac{\rho_1}{a} \left(\frac{1}{2} - e^{-a \frac{\delta t}{\tau_e}} \right) \right]; \quad (3.51)$$

$$\bar{Z}_3^{a_m} = N\tau_e \left\{ \begin{array}{l} 1 - \frac{\delta t}{\tau_e} B_1 - \frac{1}{a} \left(e^{a \frac{\delta t}{\tau_e}} - e^{-a} \right) + \frac{1}{2a} + \\ + \rho \left[-\frac{t_1}{\tau_e} + \frac{1}{a} \left(e^{-a \frac{\delta t}{\tau_e}} - \frac{1}{2} \right) \right] \end{array} \right\}; \quad (3.52)$$

$$\bar{Z}_4^{a_m} = N\tau_0 \left[\frac{1}{a} e^{-a \left(\frac{\delta t}{\tau_e} - 1 \right)} - \frac{1}{2a} + \left(1 - \frac{t_1}{\tau_e} \right) \rho + \frac{\rho_1}{a} \left(\frac{1}{2} + e^{-a} - e^{-a \frac{\delta t}{\tau_e} - 1} \right) \right]. \quad (3.53)$$

Таким чином, нормовану, усереднену по a_m квазіавтокореляційну функцію $Z_{\text{вух}}/N\tau_e$ можна подати у вигляді

$$\frac{Z_{\text{вух}}}{N\tau_e} = \begin{cases} 1 + \frac{\delta t}{\tau_e} B_1 - \frac{\ln 2}{a} - \frac{B_1}{2a} \left(1 - 2e^{-a \left(1 - \frac{\delta t}{\tau_e} \right)} \right) + \rho \frac{e^{-a}}{a}; & -1 + \frac{\ln 2}{a} \leq \frac{\delta t}{\tau_e} \leq 0; \\ 1 - \frac{\ln 2}{a} + \frac{B_1}{2a} \left(1 - 2e^{-a \frac{\delta t}{\tau_e}} \right) + \frac{e^{-a}}{a} - \frac{\ln 2}{a} \rho_1; & 0 \leq \frac{\delta t}{\tau_e} \leq \frac{\ln 2}{a}; \\ 1 - \frac{\delta t}{\tau_e} B_1 + \frac{B_1}{2a} \left(1 - 2e^{-a \frac{\delta t}{\tau_e}} \right) + \frac{e^{-a}}{a} - \frac{\ln 2}{a} \rho_1; & \frac{\ln 2}{a} \leq \frac{\delta t}{\tau_e} \leq 1; \\ \frac{B_1}{2a} \left(2e^{-a \left(\frac{\delta t}{\tau_e} - 1 \right)} - 1 \right) + \rho_1 \left(1 - \frac{\ln 2}{a} + \frac{e^{-a}}{a} \right); & 1 \leq \frac{\delta t}{\tau_e} \leq 1 + \frac{\ln 2}{a}. \end{cases} \quad (3.54)$$

Функція (3.54) легко програмується й обчислюється для різних a і ρ . На рис. 3.3 вона подана для різних відношень α смуги пропускання високочастотного тракту до ширини спектра ШШС.

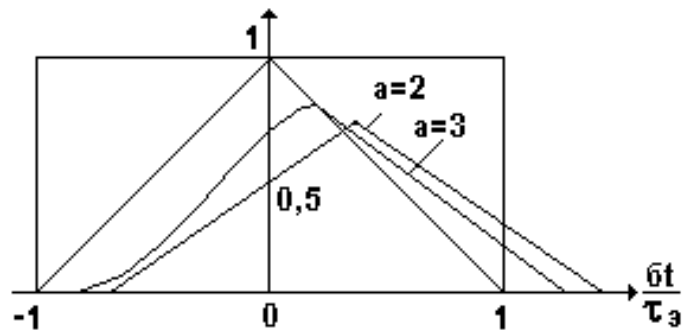


Рис. 3.3. Відлік корелятора на вихідний сигнал

У прийнятих тут позначеннях $\delta t < 0$ означає, що прийнятий сигнал відстає від опорного сигналу (рис. 3.2). Маючи це на увазі, з рис. 3.3 можна помітити такі ефекти, пов'язані зі зміною (наприклад збільшенням) величини α :

а) при збільшенні α втрати енергії зменшуються за рахунок меншого придушення частотною характеристикою тракту (наприклад ПРЧ) високочастотних складових, тоді нормована квазіавтокореляційна функція прагне до 1;

б) при $\alpha \rightarrow \infty$ зсув $t_1 = \ln 2 / \alpha$ піка квазіавтокореляційної функції зменшується до нуля;

в) при всіх α ширина вихідного ефекту «по нулях» не змінюється;

г) при всіх α квазіавтокореляційна функція близька до автокореляційної за формою, а при $\alpha \geq 8$ – і за величиною;

д) уже при $\alpha \geq 2$ квазіавтокореляційна функція має лінійний характер у бік випередження прийнятого сигналу;

е) навіть при великих α і при $\delta t < 0$ існує незначна кривизна вихідного ефекту, зумовлена перехідним процесом наростання напруги після інверсії фази на π ;

ж) величина квазіавтокореляційної функції на відповідному рівні при кожному α не більше ширини функції автокореляції неспотвореного сигналу.

На підставі (3.54) і наведених зауважень можна зробити такі висновки:

1) при оптимізації системи за параметром α варто враховувати втрати енергії сигналу за рахунок обмеженості смуги пропускання радіотракту до помножувача при $\alpha < \delta$;

2) зсув квазіавтокореляційної функції на t_1 у бік запізнювання істотно залежить від α , однак при оптимізації його можна врахувати як систематичну похибку системи синхронізації;

3) із властивостей «д» і «е» випливає, що при $\alpha < \delta$ відбудеться додатковий зсув опорного сигналу у бік випередження через кривизну на великих рівнях автокореляційної функції, що бере участь у створенні дискримінаційної характеристики каналу синхронізації. Цей зсув істотний тільки при малому відношенні сигнал/шум;

4) із властивості «ж» випливає, що якщо при якому-небудь α компенсувати втрати енергії за рахунок збільшення потужності сигналу й врахувати зсув, то якість оцінки затримки й умови розрішення за дальністю погіршаться незначно;

5) із властивостей «а», «г» випливає, що при врахуванні зсуву квазіавтокореляційну функцію сигналу залежно від параметрів α і δt можна подати у вигляді

$$Z_{\text{вих}} = N\tau_e [1 - \varphi(a, 0)] \left(1 - \left| \frac{\delta t}{\tau_e} - \frac{\ln 2}{a} \right| \right), \quad (3.55)$$

$$\text{де } 1 - I(a, 0) = 1 - \frac{\ln 2}{a} - \frac{B}{2a} + \frac{e^{-a}}{a}.$$

Тому в оптимізації системи можуть брати участь факторизовані функції від параметрів α і δt , якщо припустити, що систематична похибка синхронізації $\ln 2/\alpha$ буде усунута при настроюванні системи.

3.3. Вплив похибки кутового супроводу антени на потужність сигналу й оптимальне співвідношення між її спрямованістю і якістю кутової стабілізації променя

Питання про вплив кутової нестабільності антени і про оптимальне співвідношення її спрямованості до величини її тілесних флуктуацій має особливе значення при створенні систем із протяжними радіолініями, які використовують сигнали великої тривалості, наприклад широкосмугові шумоподібні сигнали.

Необхідність забезпечення заданого енергетичного потенціалу радіолінії призвела до збільшення коефіцієнта спрямованої дії (КСД) застосовуваних антен, а отже, до більш жорстких вимог до якості систем кутової стабілізації антен.

Спочатку розглянемо для простоти однобічну радіолінію, у якій одна з антен ідеально стабілізована, а інша – тільки за одним кутом θ_y . За іншим кутом θ_x антена флюктує. Нехай дисперсія флуктуацій – D_x . У режимі автосупроводу за цим кутом звичайно дисперсія флуктуацій менше ширини діаграми спрямованості антени по половині потужності. Для цього випадку в межах максимуму хорошою апроксимацією діаграми спрямованості $G(\theta_x)$ за полем для реальних антен може бути закон

$$G(\theta_x) = \exp\left(-\frac{\theta_x^2}{2\Delta\theta_x^2}\right). \quad (3.56)$$

У випадку автосупроводу за кутом через багатопричинність збурювань і велику сталу часу системи стабілізації за кутом відхилення $\theta_x(t)$ максимуму діаграми спрямованості від напрямку на ціль є випадковий центрований стаціонарний процес із нормально розподіленим відхиленням

$$p(\theta_x) = \frac{1}{\sqrt{2\pi D_x}} \exp\left(-\frac{\theta_x^2}{2D_x}\right). \quad (3.57)$$

Знайдемо щільність розподілу ймовірності рівня сигналу за полем $p(G)$

$$p(G) = 2p(\theta_x) \left| \frac{\partial \theta_x}{\partial G} \right|. \quad (3.58)$$

Оскільки

$$\theta_x = \sqrt{2\Delta\theta_x \ln \frac{1}{G}}, \quad (3.59)$$

то

$$p(G) = \sqrt{\alpha_x / \pi} \frac{G^{\alpha-1}}{\sqrt{\ln \frac{1}{G}}}, \quad (3.60)$$

де $\alpha_x = \frac{\Delta\theta_x^2}{D_x}$.

Сімейство $\rho(\alpha_x, G)$ подане на рис. 3.4. Звідси математичне сподівання рівня сигналу за полем дорівнює

$$M[G] = \sqrt{\frac{\alpha_x}{\alpha_x + 1}}, \quad (3.61)$$

а n -й початковий момент

$$M[G^n] = \sqrt{\frac{\alpha_x}{\alpha_x + n}}. \quad (3.62)$$

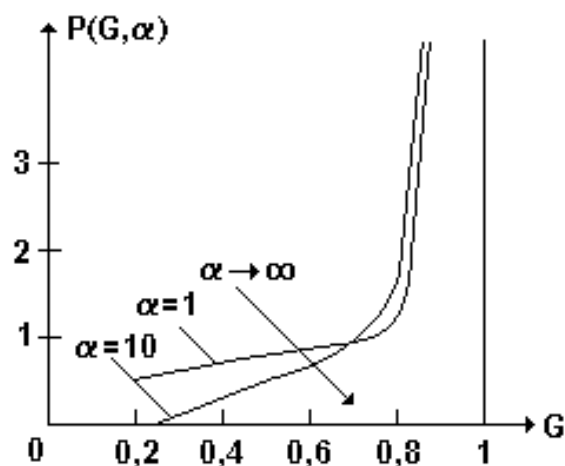


Рис. 3.4. Щільність розподілу ймовірності рівня сигналу у стабілізованій антені

Залежності $M[G]$ і $M[G^2]$ подані на рис. 3.5.

З виразів (3.61), (3.62), а також з рис. 3.5 видно, що врахування дестабілізуючих факторів доцільне уже при $\alpha \leq 8$, а при $\alpha_x \approx 1$ воно просто необхідне. Оскільки при $\alpha_x = 8$ середні енергетичні втрати становлять 10 %, а при $\alpha_x = 1$ – 50 %. Очевидно, що зі зменшенням α_x зростає роль мультиплікативної завади за рахунок флуктуацій діаграми спрямованості антени, які мають розподіл (3.60), обумовлений формами розподілу флуктуацій і діаграми спрямованості антени.

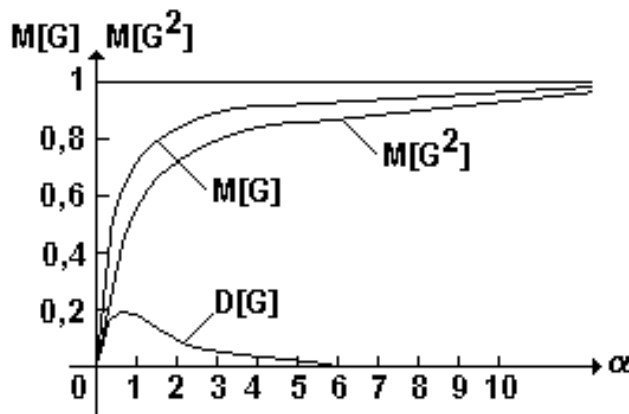


Рис. 3.5. Математичне сподівання рівня й потужності сигналу та його дисперсія

На великих відстанях, коли при оптимальній формі сигналу й максимальній потужності передавача потрібно, крім того, збільшення КСД антени, тобто зменшення $\Delta\theta$, доводиться висувати більш жорсткі вимоги до системи кутової стабілізації діаграми спрямованості антени. Точність виконання поворотної механічної частини системи й допуски на флуктуацію діаграми спрямованості антени також визначаються виразами (3.61), (3.62). При збільшенні КСД, ускладненні й обваженні поворотної частини антени час кореляції рівня сигналу також зростає. Формули (3.61), (3.62) описують середні енергетичні втрати за проміжок часу, що перевищує час кореляції.

Максимальна дальність радіолінії R_{max} з урахуванням кутових флуктуацій при заданому максимальному енергетичному потенціалі й параметрах радіолінії таким чином залежить від параметра α .

Для зазначеної моделі однобічної радіолінії

$$R_{max}^2 = \frac{P_{прд} G_{0прм} G_{0прд} \lambda^2}{P_{спрм} 16\pi^2} M[G_{прд}^2], \quad (3.63)$$

де $P_{прд}$ – потужність передавача;

$P_{спрм}$ – реальна чутливість приймача;

$G_{0прд}$ – КСД передавальної антени;

$G_{0прм}$ – КСД приймальної антени;

λ – довжина хвилі.

Оскільки КСД антени ($G_{0прд}$) можна представити у вигляді

$$G_{0прд} = \frac{K_1}{\Delta\theta_{xпрд}^2 \Delta\theta_{yпрд}^2} \quad (3.64)$$

для гостроспрямованих антен, вираз (3.63) з урахуванням (3.64) можна записати таким чином (рис. 3.6) :

$$R_{max}^2 = \frac{K_2}{\sqrt{\Delta\theta_{xпрд}^2 (\Delta\theta_{xпрд}^2 + 2D_{xпрд})}}, \quad (3.65)$$

де K_1, K_2 – постійні;

$\Delta\theta_{xпрд}^2, \Delta\theta_{yпрд}^2$, як і раніше, ширина діаграми спрямованості передавальної антени за потужністю в стабілізованих площинах відповідно.

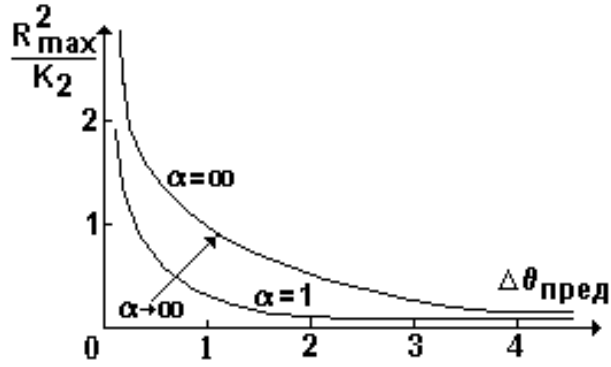


Рис. 3.6. Вплив діаграми спрямованості на відстань зв'язку

З (3.65) випливає, що залежність R_{max}^2 від $\Delta\theta_{x\text{прд}}$ сильніша, ніж залежність R_{max}^2 від $D_{x\text{прд}}$.

Дійсно,

$$\frac{\partial(R_{max}^2)}{\partial(\Delta\theta_{x\text{прд}}^2)} = -K_2 (\Delta\theta_{x\text{прд}}^4 + 2\Delta\theta_{x\text{прд}}^2 D_{x\text{прд}})^{-\frac{3}{2}} (\theta_{x\text{прд}}^2 + D_{x\text{прд}});$$

$$\frac{\partial(R_{max}^2)}{\partial(D_{x\text{прд}})} = -K_2 (\Delta\theta_{x\text{прд}}^4 + 2\Delta\theta_{x\text{прд}}^2 D_{x\text{прд}})^{-\frac{3}{2}} \theta_{x\text{прд}}^2.$$

Тому завжди

$$\left| \frac{\partial(R_{max}^2)}{\partial\Delta\theta_{x\text{прд}}^2} \right| > \left| \frac{\partial(R_{max}^2)}{\partial D_{x\text{прд}}} \right|.$$

Таким чином, для збільшення дальності дії командної радіолінії доцільно переважно зменшувати ширину ДС антени в порівнянні з похибкою стабілізації за кутом. Якщо переписати (3.65) у вигляді

$$R_{max}^2 = \frac{K_2}{\Delta\theta_{x\text{прд}}^2 \sqrt{1 + \frac{2}{\alpha}}}, \quad (3.66)$$

то стане ясно, що доцільність використання α_x більше 8 зменшується, тому що подальше зменшення D_x мало позначається на збільшенні R_{max}^2 . При $\alpha_{x0} \geq 8$ втрати енергетичного потенціалу $1-M[G^2]$ становлять не більше 11 %.

Усе викладене неважко поширити на реальні радіолінії, де кутові флуктуації ДС антен відбуваються незалежно за двома кутами на двох кінцях радіолінії. У цьому випадку середнє значення R_{max}^2 залежить від $\prod_{i=1}^4 M[G_i(\Delta\theta_i)]$ для однобічної радіолінії й середнє значення R_{max}^2 залежить від $\prod_{i=1}^8 M[G_i(\Delta\theta_i)]$ для радіоліній з ретрансльованою активною відповіддю.

Якщо антена використовується, крім того, для пеленгації цілі в амплітудному або моноімпульсному методі, то D_x і α_x вибираються з умови забезпечення необхідної точності вимірювання кута $\delta\theta_x^{-2}$, тобто з умов $D_x(\delta\theta_x)^2$ і $\Delta\theta_x(\delta\theta_x)^2$. Однак при цьому можливий випадок сумарності D_x і $\Delta\theta_x$, звідки впливає необхідність урахувувати втрати енергетичного потенціалу.

При формалізації задачі оптимізації радіотехнічних систем розглянутий ефект буде врахований у дисперсії D_x похибки кутового супроводу; $\Delta\theta_x$ буде врахована в КСД антени.

3.4. Вплив невизначеності в знанні частоти на якість кореляційної обробки сигналу, фазоманіпульованого за законом псевдовипадкової послідовності

Щоб уникнути ефектів, пов'язаних з реальним інтегруванням, розглянемо кореляційну обробку сигналу з ідеальним інтегратором.

Запишемо випромінюваний сигнал у вигляді

$$S(t) = S_0 \sum_{i=1}^N I_{\tau_e}(t - i\tau_e) \cos[a_i\pi + \omega_0 t + \varphi(t)]. \quad (3.67)$$

Опорний сигнал має аналогічний вигляд.
 Припустимо, що тактова частота незалежна від несучої.
 Прийнятий сигнал має вигляд

$$S(t - t_3) = S_0 \sum_{i=1}^N I_{\tau_{ei}} \left(t - \sum_{k=1}^i \tau_{ek} \right) \cos[a_i \pi + \omega(t - t_3) + \varphi(t - \tau_3)], \quad (3.68)$$

де τ_{ei} залежить від руху цілі, нестабільності частоти генератора тактових імпульсів (ГТІ) приймача щодо опорного ГТІ, t_3 – від дальності.

Механізм впливу ефекту Доплера полягає в тому, що відповідно до радіальної швидкості змінюється несуча частота й відповідно змінюється тривалість усіх елементів сигналу (рис. 3.7). Тому ті самі елементи зсуваються щодо опорних тим більше, чим ближче до кінця періоду.

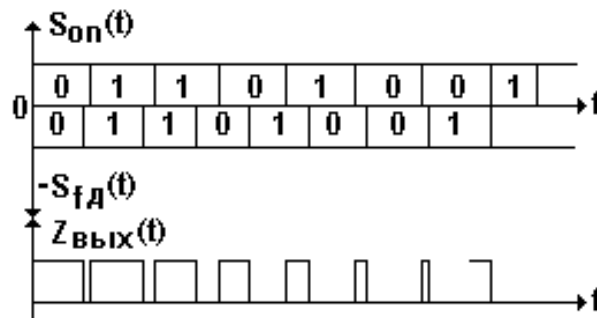


Рис. 3.7. Вплив частоти Доплера на якість приймання послідовного сигналу

З рис. 3.7 видно, що якщо об'єкти зближаються з постійною радіальною швидкістю, при якій тривалості всіх інтервалів послідовності коротшають, то коротшає й тривалість елементів у порівнянні з вихідною. Якщо тривалість усієї послідовності коротшає на один елемент, то, як видно з рис.3.7, могла б втратитись половина енергії сигналу. Але насправді втрачається приблизно чверть енергії сигналу, тому що приблизно половина енергії частини сигналу, що не збігається, все-таки синфазна.

У реальних системах тактова синхронізація й синхронізація по несучій частоті здійснюються окремими пристроями й тому вони незалежні.

Припустимо, що існує зв'язок з об'єктом, $\delta t_3 = 0$, і згортка (вихідний ефект) погіршується тільки за рахунок ефекту Доплера при тактовій синхронізації або за рахунок непередбаченої зміни частоти, наприклад вандера, або за рахунок динамічної похибки ФАПЧ. Канал передачі дискретної інформації має смугу, що не перевищує $\Pi_{\text{фачч}} \approx \frac{1}{T}$, де T – період послідовності.

$$[\omega_{\partial}(T) - \omega_{\partial}(0)] \leq \frac{2\pi}{T} = \Pi_{\text{фачч}} 2\pi. \quad (3.69)$$

А підсистема фазового автопідстроювання частоти (ФАПЧ) в значно точніша. Тому зміни частоти $\Delta\omega(t)$ досить малі. Особливістю ефекту, зображеного на рис. 3.7, є те, що з кожним тактом (елементом) синфазна частина сигналу втрачається на величину τ_e . А ті інтервали часу, де символи не збігаються, можуть бути синфазними чи ні, залежно від коду. Оскільки ймовірність збігу з елементом символу одного значення дорівнює $\frac{1}{2}$, то енергія сигналу за один період відповідно зменшується. Існують дві реальні схеми обробки приймання, або приймання фазоманіпульованого за законом послідовності сигналу. Перший спосіб – це використання фазового детектора, при якому на виході буде сигнал, маніпульований за амплітудою. Другий спосіб – інвертування елементів, зсунутих по фазі на π . При цьому спектр сигналу звужується в N разів, спектр вузькосмугової завади навпаки, розширюється в N разів. А оскільки фільтрація сигналу здійснюється у вузькій смузі, то завада послаблюється в N разів. Будемо тому мати на увазі другий спосіб.

Тоді енергія синфазної частини сигналу, якщо мати на увазі, що рис.3.7 являє собою тільки інтервали елементів радіосигналу, дорівнює

$$Z_{вих} \approx \sum_{i=1}^N \int_{(i-1)\tau_e}^{i\tau_e - a_i \tau_{ек}(2i-1)} dt = \sum_{i=1}^N (\tau_e - a_i \tau_{ек}(2i-1)) \quad (3.70)$$

при

$$\left| i\tau_e - \sum_{k=0}^i \tau_{ек} \right| \leq \tau_e, \quad (3.71)$$

де середнє за ймовірністю значення $M[a_k] = 1/2$. Тоді з виразу (3.70) отримаємо:

$$Z_{вих} \approx \sum_{i=1}^N (\tau_e - \frac{1}{2} \tau_{ек}(2i-1)), \quad (3.72)$$

$$\text{де } \frac{\tau_{ек}}{\tau_e} \approx \frac{f_{\partial}}{f_T} = \frac{f_{\partial} T}{N} = \frac{f_{\partial}}{N\Pi}. \quad (3.73)$$

Тоді (3.72) можна визначити як

$$Z_{вих} \approx N\tau_e \left(1 - \frac{\tau_{ек}}{2\tau_e} \sum_{i=1}^N (2i-1) \right) = N\tau_e \left(1 - \frac{\tau_{ек}}{2\tau_e} (N-2) \right). \quad (3.74)$$

З огляду на (3.73), також у вигляді

$$\Delta\omega_{\partial}\tau_e \leq \frac{2\pi}{N}, \quad (3.75)$$

(3.74) можна подати таким чином:

$$Z_{вих} \approx N\tau_e \left(1 - \frac{1}{2} \frac{\omega_{\partial}}{\omega_T} (N-2) \right). \quad (3.76)$$

У випадку довільного руху об'єктів передачі й приймання сигналу формулу (3.76) можна узагальнити.

За умови (3.75), коли $0 \leq i\tau_e \leq T$ при $\forall i \in (1, n)$, величину

$\frac{\omega_{\partial}}{\omega_T}$ можна записати як

$$\frac{\omega_{\partial}(t)}{\omega_T} = \frac{1}{C} \left(\dot{R}_0 + \frac{\ddot{R}(t)}{1!} + \frac{\ddot{R}t^2}{2!} + \dots \right). \quad (3.77)$$

Така форма запису корисна тоді, коли відомі закони руху об'єктів і коли частоту сигналу відстежує ФАПЧ із необхідним астатизмом за будь-якою похідною дальності. Тому під $R^{(k)}$ тоді можна розуміти або похідну від дальності, або відповідну динамічну похибку спостереження.

Оскільки можна дискретизувати $t = i\tau_e$, то

$$\frac{\omega_{\partial i}}{\omega_T} = \frac{1}{C} \sum_{k=1}^{\infty} R^{(k)} \frac{(i\tau_e)^{k-1}}{(k-1)!}, \quad (3.78)$$

Тоді (3.76) можна записати у вигляді

$$Z_{\text{вих}} = N\tau_e \left(1 - \frac{1}{2C} \sum_{k=1}^{\infty} R^{(k)} \frac{(i\tau_e)^{k-1} (N-2)}{(k-1)!} \right). \quad (3.79)$$

З (3.79) випливає, що кількість періодів несучої частоти $\omega_0\tau_e$ в одному елементі псевдовипадкової послідовності велика,

$$\text{оскільки } \max \frac{\omega_{\partial}}{\omega_T} = \max \frac{\nu_r}{C} \approx \frac{10^4}{3 \cdot 10^8} \approx 10^{-4}.$$

Оскільки при $\omega_{\partial} = 0$ $Z_{\text{вих}} = Z_0 = N\tau_e$, то формула (3.79) набуває виду у формі втрат енергії α за період сигналу.

$$\alpha = 1 - \frac{Z_{\text{вих}}}{Z_0} = \frac{1}{2C} \left| \sum_{k=1}^{\infty} \frac{R^{(k)} \tau_e^{k-1} N^{k-1}}{(k-1)!} \right| \quad (3.80)$$

і

$$\sum_{k=1}^{\infty} \frac{R^{(k)} \tau_e^{k-1}}{2C(k-1)!} N^{k-1} - \alpha_0 = 0, \quad (3.81)$$

де $\alpha_0 < 1$ – коефіцієнт, що характеризує відносні втрати за рахунок даного ефекту.

Наприклад, задаючись втратами, для випадку довільного руху можна визначити позитивний корінь алгебраїчного рівняння (3.81) відповідного ступеня, тобто $N = N_{max}$, при якому спостерігаються втрати α_0 .

Відповідно до принципу дисипативності (обмеженості) енергії значення похідних звичайно зменшуються. Кожен член ряду Тейлора обмежує число елементів послідовності, якщо похідні позитивні (рис.3.8).

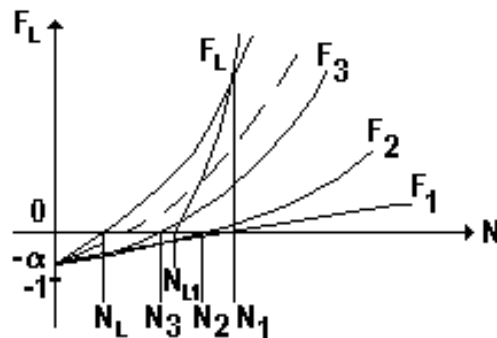


Рис. 3.8. Символічна залежність усіченого ряду від числа елементів сигналу

Іноді достатню збіжність до позитивного кореня може дати метод дотичних.

Для розрахунку значення припустимого числа елементів N_{ℓ_n} можна використати ітераційний вираз для пошуку кореня рівняння порядку l і для n -го наближення можна записати у вигляді

$$N_{\ell(n)} = N_{\ell(n-1)} - \frac{\sum_{k=1}^{\ell} \frac{R^k \tau_e^{k-1} N_{\ell(n-1)}^k}{2C(k+1)!} - \alpha}{\sum_{k=1}^{\ell} \frac{R^k \tau_e^{k-1} N_{\ell(n-1)}^{k-1}}{2C(k+1)!}}. \quad (3.82)$$

З виразу (3.77) видно, що при зменшенні значення $R^{(k)}$, зі збільшенням k і при зменшенні τ_e збіжність ітерацій N_{ℓ} у (3.82)

поліпшується й практично можна обмежитися першими наближеннями ($n=1, 2$):

$$N_{\ell_0} = N_1 = \frac{2\alpha C}{\dot{R}}. \quad (3.83)$$

Для випадку, коли $R^{(k)}\tau^k < R^{(k-1)}\tau^{k-1}$,

$$N_{\ell(n)} \approx N_{\ell(n-1)} \left(1 - \frac{1}{\ell}\right) + \frac{\alpha}{\sum_{k=1}^{\ell} \frac{R^{(k)}\tau_e^{k-1} k N_{\ell(n-1)}^{k-1}}{2C(k+1)!}} - \frac{1}{\ell R^{(\ell)}} \sum_{k=1}^{\ell-1} R^{(\ell-k)} \tau_e^k \prod_{i=1}^k (\ell - i + 2). \quad (3.84)$$

Таким чином, якщо здійснити кореляційну обробку ФМ ПВП ідеальним інтегратором із часом інтегрування T , то навіть при дотриманні умови (3.75), що виключає зрив автосупроводу за частотою, будуть спостерігатися втрати енергії сигналу, пов'язані з ефектом Доплера. При цьому втрати визначаються в основному зміною тривалості елементів сигналу для випадку, коли тактова частота не залежить від несучої. Втрати будуть залежати від різниці $\omega_o(t)$ несучої прийнятої частоти й частоти $\omega_{n_2}(t)$ генератора, що перестроюється. Розрахунок буде той же, тільки замість $R^{(k)}$ варто підставляти динамічну похибку ($R^{(k)} - R_{фанч}^{(k)}$), що відображає відповідний астатизм ФАПЧ. Крім того, наявність доплерівського ефекту для зазначених випадків накладає обмеження зверху на кількість елементів сигналу (3.80). При цьому максимально можлива кількість елементів (база) сигналу значно залежить від виду (траєкторії руху цілі) і зменшується тим більше, чим більше величина й ступінь похідних $R^{(k)}$ (або динамічних похибок $\Delta R^{(k)} = R^{(k)} - R_{фанч}^{(k)}$ ФАПЧ), тобто при рівних величинах $R^{(k)}(\Delta R^{(k)})$ «внесок» похідних у втрати енергії сигналу пропорційний N_l , де l – порядок усічення ряду. При фіксованому $N < N_{max}$ величина вихідного ефекту (3.76) тим більше, чим менше величини й ступені похідних R^k і чим вище порядок астатизму ФАПЧ і менше динамічні похибки $\Delta R^{(k)}$ при тактовій частоті, що залежить від несучої.

Наприклад, якщо $\alpha=0,5$ і $R \approx 10$ км/с, то

$$N_{max} = \frac{2\alpha C}{R} = \frac{2 \cdot \frac{1}{2} \cdot 3 \cdot 10^8}{10^4} = 3 \cdot 10^4$$

Якщо $\alpha=0,5$ і $\Delta R^{(1)} = 0,5$ м/с (похибка спостереження за радіальною швидкістю цілі), то

$$N_{max} = \frac{2\alpha C}{\Delta R^{(1)}} = \frac{2 \cdot \frac{1}{2} \cdot 3 \cdot 10^8}{0,5} = 6 \cdot 10^8.$$

З формули (3.80) і наведених прикладів випливає, що оцінювання й відстеження частоти сигналу істотно збільшують базу широкосмугового сигналу, що припустимо використовувати в радіотехнічній системі із заданими втратами в енергетичному потенціалі.

3.5. Вплив нестабільності тактової частоти на якість кореляційної обробки фазоманіпульованого сигналу

Коли стабільності тактової й несучої частоти незалежні, тобто обумовлені різними причинами, і виконується умова автосупроводу за тактовою частотою

$$\omega_\delta(T) - \omega_\delta(0) \ll \frac{2\pi}{T} \qquad \omega_0 \tau_e \leq 10^5,$$

то вихідний ефект можна записати у вигляді

$$Z_{вих} \approx \sum_{i=1}^N \left(\int_{(i-1)\tau_e}^{i\tau_e} dt - \left| \int_{i\tau_e - \delta t_j}^{i\tau_e} dt \right| \right) = N\tau_e - \sum_{i=1}^N |\delta t_j|, \qquad (3.85)$$

де $\delta t_j = \tau_{ej} - \tau_e$ – нормально розподілені, центровані щодо ідеальних тактових моментів величини з дисперсіями σ_j^2 .

Втрати енергії сигналу на одному тактовому інтервалі становлять $|\delta_j|$.

Тоді щільність розподілу випадкових втрат $|\delta_j|$ на тактовому інтервалі можна визначити таким чином:

$$P(|\delta_j|) = \sqrt{\frac{2}{\pi\sigma_j^2}} e^{-\frac{|\delta_j|^2}{2\sigma_j^2}}, \quad (3.86)$$

при $|\delta_j| \in [0, \infty)$, з математичним сподіванням

$$M[|\delta_j|] = 2\sqrt{\frac{2}{\pi}}\sigma_j, \quad (3.87)$$

де σ_j^2 – як і раніше, дисперсія розподілу джитера й втрат енергії на інтервалі.

Отже, середнє на всьому інтервалі спостереження значення енергії вихідного сигналу, яке позначається $\bar{Z}_{вих}^{\delta t_i}$, можна одержати з (3.84):

$$\bar{Z}_{вих} = N\tau_e - \sum_{i=1}^N M[|\delta_j|] = N(\tau_e - 2\sqrt{\frac{2}{\pi}}\sigma_j). \quad (3.88)$$

Оскільки $N\tau_e = Z_0$ і $\sigma_j = \sigma_{j-1} = \sigma$, (3.88) можна подати у вигляді

$$\bar{Z}_{вих} = Z_0 \left(1 - 2\sqrt{\frac{2}{\pi}} \frac{\sigma}{\tau_e} \right). \quad (3.89)$$

Для $\sigma \ll \tau_e$

$$\frac{\sigma}{\tau_e} \approx \frac{\Delta f_T}{f_T}, \quad (3.90)$$

де $\frac{\Delta f_T}{f_T}$ – короткочасна відносна (неврахована) нормально розподілена нестабільність частоти – джитер, із шириною спектра більше 10 Гц.

Дійсно, з точністю 10 %

$$f_T - f_{Tj} = \frac{1}{\tau_e} - \frac{1}{\tau_e + \delta t_j} \approx \frac{\delta t_j}{\tau_e^2}.$$

Тоді $\frac{\Delta f_{Tj}}{f_T} = \frac{\delta t_j}{\tau_e}$, оскільки $\tau_e = \frac{1}{f_T}$.

Нестабільність генераторів тактової частоти обмежує можливості застосування цифрових фільтрів, ліній затримок та ін.

Короткочасна нестабільність тактової частоти порівняно мало впливає на втрати енергетичного потенціалу. Вона може бути обумовлена як нестабільністю передавального генераторного обладнання, так і впливом завад.

3.6. Вплив довготривалої часової затримки на вихідний ефект приймача

Якщо приймач працює в режимі спостереження за часом затримки, то похибка синхронізації позначиться на середній величині вихідного ефекту сигналу, що використовується для передачі інформації або вимірювання радіальної швидкості.

Визначимо середнє значення вихідного ефекту в припущенні, що протягом періоду ФМ ПВП сигналу похибка синхронізації змінюється незначно. У режимі синхронізації в

інтервалі затримок $-1 \leq \frac{\delta t}{\tau_e} \leq 1$ нормовану автокореляційну функцію сигналу можна подати у вигляді:

$$\frac{Z}{Z_c} = \rho = 1 - \frac{|\delta t|}{\tau_e}. \quad (3.91)$$

Оскільки канал часової синхронізації досить вузькосмуговий, з великою ймовірністю можна вважати, що похибка синхронізації – нормальний стаціонарний центрований процес із розподілом

$$P\left(\frac{\delta t}{\tau_e}\right) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left[-\left(\frac{\delta t}{\tau_e}\right)^2 \frac{1}{2\sigma_c^2}\right], \quad (3.92)$$

$$\text{Де } \sigma_c^2 = M\left[\left(\frac{\delta t}{\tau_e}\right)^2\right].$$

У цьому випадку, з огляду на (3.91) і (3.92), щільність імовірності значення ρ визначається як

$$P(\rho) = 2p\left(\frac{\delta t}{\tau_e}\right) \frac{\partial \left|\frac{\delta t}{\tau_e}\right|}{\partial |\rho|} = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sigma_c} e^{-\frac{(1-\rho)^2}{2\sigma_c^2}}. \quad (3.93)$$

Тоді можна знайти математичне сподівання $M[\rho]$ середнього рівня автокореляційної функції ρ :

$$M[\rho] = \int_0^1 \sqrt{\frac{2}{\pi}} \frac{\rho}{\sigma_c} e^{-\frac{(1-\rho)^2}{2\sigma_c^2}} d\rho = \operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma_c}\right) - 2\sqrt{\frac{2}{\pi}}\sigma_c \left(1 - e^{-\frac{1}{2\sigma_c^2}}\right).$$

Для малих дисперсій вандеру формула для середнього ρ записується простіше:

$$M[\rho] \approx 1 - 2\sqrt{\frac{2}{\pi}}\sigma_c. \quad (3.94)$$

Оскільки середнє за часом (за достатню кількість періодів сигналу), протягом яких процес $\delta t(t)$ вважається стаціонарним,

рівним середньому за ансамблем, то втрати вихідного ефекту можна розраховувати за (3.94). При цьому систематичні похибки вважають врахованими, а динамічні – малими. Середньоквадратична похибка вандеру σ_c реальних систем зв'язку досягає $10^{-1} \div 10^{-2}$. Тому втрати енергії $2\sqrt{\frac{2}{\pi}}\sigma_c$ становлять 17,6 %. Такі втрати варто враховувати в системах, у яких спостерігаються інші численні втрати й загасання.

3.7. Одночасний вплив розстроювань, збурювань і неідеальностей на вихідний ефект приймача

Система, що використовує ШШС типу ФМ ПВП сигнал, може призначатися як для передачі цифрової або аналогової інформації, так і для вимірювання параметрів руху цілі, що впливають на параметри сигналу. Якщо припустити, що система працює з когерентним сигналом і здійснює його когерентну обробку, то навіть у випадку передачі інформації вона відслідковує два параметри: фазу й затримку.

Нехай вихідний сигнал має вигляд

$$S(t - \delta t) = G_\mu S_0 I_T (1 - \delta t) \sum_{k=1}^N I_{\tau_e} [t - \delta t - (k - 1)\tau_e] e^{j(\omega_0 - \omega_\partial)(t - \delta t) + \psi_c + a_k \pi}, \quad (3.95)$$

де $I_T(t)$ – прямокутна зрізаюча функція тривалості, яка дорівнює періоду сигналу T ;

$I_{\tau_e}(t)$ – прямокутна функція тривалості, яка дорівнює тривалості елемента τ_e ;

S_0 – амплітуда;

G_μ – мультиплікативний множник, що враховує загасання амплітуди при поширенні радіохвиль, при обробці й т.д.;

ψ_0 – сумарна похибка за фазою за рахунок флуктуації фазової діаграми спрямованості, змін режимів роботи РТС і т.д.;

a_k – алгоритм – (код) ФМ ПВП сигналу;

ω_∂ – як і раніше, розстроювання частоти.

На виході високочастотного тракту до помножувача сигнал спотвориться і буде дорівнювати

$$S_{\lambda}(t) \approx \mu S_0 K_{\text{вч}} I_T(t - \delta t) \sum_{k=1}^N U_{\tau_e} [t - \delta t - (k-1)\tau_e] e^{j(\omega_0 - \omega_0)(t - \delta t) + a_k \pi + \psi_0}, \quad (3.96)$$

де $K_{\text{вч}}$ – коефіцієнт передачі високочастотного тракту;

$$U_{\tau_e} = \begin{cases} |1 - e^{-\Pi_{\text{вч}}(t-t_1)}| & a_m = a_k \oplus a_{k-1} = 1 \\ I_{\tau_e}(t - t_1) & a_m = 0 \quad t \in [0, \tau_e]. \end{cases}$$

Опорний сигнал має вигляд

$$S_{\text{он}}(t) = I_T(t) \sum_{i=1}^N I_{\tau_e} [t - (i-1)\tau_e] e^{j(a_i \pi + \omega_0 t)}. \quad (3.97)$$

Припустимо, що приймач здійснює кореляційну обробку сигналу з ідеальним інтегруванням.

Тоді на виході корелятора в момент $t=T$ згідно з (3.96) і з (3.97) одержимо

$$Z_{\text{вих}} = \int_0^T [\dot{S}_{\lambda}(t - \delta t) + \dot{n}(t)] S_0(t) dt, \quad (3.98)$$

де $n(t)$ – нормальний білий стаціонарний шум.

Тут і далі передбачається малість взаємокореляційної функції шуму й сигналу в порівнянні з автокореляційною функцією сигналу, що необхідно для надійних вимірювань дальності й швидкості.

Тому

$$Z_{\text{вих}} \approx \int_0^T \dot{S}_{\lambda}(t - \delta t) S_0(t) dt. \quad (3.99)$$

З огляду на те, що

$$ReU ReU = Re\left(\frac{u^*U}{2} + \frac{uU^*}{2}\right),$$

можна записати

$$Z_{вих} = \frac{1}{2} Re\left(\int_0^{T-\delta_0} \dot{S}_1(t-\delta t)\dot{S}_0^*(t)dt + \int_0^{T-\delta} \dot{S}_1^*(t-\delta t)\dot{S}_0(t)dt\right) = \frac{1}{2} Re(I_{1Z} + I_{2Z}).$$

При цьому

$$I_{2Z} = \int_0^{T-\delta} I_T(t-\delta t)I_T(t)\sum_{k=1}^N \sum_{i=\left[\frac{\delta t}{\tau_e}\right]}^N U_{\tau e} I_{\tau e i} [t - (i-1)\tau_e] \times e^{j[2\omega t - \omega_0 t + \psi_0 + (a_k \oplus a_i)\pi]} dt \approx 0.$$

Оскільки обвідна крива змінюється повільно й коливання осцилюють із частотою $2\omega_0$.

Тоді

$$Z_{вих} = \frac{1}{2} Re I_{1Z},$$

де

$$I_{1Z} = \int_0^T \sum_{k=1}^N \sum_{i=1}^N U_{\tau e k} I_{\tau e i} e^{j(-\omega t - \omega_0 t + \psi_0 + a_m \pi)} dt, \quad (3.100)$$

тому що обробка сигналу і вимірювання параметрів здійснюються в режимі синхронізації при відповідно малому часі запізнювання $\delta t < \tau_e$ і малих розстроюваннях.

Якщо припускати, що сумарний набіг фаз $\psi_0 = \sum_i \psi_i$ створюється просторовим коливанням чотирьох незалежних антен і фазових діаграм спрямованості, то $\psi_i = \xi_i \cdot \theta_i$.

Через те, що θ_i розподілені за нормальним законом, ψ_0 має нормальний закон розподілу з 8 степенями вільності, з нульовим

математичним сподіванням і дисперсією $\sum_{i=1}^8 \xi_i^2 D_{oi}$. Тут D_{oi} – дисперсія похибки автосупроводу за кутом.

Припустимо також, що спектр флуктуацій частоти розстроювання ω_{∂} і похибки синхронізації δt досить вузькі в порівнянні зі смугами відповідно ФАПЧ і каналу синхронізації, тобто, що час кореляції процесів перевершує часи інтегрування в каналах. У реальних системах це звичайно виконується. Тоді можна вважати, що ψ_c , ω_{∂} і δt незмінні протягом періоду ФМ ПВП сигналу.

У цьому випадку $I_{\lambda Z}$ відповідно до характеру (3.99) буде кусково-гладкою функцією часу

$$I_{\lambda Z} = e^{j\psi_0} \int_0^T \sum_{k=1}^N \sum_{i=1}^N U_{\tau_{ei}} \cdot I_{\tau_{ei}} \cdot e^{j(a_k - a_i)\pi} dt. \quad (3.101)$$

Визначимо $Z_{вих}$ для чотирьох інтервалів затримки δt у припущенні малості знакової кореляційної функції ρ (3.96), що практично завжди виконується для безперервного ФМ ПВП сигналу великої тривалості.

На інтервалі $-\tau_e + t_{\lambda} \leq \delta t < 0$ (рис. 3.2)

$$Z_{вих} = \frac{1}{2} Re \sum_{i=1}^N e^{j\psi_0} \left[\int_{(k-1)\tau_e + t_1}^{k\tau_e - \delta t} \left(1 - e^{-\Pi_{\partial} [t - (k-1)\tau_e]}\right)^{a_m} e^{-j\omega_{\partial} t} + \int_{(k-1)\tau_e}^{(k-1)\tau_e + t_1} \left(2e^{-\Pi_{\partial} (k-1)\tau_e} - 1\right)^{a_m} e^{-j\omega_{\partial} t} dt \right]. \quad (3.102)$$

Звичайно частота появи 1 або 0 для псевдовипадкових кодів дорівнює

$$p^*(a_k = 1) = p^*(a_k = 0) = \frac{1}{2}.$$

Середнє за ансамблем a_m для фіксованої позиції «до» визначається як

$$\bar{a}_m^{a_m} = 1 \cdot p(a_m = 1) + 0 \cdot p(a_m = 0) = \frac{1}{2}, \quad (3.103)$$

де $\bar{a}_m^{a_m}$ – означає усереднення (a_m) за ансамблем a_m ,
 $p(a_m) = p(a_k)$, оскільки $a_m = a_k \oplus a_{k-1}$ має
 властивості подібні a_k .

З огляду на зроблені зауваження й співвідношення (3.96) і
 (3.103), можна знайти (3.102), нехтуючи доданком третього
 порядку малості

$$Z_{вих} = \frac{T}{2} Re \left\{ \left[\begin{array}{l} 1 - \frac{|\delta t|}{\tau_e} - \frac{\ln 2}{a} - \frac{1}{2a} \left(1 - 2e^{-a} \left(1 - \frac{\delta t}{\tau_e} \right) \right) + \\ + j\omega_\delta \tau_e \frac{N}{2} - \frac{N^2}{6} \omega_\delta^2 \tau_e^2 \end{array} \right] e^{j\psi_0} \right\}. \quad (3.104)$$

Аналогічно визначаються I_λ для випадків

$$t_1 \leq \delta t \leq \tau_e \quad i \quad \tau_e \leq \delta t \leq \tau_e + t_1.$$

Далі, зауважуючи, що доданки $j\omega_\delta \tau_e \frac{N}{2} - \frac{N^2}{6} \omega_\delta^2 \tau_e^2$ і
 співмножник $e^{j\omega}$ з (3.104) наявні у виразі для $z_{вих}$, на будь-якому
 інтервалі затримки δt можна аналогічно (3.67), використовуючи
 такі ж міркування, записати вираз для $z_{вих}$ у вигляді

$$Z_{вих} = \frac{T}{2} [1 - I(a, 0)] \left(1 - \frac{|\delta t|}{\tau_e} - \frac{\ln 2}{a} \right) Re \left(1 + j\omega_\delta \tau_e \frac{N}{2} - \frac{N^2}{6} \omega_\delta^2 \tau_e^2 \right) e^{j\psi_0}. \quad (3.105)$$

Необхідно зазначити, що вираз (3.105) справедливий для
 систем з розімкнутими й замкнутими зворотними зв'язками в
 каналах спостереження за δt і ω_δ , якщо під відносними
 розстроюваннями розуміти або просто розстроювання, або
 відповідні похибки спостереження за параметрами.

Розглянемо, як розподілені зусилля елементів реальних схем щодо формування вихідного ефекту (3.105) або функції невизначеності в околі малих розстроювань.

Нехай канал вимірювання реальної швидкості має структуру (рис. 3.9), де ВЧТ – високочастотний тракт, X – помножувач, ППЧ – підсилювач проміжної частоти, ФД – фазовий детектор, ФНЧ – фільтр низьких частот, ПГ – перестроюваний генератор.

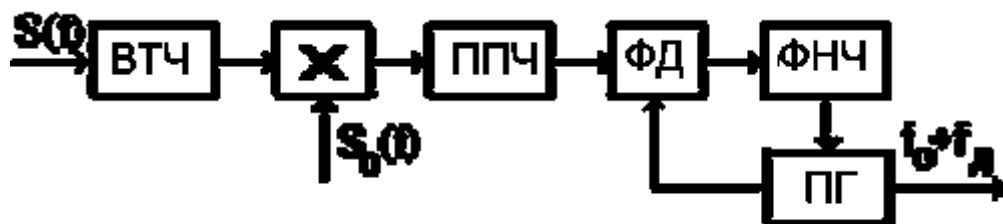


Рис. 3.9

У такому каналі вимірювання частоти операція перемножування прийнятого й опорного сигналів розділена на дві: операції стиску спектра (перемножування прийнятого сигналу на ПВП) і операцію перемножування коливань на фазовому детекторі. Інтегрування перемножених коливань замінюється фільтрацією частково вузькосмуговим ППЧ перед фазовим детектором і основним фільтром нижніх частот. Відома схема (рис. 3.9) здійснює квазіоптимальну обробку сигналу. Вона здатна видавати оцінки, близькі до потенційних. Неідеальності враховані формулою (3.105). Якщо, крім того, врахувати ще неідеальність інтегрування (блоками ППЧ і ФНЧ), випадковість ψ_0 , настроювання систем і усунення систематичних похибок $\bar{\psi}_0^{\psi_0}$, $\frac{\ln 2}{a}$, то вираз (3.105) можна подати у вигляді

$$Z_1 = \mu Z_0 [1 - \varphi(a)] f(TP_{\text{фнч}}) \left(1 - \frac{|\delta t|}{\tau_e}\right) \left(1 - \frac{1}{2} \bar{\psi}_{\text{од}}^{\psi_0}\right) \left(1 - \frac{N^2}{6} \omega_{\text{д}}^2 \tau_3^2\right), \quad (3.106)$$

де $\bar{\psi}_0 = \sum_i \left[\overline{|\psi_0 - \bar{\psi}_0^{\psi_0}|^2} \right]^{\psi_0} = 2 \sum_{i=1}^{\delta} \xi_i D_{oi};$

$\bar{\psi}_{o\partial}$ – похибка, згладжена системою ФАПЧ;

Z_0 – енергія сигналу без втрат, $Z_0 = \frac{T}{2}$;

$f(T\Pi_{\text{фнч}})$ – відносні втрати енергії за рахунок неідеальності інтегрування;

$\Pi_{\text{фнч}}$ – смуга фільтра нижніх частот.

В (3.90) показано, що внеском у δt за рахунок нестабільності існуючих стандартів частоти можна знехтувати й похибка каналу синхронізації δt обумовлюється в основному рухом рухомих об'єктів і астатизмом каналу синхронізації. Оскільки для реальних сигналів і параметрів руху об'єктів добуток $R^{(k)}\tau_e^{k-1}$ (k -ї похідної від дальності на τ_e^{k-1}) швидко зменшується, ряд (3.97) швидко сходиться й у першому наближенні варто обмежитися лише $R^{(1)}$.

Тоді δt_{∂} з (3.78) і (3.83) впливає

$$\delta t_{\partial} = \tau_e \frac{\Delta R^{(1)} N}{2C}.$$

При цьому припускаючи, що синхронізація здійснюється шляхом зміни тактової частоти $f(R)$, визначається різниця величини максимальної швидкості об'єкта й того її значення, що ще відслідковує канал синхронізації.

Таким чином, при кореляційній обробці ШШС на формування функції невизначеності (вихідного сигналу) впливають розстроювання і неідеальності приймача.

При цьому необхідно відзначити таке.

1. Вихідний ефект як багатовимірна функція розстроювань і неідеальностей може бути поданий у факторизованому вигляді (у вигляді добуток одновимірних функцій від розстроювань і неідеальностей).

2. Функції від розстроювань і неідеальностей мають монотонний характер.

3. Умова забезпечення необхідного рівня Z_I (3.106) висуває більш жорсткі вимоги щодо розстроювань і неідеальностей, ніж

зроблені припущення про їх малість. Тобто для того, щоб при $\alpha > 1$ $\frac{Z_1}{Z_0} \geq \frac{\lambda}{\alpha}$, необхідно, щоб при рівному внеску у втрати вихідного ефекту

$$\varphi(a) \leq 1 - \frac{1}{\sqrt{\alpha}}, \quad \left| \frac{\delta t_{\partial}}{\tau_e} \right| \leq 1 - \frac{1}{\alpha^{1/5}}, \quad \psi_0 \leq 2\sqrt{1 - \frac{1}{\sqrt{\alpha}}},$$

$$\omega_{\partial} \leq \sqrt{\frac{6}{5}} \frac{2\sqrt{1 - \alpha^{-1/5}}}{T}, \quad ТП_{\text{фнч}} \leq \ln \alpha^{1/5}, \quad a \geq 5$$

наприклад, при $a=2$

$$I(a) \leq 0,105; \quad \left| \frac{\delta t_{\partial}}{\tau_e} \right| \leq 0,105; \quad \bar{\psi}_{\partial}^2 \leq 0,438,$$

$$\omega_{\partial} \leq \frac{0,48}{T}; \quad ТП_{\text{фнч}} < 0,09.$$

4. Найжорсткіші вимоги висуваються до смуги пропускання фільтра нижніх частот і розстроювання ω_{∂} по частоті, а найменш вибагливі – до дисперсії сумарної фазової похибки ψ_{∂} .

ВИСНОВКИ

У результаті розгляду впливу паразитних ефектів на кореляційну обробку сигналу можна зробити нижчеподані висновки.

1. Обмеження результуючої смуги високочастотного тракту від передавача до помножувача приймача приводить до постійної затримки, пов'язаної з тривалістю перехідних процесів, втратами енергії й до незначних спотворень згортки прийнятого й опорного сигналів. Постійне запізнювання враховується в каналі синхронізації, тому при оптимізації системи варто брати до уваги лише втрати рівня вихідного ефекту залежно від відношення α результуючої смуги високочастотного тракту до ширини спектра широкосмугового шумоподібного сигналу (ШШС).

2. У режимі автосупроводу за кутами флуктуації максимуму діаграми спрямованості антени при моноімпульсному методі

навколо напрямку на об'єкт приводять до середніх втрат потужності сигналу, що залежить від відношення α ширини діаграми за потужністю до дисперсії сталої похибки супроводу за кутами. Втрати істотні вже при $\alpha \leq 8$.

3. За умови стійкого спостереження системи ФАПЧ за частотою ефект Доплера сильніше позначається на втратах енергії за рахунок зміни тактової частоти (тривалості елементів) і слабше за рахунок зміни несучої частоти. Тому взаємний зв'язок тактової й несучої частот для визначення втрат при даному ефекті значення не має, відіграє роль лише динамічна похибка спостереження за частотою, астатизм фільтра й відповідно величина радіальної швидкості і її похідних. Втрати енергії значно зростають зі збільшенням кількості елементів, зі зменшенням порядку астатизму й при складному русі об'єкта з більшою величиною похідних вищих порядків.

4. Втрати енергії вихідного сигналу зростають пропорційно короткочасній нестабільності тактової частоти й кореню квадратному з кількості елементів ШШС. Однак лише для більшої кількості елементів ШШС ($N \geq 10^{10}$) і значних шумів при простих засобах стабілізації тактової частоти $\left(\frac{\Delta f_T}{f_T} \approx 10^{-5} \right)$

втрати енергії можуть бути істотними.

5. При великій кількості елементів ШШС і відносно малих динамічних похибках відстеження частоти в каналі синхронізації втрати енергії за рахунок ефекту Доплера й нестабільності частоти мають складну залежність. Однак при сучасних стандартах частоти, малих шумах у каналі синхронізації й застосовуваній кількості елементів можна брати до уваги лише втрати енергії за рахунок сумарних похибок, включаючи динамічні похибки при наявності складного руху об'єкта або недостатнього порядку астатизму каналу синхронізації.

6. Для малих розстроювань за частотою й за часовою затримкою вихідний сигнал (функцію невизначеності) можна подати у вигляді добутку одновимірних функцій від аргументів, що характеризують розстроювання і неідеальності (3.106), що важливо при формалізації задач оптимізації систем, які використовують ШШС.

7. Результати розділу можуть виявитися корисними для таких інженерних розрахунків:

а) для розрахунку впливу обмеженої смуги пропускання на величину вихідного фазоманіпульованого за законом псевдовипадкової послідовності сигналу при різних розстроюваннях у каналі синхронізації й різних коефіцієнтів знакової автокореляції ПВП (при розстроюванні на тривалість елемента);

б) для розрахунку втрат енергетичного потенціалу при просторових флуктуаціях діаграми спрямованості антени в режимі автосупроводу за кутами, для розрахунку відповідного розподілу рівня сигналу й моментів, розподілу для визначення оптимального співвідношення між дисперсією похибки автосупроводу за кутом і спрямованістю антени, що дозволяє досягти максимальної дальності радіолінії;

в) для розрахунку втрат енергій сигналу за рахунок руху об'єкта при кореляційній обробці сигналу й спостереження за затримкою з динамічними похибками; отримані рекурентні співвідношення для розрахунку максимально припустимої кількості елементів ФМ ПВП сигналу при неочікуваному законі руху;

г) для розрахунку втрат енергії сигналу за рахунок нестабільності тактової частоти в каналі синхронізації; для розрахунку максимально припустимої кількості елементів сигналу при заданій нестабільності тактової частоти каналу синхронізації;

д) для розрахунку втрат енергії сигналу за рахунок одночасного впливу динамічних похибок каналу синхронізації при неочікуваному русі об'єкта й нестабільності тактової частоти;

е) для розрахунку втрат енергії сигналу за рахунок впливу похибки синхронізації з нормальним розподілом;

ж) для розрахунку втрат енергії сигналу при кореляційній обробці за рахунок одночасного впливу розстроювань і неідеальностей;

з) для оцінки вимог до розстроювань і неідеальностей при заданих втратах енергії.

Розглянутими в розділі ефектами далеко не обмежується список факторів, які варто врахувати при синтезі РТС. У даному розділі наведені лише істотні результати або найменш висвітлені у відомій літературі.

4. ПАРАМЕТРИЧНИЙ СИНТЕЗ ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ РАДІОТЕХНІЧНИХ СИСТЕМ

У даному розділі містяться відповіді на такі питання:

- який критерій оптимальності і який склад показників якості варто використовувати при оптимальній побудові інформаційно-вимірювальних радіотехнічних систем (ІВРТС);
- як визначити оптимальні технічні параметри, і що робити із сигналами й вимірювальними структурами;
- які розглядати ефекти обробки сигналу;
- як урахувати параметри ефектів і системи при оптимізації;
- за якими формулами розраховувати оптимальні параметри;
- який вигаш при цьому варто очікувати й т. д.

Викладені методи оптимізації, розрахункові співвідношення й алгоритми дозволяють одержати оптимальний вигляд ІВРТС на етапі ескізного проектування. Отримані рішення про оптимальні технічні параметри, сигнали й вимірювальні структури повинні служити вихідними даними для схемного проектування функціональних елементів, підсистем і вузлів ІВРТС.

Завдання на прочитання

1. Зрозуміти, як у даній задачі ураховуються практично усі показники якості системи й усі технічні параметри.
2. Як ураховується вартість системи?
3. Які властивості має вартість?
4. Зрозуміти, що зміниться в задачі, якщо критерієм якості стане завадостійкість передачі інформації.
5. Оцінити переваги викладеного методу розв'язання задачі.

4.1. Показники якості радіотехнічних систем

У розд. 2 показано, що найбільш об'єктивним критерієм оптимальності можна вважати умовний критерій, що враховує сукупність показників якості, поданих у ТТВ.

До сучасних радіотехнічних систем висувають безліч таких тактико-технічних вимог (ТТВ), виконання яких забезпечує

створеній системі успішне функціонування й вирішення поставлених завдань. До основних показників якості РТС, які так чи інакше враховуються при проектуванні, насамперед варто віднести:

- 1) точність вимірювань параметрів руху об'єкта;
- 2) завадостійкість передачі інформації прямого й зворотного каналів;
- 3) точність звірення часу й прив'язки шкал;
- 4) час вимірювань і передачі інформації;
- 5) надійність функціонування;
- 6) електромагнітна сумісність;
- 7) швидкість передачі інформації;
- 8) пропускна здатність системи;
- 9) живучість;
- 10) завадозахищеність;
- 11) економічність;
- 12) вага апаратури і її складових частин;
- 13) обсяг апаратури і її складових частин;
- 14) інші показники якості.

При оптимізації РТС на етапі проектування, а також для оцінки зазначених показників якості зручніше виражати їх у конкретній числовій формі у вигляді вектора, що досить повно описує якість виконання основної задачі. Вектор наведених вище якісних показників можна використовувати для оптимізації будь-яких РТС. Показники якості є взаємозалежними і взаємообумовленими. Тому більш повний їхній облік приводить прийнятну модель системи до більшої відповідності з реальною РТС. У цьому сенсі розв'язання задачі оптимізації за критеріями 1-14 рівноцінне глобальному розв'язанню при фіксованих значеннях 13-ти незалежних критеріїв або нульовому наближенню при наявності зв'язків між критеріями.

У розділі сформульована й вирішена задача оптимізації однопараметричної вимірювальної системи за критерієм мінімуму похибки вимірювання параметра руху абонента при обмежених асигнуваннях на систему. У даному розділі буде формалізована й розв'язана задача оптимізації систем на множині технічних, сигнальних і структурних параметрів, тобто задача

параметричного синтезу систем з урахуванням основних критеріїв.

Ідею оптимізації в спрощеному вигляді можна викласти таким чином. Передбачається, що є кореляція або функціональна залежність між основними технічними параметрами й вартістю системи, тобто, що при поліпшенні якості вартість системи, загалом кажучи, монотонно збільшується. Наприклад, при збільшенні потужності або при зменшенні нестабільності частоти вартість системи зростає. Задані вимоги до якісних показників системи є обмеженнями на ті технічні параметри, від яких залежать ці якісні показники. Наприклад, похибка передачі інформації або вимірювань параметрів руху залежить від відношення сигнал/шум на виході каналів. У свою чергу відношення сигнал/шум залежить від таких технічних параметрів, як потужність передавача, коефіцієнт підсилення антени, коефіцієнт шуму приймача й т.д.

Якщо для системи використовувати граничні, найкращі, технічно реалізовані значення всіх параметрів, то вартість такої системи буде надзвичайно велика і навряд чи виправдана. До показників якості РТС, що працюють на малих відстанях, не висувають високих вимог, оскільки реалізація граничних значень параметрів являє собою науково-технічне досягнення й вимагає найвищих і тому невиправданих асигнувань, крім того, поліпшення параметрів незначно й мало позначається на якісному показнику.

Навіть при слабких вимогах до якісних показників системи далеко не однаково, які технічні показники вибирати. Якщо один з технічних параметрів вибирати технічно реалізованим, то сумарні витрати на системи можуть істотно зрости, у той час як якісні показники зміняться не істотно. Наприклад, якщо взяти граничний коефіцієнт підсилення антени, це вимагатиме додаткових асигнувань на антену за рахунок асигнувань на передавач й інші блоки, що дозволить знизити потужність передавача й інші параметри. Але добуток потужності передавача на коефіцієнт підсилення антени повинен при цьому зменшитися при тих же асигнуваннях на забезпечення цих параметрів. Те ж відбудеться при граничній потужності передавача. Тому очевидно, що існують оптимальні значення

параметрів, при яких досягається максимальне відношення сигнал/шум при заданих асигнуваннях на систему. Задача синтезу параметрів полягає в тому, щоб знайти цей мінімум і значення оптимальних параметрів системи й усіх її каналів.

4.2. Вартість блоків системи як обмеження на відповідні технічні параметри

В умовах обмеженості асигнувань на створення цієї системи велике значення має оптимальний розподіл засобів, відпущених на створення РТС, що забезпечує найвищу точність вимірювача при заданих витратах. Дійсно, при сучасному стані науки й техніки найвища точність може бути реалізована лише системою, що має гранично кращі параметри. Однак вартість системи значно зростає в порівнянні з ледве менш точними, тому що граничне значення кожного параметра є сучасним науково-технічним досягненням. Очевидно, що вартість має ту ж значущість, що й точність.

Вартість системи залежить від часу, номера серії, року випуску й т.д., а період «морального старіння» техніки порівняний з терміном служби. Однак вартість системи незмінна на відповідному досить короткому інтервалі часу поступового еволюційного розвитку техніки й виробництва й може бути перерахована до моменту готовності виробу до складання. При революційному розвитку техніки й виробництва, тобто при винаході нових елементів, принципів роботи апаратури, методики вимірювань, способів виробництва, оптимальність технічних параметрів змінюється, оскільки, як вартість системи, що забезпечує при правильній реалізації системи задану похибку вимірювання, так і, власне, похибка вимірювань при обмеженій вартості реалізації системи може істотно змінюватись. Потреба в оптимізації тим більш необхідна, якщо вимірювачі, що оптимізуються, особливо для великих відстаней, є складними системами великої вартості, на яких ті самі революціонізуючі фактори позначаються в меншій мірі, ніж на малих системах. Крім того, подібна апаратура може бути унікальною і малосерійною. Остання обставина свідчить про те, що в цьому

випадку ускладнюється або зовсім виключається звичайна оптимізація апаратури за рахунок серійності виробництва, розрахована на здорову інженерну інтуїцію (метод «проб і помилок»).

При цьому залежно від того, які елементи й вузли використовує завод-виготовлювач системи, будемо вважати, що статистика собівартостей або цін цих елементів і вузлів узятая на відповідних заводах-постачальниках.

Постановка задачі оптимальної побудови системи для вимірювання радіальної швидкості за критерієм якості полягає в тому, що у вигляді багатовимірної цільової функції, що виражає точність вимірювання частоти Доплера і що залежить від розстроювань і неідеальностей системи, формалізуються вимоги до вузлів, блоків системи, які набувають певного сенсу (розв'язання) лише при наявності обмежень і зв'язків між параметрами.

У техніці завжди є діалектичне протиріччя між технічними параметрами системи, що описують якість відповідних блоків, і вартістю розроблення й виробництва відповідної частини системи. Причому, поліпшуючи технічні характеристики, ми змушені відповідно збільшувати асигнування на систему.

Як показано в літературі [18-20], між параметрами системи і її вартістю існує висока кореляція або навіть відома функціональна залежність. Це дає підстави встановити характер залежності технічних параметрів системи від її вартості як обмежень і шукати глобальний екстремум цільової функції.

Якщо таких залежностей нема, то їх можна одержати за відомими прайс-листами комплектуючих функціональних елементів. Ці дані підлягають обробці з використанням методу перетворення нечітких множин вартості у випадкову величину [4]. Алгоритм перетворення поданий далі.

Отримання техніко-економічних залежностей

1. Для меншого впливу на техніко-економічну статистику технічні параметри, не враховані у даній задачі синтезу і фіксуються із заданою точністю, а статистика збирається для систем одного призначення.

2. Спочатку оцінюється коефіцієнт кореляції ціни і параметрів підсилення та коефіцієнт шуму. Якщо коефіцієнт кореляції не менший 0,6, то в задачі синтезу враховуються ці параметри.

3. Усі параметри потрібно подати у вигляді «чим вони менші, тим краще» для спрощення задачі і для отримання вищевказаних переваг при розв'язанні. Коефіцієнти підсилення подані параметрами L , наприклад $L=I/Ш$.

4. На площині ціна-параметр відображаються і нумеруються всі точки, що належать реальним системам.

5. Вручну чи за вибраним критерієм близькості до осі координат відбраковуються спекулятивні та некращі функціональні елементи систем.

6. Використовується метод найменших квадратів (МНК) для пошуку техніко-економічних залежностей.

7. Формується функція зв'язку, або техніко-економічні обмеження та задача синтезу в цілому.

8. Згідно з принципами У. Оккама, Р. Декарта і Р. Беллмана задача спрощується, тобто розв'язується спочатку для одного каскаду, а потім для двох і т. д. Потім задача зводиться до сепарабельного блокового програмування [50]. Вона має складне і громіздке розв'язання у квадратурах, яке як і у динамічному програмуванні отримується з кінця. Однак воно дає значні переваги.

Розв'язання ж задачі дозволяє знайти не тільки оптимальний розподіл коштів між блоками, але й оптимальні технічні параметри РТС, які визначають оптимальний у цьому сенсі вибір блоків при проектуванні системи.

Оптимізація РТС вимірювання радіальної швидкості виконується за критерієм точності, що обчислюється дискримінаційним методом у режимі автосупроводу, тому синтез пошукових та інших систем повинен здійснюватися аналогічно з урахуванням інших критеріїв, наприклад, максимальної ймовірності виявлення сигналу при заданому часі пошуку при заданій ймовірності виявлення й т.п.

Для випадку, коли задача розв'язана, вибір розв'язання проілюструємо (рис. 4.1) на прикладі залежності середньоквадратичного значення вартості стандартів частоти, що

використовуються як задавальний генератор (ЗГ), від короткочасної (за 1 с) нестабільності частоти $\left(\frac{\Delta f}{f_0}\right)$.

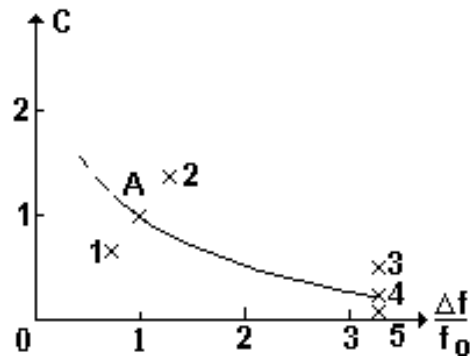


Рис. 4.1. Вид залежності вартості задавального генератора від нестабільності частоти

Нехай точки 1-5 відповідають певним сучасним стандартам частоти, а залежність $C=C(\Delta f_{K_x} / f_0)$ може бути визначена таким чином.

Крива C є кривою середньоквадратичної регресії вартості від параметра відносної нестабільності частоти, отримана шляхом згладжування відомих значень параметрів для відомих зразків (стандартів). Неповна еквівалентність стандартів частоти може пояснюватись розходженнями ваг, об'ємів, експлуатаційними умовами, способами визначення собівартості, технологічними умовами й т.д. Якщо в результаті оптимізації виявлено, що оптимум точності системи досягається в точці А, то варто вибрати той стандарт, що забезпечує кращий параметр (наприклад, 1 або 2 залежно від інших якостей: надійності, зручності в експлуатації, об'єму й т.д.).

Якщо обрано стандарт 1, то при подальшому проектуванні й розробленні доцільно знижувати вимоги до параметра й поліпшувати інші показники так, щоб вартість була не більше вартості в точці А. Коли знизити вимоги не можна, варто обмежуватися вибором стандарту, близького до оптимального.

У випадку, коли більша вартість досягається не за рахунок покращення якості стандарту (точка 2), при розрахунку кривої її не слід враховувати.

Якщо спробувати визначити оптимум цільової функції для набору реалізацій параметра (таких як точки 1-5), то легко виявити, що процедура пошуку оптимуму шляхом перебору комбінацій реалізацій для всіх обмежень різко ускладнюється з ростом розмірності задачі й числа реалізацій цього параметра.

Дійсно, при числі реалізацій i -го параметра m_i і числі параметрів N_c системи кількість обчислень цільової функції при заданих значеннях параметрів і кількість порівнянь по величині дорівнює $M = \prod_{i=1}^N m_i$, наприклад, при $m_i = m > 10$ і $N_c > 14$ $M > 10^{14}$.

Якщо час розрахунку цільової функції дорівнює 1 мкс, то необхідність обчислень зникає, тому що потрібно більше 10 років безперервного рахунку. Тому подання обмежень за вартістю на параметри у вигляді безперервних функцій середньоквадратичної регресії є кращим у цих умовах способом формалізації задачі.

Вартості C_i блоків системи залежать не тільки від необхідних J_i , а також від інших критеріїв, наприклад, надійності, довговічності, ваги й т. п. Тому C_i повинні визначатися при можливо рівних умовах і однаковому призначенні блоків.

Вартість у процесі оптимізації системи є багатовимірною функцією регресії всіх фазових параметрів, які характеризують розстроювання, неідеальності й інші технічні параметри блоків системи, тобто вартість являє собою вибіркоче математичне сподівання ансамблю вартостей при фіксованих значеннях фазових параметрів і мінімальних середньоквадратичних відхиленнях.

Як і будь-яку безперервну й гладку функцію, що має похідні багатьох порядків, її можна подати у вигляді ряду Тейлора в межах точки (X_{i0}, \dots, X_{j0})

$$C(X_1, \dots, X_l) = C(X_{i0}, \dots, X_{j0}) + \sum_{i=1}^l C'_{i0} (X_i - X_{i0}) + \\ + \sum_{ij}^l C''_{i0} \frac{1}{2} (X_i - X_{i0})(X_j - X_{j0}) + \dots,$$

$$\text{де } C'_{i0} = \frac{\partial C}{\partial X_i} / X_i = X_{i0}, C''_{i0} = \frac{\partial^2 C}{\partial X_i \partial X_j} / X_i = X_{i0}, X_j = X_{j0}.$$

Коефіцієнти розкладання можна знайти методом найменших квадратів. Чим більше вибірових значень C і краще їхнє групування навколо математичного сподівання, тим менше невизначеності в об'єктивному законі залежності вартості від фазових параметрів.

На практиці завжди може трапитися, що відомостей про залежності вартості від яких-небудь параметрів немає або статистика незадовільна – немає достатнього числа вибірових значень при великій дисперсії й немає можливості її заповнити і нею скористатися.

Тоді природно оптимізувати систему за іншими параметрами, вибравши таке значення параметра, що не оптимізується, яке частіше використовується в сучасних системах того ж класу. При такій оптимізації якісний показник системи, швидше за все, буде гірше, ніж при повній оптимізації, і тим гірше, коли більш значущий параметр виключається. Імовірно погіршення якісного показника системи буде істотнішим, якщо величини його частотної похідної за параметром, що виключається, більше, якщо припустимий діапазон цього параметра (діапазон використання) більше і якщо, нарешті, ступінь «угадкування» оптимальності параметра при виборі фіксованого значення гірше. Звідси випливає, що якщо малі значення величини, похідної за параметром, і самого параметра, то цей параметр можна виключити, замінивши постійним значенням. Однак, якщо таких параметрів виявиться так багато, що вони будуть визначати істотну частину вартості або вносити істотні погіршення в якісний показник, то виключати їх потрібно коректно.

Критерій економічної ефективності, або вартість, є обмеженням на параметри і має переваги у порівнянні з багатьма іншими критеріями.

Перша перевага – адитивність: вартість системи є сума вартостей її окремих частин і блоків. Ця властивість дозволяє в багатьох випадках спростити задачу.

Наступна перевага – «тотожність ефекту»: за допомогою вартості можна порівняти внесок параметрів у показник якості системи. Вартість, обумовлена суспільно необхідним робочим часом для виготовлення виробу й інших факторів, є мірою, просто й природно пов'язаною з виробами, що реалізують параметри різного найменування. В остаточному підсумку, вона також характеризує виріб, технологію його виробництва і є функцією, що підлягає вивченню в даному аспекті. Залежність вартості або (собівартості) від фазових параметрів, розглянута в даній задачі, являє собою абстраговану, виражену в грошових або в інших одиницях інформацію про стан і розвиток виробництва тих засобів, які забезпечують технічні параметри. Вона відбиває в собі всю «передісторію» розвитку й можливості виробництва в сьогоденні, на якому повинне опиратися інженерне рішення в доступному для огляду майбутньому. Ось чому таке прогнозування властивостей створюваних оптимальних систем повинне здійснюватись на час, менший інтервалу між появами великих винаходів, здатних істотно вплинути на структуру, склад або вартість системи.

4.3. Структура системи, що оптимізується, і позначення її технічних параметрів

При виборі структури системи, що оптимізується для вимірювань радіальної швидкості рухомих об'єктів, можна користуватися результатами інженерного синтезу систем, отриманих, наприклад, методом евристичних рішень (не без упередженості), але, що містять у собі досвід минулого, що не завжди дозволяє досягти оптимальності. Через помилку Ф.М. Вудворда [21] статистичний синтез РТС також не правомірний. Тому синтез системи пропонується за умовними критеріями якості [2-4]. Розглянемо структуру системи з найпростішим дискримінаційним вимірювачем швидкості цілі для оптимізації її параметрів за точнісними й економічними показниками, що враховує також ряд додаткових вимог: умови електромагнітної сумісності й завадостійкості, кращу розв'язку передавальних і приймальних каналів і т. д. Перераховані в підрозд. 4.1 суперечливі вимоги можуть виконуватися,

наприклад, при використанні ширококутових шумоподібних сигналів (ШШС). Будемо розглядати, зокрема, фазоманіпульовані за законом псевдовипадкової послідовності сигнали, не обмежуючи цільності розгляду. Необхідно, щоб система була вимірювачем високої точності і щоб промахи були малоїмовірні, що визначається власне умовою.

Тоді при уважному розгляді (3.106) система вимірювача, що стежить, збігається з реальною (рис. 3.9.) Це означає, що реальна система при заданому відношенні сигнал/шум q , за умови $q \gg 1$ здатна давати ефективну оцінку розстроюванню за частотою $F\delta$. Крім того, у реальній системі враховані всі умови реальної роботи радіолінії: вживаються заходи для боротьби із власними шумами приймача й антен, поліпшується вибірковість за несучою частотою за рахунок її перетворення, поліпшується розв'язка зондувальних і прийнятого сигналів за допомогою рознесення несучих частот і використання окремих антен і т. д. Тобто враховуються різного роду неідеальності системи й вживаються заходи для їхнього ослаблення.

Таким чином, структуру доплерівського вимірювача, що стежить, вважаємо тут відомою, дискримінаторною, а схема системи на рис.4.2 враховує основні особливості реальних систем.

Система оптимізується таким чином, щоб знайти оптимальний розподіл асигнувань по блоках системи або оптимальні параметри, з погляду мінімуму дисперсії похибки вимірювань радіальної швидкості, при обмежених асигнуваннях на досягнення цих параметрів. Далі показано, як у даній задачі врахувати згадані показники якості. Технічні параметри кількісно описують певну якість відповідних блоків. Тому знайдені в результаті розв'язання параметри повинні бути вихідними даними для детального розроблення принципової схеми блоків або для вибору при проектуванні стандартних блоків з оптимальними параметрами.

Назвемо й позначимо основні параметри системи, які є в цільовій функції і за якими буде виконуватись оптимізація.

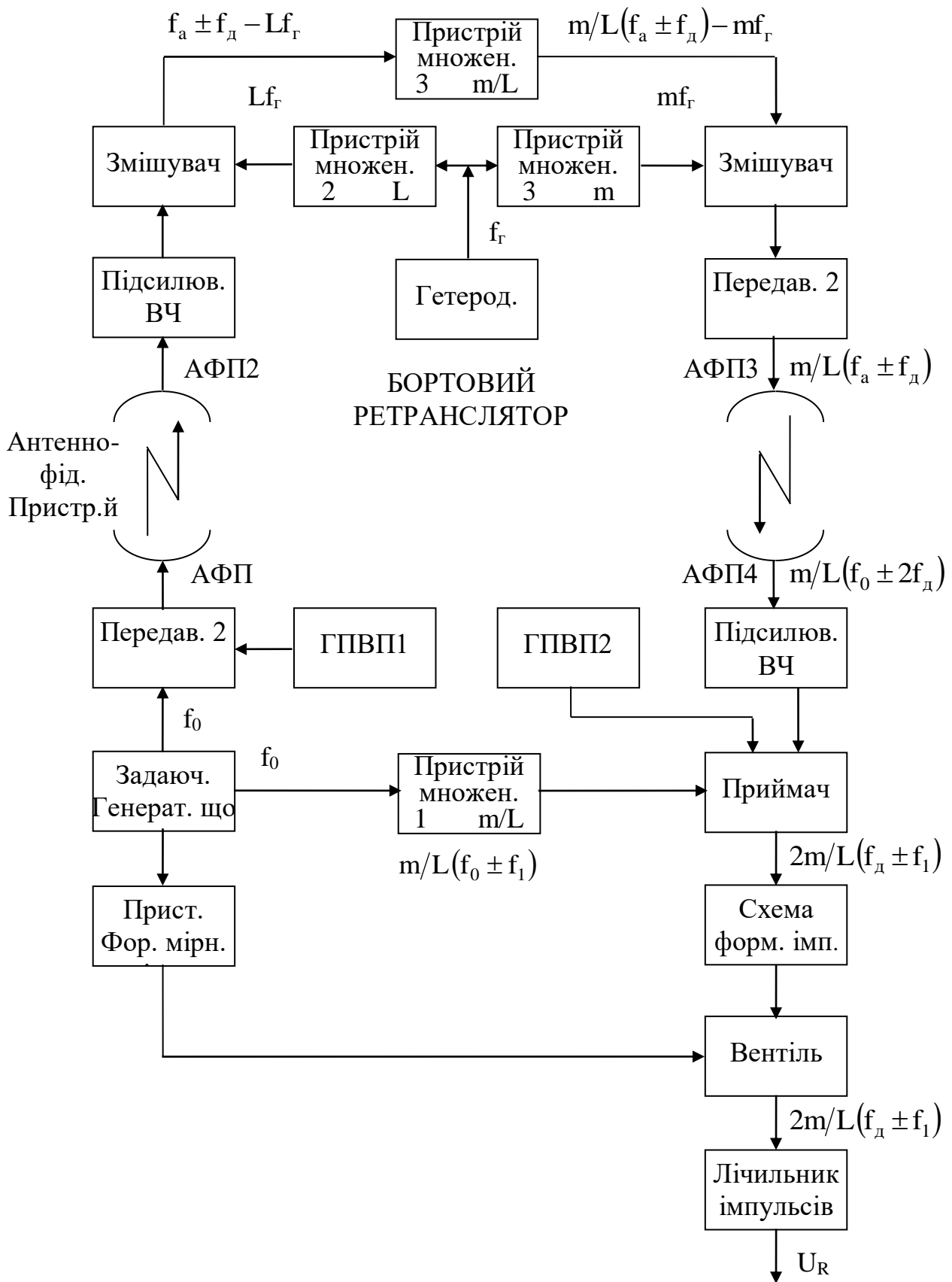


Рис.4.2. Структура вимірювальної доплерівської системи

Кожна антена (A1 і A2 – приймальна й передавальна наземні, A3 і A4 – приймальна і передавальна бортові) може описуватися такими параметрами: коефіцієнтом спрямованої дії G_{1-4} ; коефіцієнтом корисної дії η_{1-4} ; точністю виготовлення антен – коефіцієнтом J_{1-4} , що визначає величину фазових похибок Ψ_i у розкритті антени; шириною діаграм спрямованості у двох площинах $\Delta\theta_{1i}, \Delta\theta_{2i}$; втратами коефіцієнта спрямованої дії за рахунок впливу землі на діаграму спрямованості L_{x1-4} і впливу корпусу рухомого об'єкта на діаграму спрямованості L_{x3-4} ; втратами підсилення антени за рахунок непогодженості поляризації $\chi_{1,2}$ з шумовою температурою антени T_{oa} .

Система автосупроводу за кутами визначається дисперсією похибки D_{1-4} , втратами підсилення антен за рахунок нестабільностей утримання кутів $G_{n1,4}$. Фідерні пристрої визначаються коефіцієнтом згасання L_3 і коефіцієнтом хвилі, що біжить, $_{1-4}I-4$.

Підсилювачі високої частоти ПВЧ1 і ПВЧ2 приймачів визначаються смугою пропускання $P_{пвч}$, коефіцієнтом підсилення $K_{пвч}$ за потужністю, коефіцієнтом шуму $Ш_{пвч}$.

Змішувачі ЗМ1-3 можуть описуватися коефіцієнтом перетворення (або передачі за потужністю або за напругою) L_{npi} і коефіцієнтом шуму $Ш_{npi}$.

Пристрої помножування описуються коефіцієнтом передачі L_{npi} і коефіцієнтом шуму $Ш_{npi}$.

Підсилювачі проміжних частот (ППЧ) визначаються смугою пропускання $P_{пч1,2}$, коефіцієнтом підсилення $K_{пч1,2}$, втратами обробки за рахунок дрейфу підсилення частоти, коефіцієнтом $L_{\delta 1,2}$. Фазова система автопідстроювання частоти визначається (у режимі супроводу за частотою) результуючою смугою, у тому числі смугою фільтра нижніх частот $P_{фнч}$, дисперсією похибки вимірювання D_{wd} , що служить у цій задачі цільовою функцією, дисперсією похибки за рахунок динамічного й іншого ефектів $\Delta R^{(k)}$, втратами післядетекторної обробки L_o .

Канал синхронізації K_c може визначатися результуючою дисперсією похибки синхронізації δt . Задавальний генератор ЗГ являє собою стандарт частоти з подільниками і помножувачами

частоти, що визначається відносною нестабільністю частоти $\Delta f_{k_c} / f_0$.

Гетеродини Гет₁₋₂ можуть визначатися результуючою нестабільністю частоти $\Delta f_c / f_0$.

Помножувачі частоти Мн₁₋₄ визначаються коефіцієнтом шуму $Ш_{мн1-2}$ і коефіцієнтом помноження m/l .

Модулятор визначається коефіцієнтом шуму $Ш_m$ при заданій вихідній потужності і коефіцієнті підсилення.

Передавачі ПРД₁₋₂ визначаються потужністю $P_{прд1-2}$, смугою пропускання $Ппрд1-2$.

Таким чином, оптимізація системи (параметричний синтез) здійснюється за перерахованими та іншими параметрами, які є основними й необхідними для розв'язання поставленої задачі.

При цьому по можливості необхідно враховувати всі ті параметри, які роблять істотний внесок у похибку вимірювання радіальної швидкості V_r . Якщо параметри впливають лише на відношення сигнал/шум на виході системи обробки, то оптимізація параметрів РТС відноситься як до вимірювачів, що стежать, так і до вимірювачів, що не стежать повною мірою. Якщо використовується тільки показник відношення сигналу до шуму, то задача справедлива також для систем передачі інформації.

4.4. Похибка вимірювань радіальної швидкості як критерій для оптимізації системи

Похибка вимірювань радіальної швидкості або іншого параметра рухомого об'єкта є одним з показників якості системи. Тому похибка синхронізації опорного й прийнятого сигналів розглядається як параметр, а не як критерій для оптимізації. Нехай борт об'єкта працює в режимі активної ретрансляції сигналу з перенесенням його широкосмугового спектра на іншу частоту, отриману дробово-раціональним перетворенням частоти $\left(\frac{l}{m}\right)$, як це робиться в реальних системах для розв'язання передавальних і приймальних каналів. Усе викладене додатне і для пасивної ретрансляції, де часова розв'язка каналів

здійснюється за рахунок запізнювання строба або опорного складного сигналу більш ніж на елемент.

Тоді при вимірюванні радіальної швидкості $V_r = \dot{R}$ запитальним методом відносний доплерівський ефект не позначається, гравітаційна складова мала і вираз для V_r має вигляд

$$V_r = \left(\frac{l}{m} \right) \frac{f_D V_c}{2 f_0},$$

де f_0 – несуча частота;

f_D – частота Доплера;

V_c – швидкість поширення радіохвиль,

l, m – коефіцієнт множення $f_{3\Gamma}$.

Оскільки доданки незалежні, дисперсія $D \left[\frac{\Delta V_r}{V_r} \right]$

визначається як

$$D \left[\frac{\Delta V_r}{V_r} \right] = \left(\frac{l}{m} \right)^2 \left(D \left[\frac{\Delta f_D}{f_D} \right] + D \left[\frac{\Delta f}{f_0} \right] + D \left[\frac{\Delta V_c}{V_c} \right] \right), \quad (4.1)$$

де $D \left[\frac{\Delta f_D}{f_D} \right] = \frac{D[\Delta f_D]}{f_D^2}$ – дисперсія відносної похибки вимірювання частоти Доплера;

$D \left[\frac{\Delta f}{f_0} \right]$ – дисперсія сумарної похибки за рахунок розстроювань, нестабільностей частоти й інших факторів, що дають прямий внесок у похибку вимірювань радіальної швидкості;

$D \left[\frac{\Delta V_c}{V_c} \right]$ – дисперсія відносної похибки за рахунок неточного знання умов поширення радіохвиль.

При цьому з рис. 4.2 видно, що

$$D\left[\frac{\Delta f}{f_0}\right] = \frac{\Delta f_{зг}^2}{f_0^2} + \frac{\Delta f_{ng}^2}{f_{ng}^2} + \sum_{i=1}^7 \frac{\Delta f_{умн i}^2}{f_0^2} + \frac{\Delta f_{сч}^2}{f_1^2} + \frac{\Delta f_{\delta}^2}{f_0^2} + \frac{\Delta f_T^2}{f_1^2}, \quad (4.2)$$

де $\Delta f_{зг}$ – нестабільність частоти задавального генератора;

Δf_{ng} – нестабільність частоти генератора, що перестроюється (рис. 3.5);

$\Delta f_{умн i}$ – приведена до f_0 нестабільність частоти помножувачів і перемножувачів;

$\frac{\Delta f_{сч}}{f_1}$ – відносна похибка лічильника;

$\frac{\Delta f_{\delta}}{f_0}$ – приведена до f_0 відносна нестабільність за рахунок

інших блоків бортової й наземної апаратури, що включає невідомі систематичні похибки.

При визначенні дисперсії вимірювання частоти Доплера

$$D\left[\frac{\Delta f_D}{f_D}\right]$$

нас буде цікавити оцінка розстроювання частоти ω_D

при наявності випадкових і не випадкових параметрів, розстроювань, згасань і неідеальностей системи y_i .

З реалізації $u(t)$ адитивної суміші сигналу $S(t)$ з гауссовим білим шумом $n(t)$ необхідно оцінити вимірювану частоту.

Таким чином,

$$u(t) = \mu S(t, \omega_D, \delta t, y_1, \dots, y_n) + n(t),$$

де μ – мультиплікативний множник.

Для реальних умов роботи вимірювальної системи з активною відповіддю статистичну модель ФМ ПВП сигналу можна описати реальною частиною виразу

$$S(t) = \mu S_0 I_T(t - \delta t) \sum_{k=1}^N u_\tau [t - (k-1)\tau_s - \delta t] \cos[(\omega_0 - \omega_D)(t - \delta t) + a_k \pi + \psi_1]. \quad (4.3)$$

Квазіоптимальний приймач формує вихідний ефект (3.105), (3.106) [43], однозначно пов'язаний з автокореляційною функцією сигналу.

Енергія сигналу з урахуванням (4.3) визначається як

$$E(f_D, \mu, \delta t) = \int_0^T S^2(t) dt = \frac{1}{2} \mu^2 S_0^2 T \varphi_1(a), \quad (4.4)$$

$$\text{де } \varphi_1(a) = 1 - \frac{2}{a} + \frac{4}{a} e^{-a} - \frac{2}{a} e^{-2a}. \quad (4.5)$$

З огляду на (4.4) і (3.106), а також незалежність $\mu, \delta t, f_D$ вираз (4.4) можна подати у вигляді

$$E(\bar{X}) = E_0 \prod_i^{n_i} X_{Ei}, \quad (4.6)$$

де X_{Ei} – монотонні функції технічних параметрів y_i , що впливають на енергію сигналу й залежать від його обробки, $X_{Ei} = X_{Ei}(y_i)$.

Звідси в припущенні, що апіорна щільність імовірності $P(f_D)$ (цілевказівки) розподілена рівномірно у великому діапазоні, що значно перевищує смугу ФАПЧ, дисперсія похибки вимірювання частоти Доплера f_D визначається [4] як

$$D[\Delta f_D] = \frac{2\beta_a^2 \sigma_a^2}{q} = \frac{1}{\prod_{i=1}^{n_i} X_i(y_i)}, \quad (4.7)$$

Таким чином, точність вимірювань частоти f_D реального приймача завжди гірше, ніж оптимального, за рахунок впливу багатьох факторів. Дійсно, навіть якщо приймач передбачає

ідеальне усереднення за випадковими μ , δt , f_D , завжди будуть існувати інші неідеальності схеми, конструкції й випадкові збурювання в блоках РТС, яких у реальних приймачах значно більше, ніж у прийнятій моделі.

Оскільки $D[\Delta f_D]$ є випадковою величиною, що залежить від випадкових μ і δt , то як цільову функцію для стохастичного програмування (4.7) можна прийняти лише «у середньому», тобто використовуючи числові характеристики (перші початкові моменти) для μ і $|\delta t|$ з підрозд. 3.1.

При цьому

$$\bar{\mu}^{\theta_i} = \prod_{i=1}^8 \overline{G_i^2(\theta_i)^{\theta_i}} = \prod_{i=1}^8 \frac{\alpha_i}{\alpha_i + 2} = \prod_{i=1}^8 X_{Pi}, \quad (4.8)$$

де X_{Pi} – «фазові» параметри, що являють собою монотонні функції від технічних параметрів y_i (у цьому випадку α_i), які впливають на потужність сигналу.

Аналогічні співмножники у вигляді функцій, що відображені у формулах (3.1) - (3.35), наприклад,

$$\left(1 - \frac{|\delta t|}{\tau_e}\right) = 1 - \sqrt{\frac{2}{\pi}} \frac{\sigma}{\tau_e}$$

Потужність сигналу P можна визначити з рівняння радіолокації з активною відповіддю:

$$P = \frac{P_{npD1} \lambda_{3-B}^4 K \mu}{(4\pi)^4 R^4} \frac{e^2}{m^2} \chi_1 \chi_2 \prod_{i=1}^4 (g_i \varepsilon_i L_{xi}), \quad (4.9)$$

де

$$K = \prod_{i=1}^4 (L_{3i} L_{\delta si}) K_{yвч1} K_{yвч2} K_{npD1} K_{ynч1} K_{ynч2} \prod_{i=1}^3 \alpha_{смк} L_0 L_{\partial p} L_{\partial 2} L_{nep1}$$

Спектральна щільність вихідного шуму N_0 у припущенні, що на відстанях, більших 100 км, ретрансльований шум значно менше власного шуму наземного приймача і що коефіцієнт підсилення підсилювача високої частоти $K_{увч} \gg 1$ визначається як

$$N_0 = Ш_{увч1} (\lambda_0 T_{оач}) K_{увч1} K_{унч1} L_{з1} L_{см1} L_{пер1}, \quad (4.10)$$

де K_0 – постійна Больцмана.

З огляду на (4.8)- (4.10), дисперсію похибки вимірювань (4.7) можна підставити в повному вигляді

$$D[\Delta f_\delta] = \frac{6 \cdot 4^4 \pi^2 R_{max}^4}{2 \lambda_{3-\delta}^2 \lambda_{\delta-3}^2} \cdot \frac{K_0 T_{оач}}{T^3 P_{прд1} \prod_{i=1}^6 \overline{G_i(\theta_i)} \chi_1 \chi_2 \prod_{k=1}^4 (g_k \varepsilon_k) [1 - \varphi(a)]} \times \quad (4.11)$$

$$\frac{Ш_{увч1}}{\left(1 - \frac{|\delta t|}{\tau_3}\right) f(T\Pi_{фнч}) \left(1 - \frac{\overline{\psi}_0}{2}\right) \left(1 - \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma}{\tau_3}\right) \prod_{\ell=2}^4 L_{3\ell} K_{увч2} L_{см2} L_{см3} K_{прд2} K_{унч2} L_0 L_{g1} L_{g2}}$$

Очевидно, що (4.11) можна записати простіше:

$$D[\Delta f_\delta] = \frac{const}{\prod_{j=1}^{n1} X_j}$$

де, як і раніше, X_j – монотонні функції від своїх технічних параметрів, до числа яких відноситься $1/Ш_{увч1}$ і $1/T_{сач}$

$$const = const_1 \frac{R_{max}^4}{\lambda_{3-\delta}^2 \lambda_{\delta-3}^2}$$

При цьому

$$1 \leq X_{3j} \leq X_{3j\text{пред}},$$

де $X_{3j \text{ пред}}$ – граничні значення параметрів, що описують неідеальності і розстроювання.

Якщо використовувати також наявну статистику для кожного параметру нестабільності, наприклад, методом найменших квадратів, то одержимо обмеження до цільової функції (4.2) з урахуванням (4.11) у вигляді багатомірної кривої середньоквадратичної регресії

$$C = C(X_1, \dots, X_n).$$

Тоді задачу можна сформулювати в загальних позначеннях:

$$\min D \left[\frac{\Delta V_r}{V_r} \right] = \min \left[\frac{\text{const}}{\prod_{j=1}^{n_1} X_j} + \sum_{i=1}^{n_2} X_i \right] \quad (4.12)$$

при $C \leq C(X_1, \dots, X_n),$

$$X_{j \text{ пред}} \geq X_j \geq 0,$$

$$X_i \geq X_{i \text{ пред}} \quad 1 \geq X_{\varepsilon j} \geq X_{\varepsilon j \text{ пред}}.$$

Таку задачу можна розв'язати будь-яким відомим способом. Однак отриману систему функціональних рівнянь або зовсім не вдасться розв'язати в аналітичному вигляді, або розв'язання буде супроводжуватися громіздкими обчисленнями і мати числовий вигляд, особливо при великих n_1 і n_2 . Отже, необхідно використати деякі властивості вартості $C(X_1, \dots, X_N)$... Дійсно, з адитивності C випливає, що не вся вартість C залежить від X_i , а лише деяка частина, що створює якісний параметр X_i . Наприклад, випромінювану потужність наземної станції створює в основному передавач із блоками живлення.

Тому залежність $C(X_{\text{пред}})$ визначається залежністю вартості передавача $C_{\text{пред}}$ від випромінюваної потужності.

У найгіршому випадку, коли вартість блоків не розподіляється на окремі залежності від фазових параметрів, вона

все-таки буде залежати від невеликого числа параметрів. Наприклад, вартість антени залежить від ККД, КСД і коефіцієнта використання площі антени й, отже, є в цьому випадку функцією трьох змінних. Приймач при деяких спрощеннях можна розділити на субблоки, вартість кожного з яких залежить лише від одного параметра.

У загальному випадку

$$C = \sum_k C_k(X_k) + \sum_\ell C_\ell(X_{\ell 1}, X_{\ell 2}) + \sum_m C_m(X_{m1}, X_{m2}, X_{m3}) + \dots \quad (4.13)$$

Хоча вираз для C (4.13) став набагато конкретнішим на вигляд і функції $C_k(X_k)$, $C_\ell(X_{\ell 1}, X_{\ell 2})$ і $C_m(X_{m1}, X_{m2}, X_{m3})$, як правило, монотонні, але труднощі при формалізації системи обмежень, про які говорилося, не виключені.

Тому, щоб уникнути числових методів опуклого програмування й одержати аналітичний розв'язок, що важливо для розв'язання цілого класу задач, лінеаризуємо C_k , C_l і C_m . Лінеаризація тим більше виправдана, чим більша величина дисперсії кривої $C(\bar{X})$.

Поділ C_i і C_j на групи C_k , C_l і C_m в (4.13) умовний, оскільки якщо з якої-небудь причини виключити з оптимізації один параметр із C_m , то C_m варто подати в C_l і т.п.

Відповідно до (4.13) $\sum_j C_j$ можна записати так:

$$\begin{aligned} & \sum_{j=1}^p [C_{oj}(X_{oj}) + C_{oj}(X_{oj})(X_j - X_{oj})] \times \sum_{j=1}^q C_{oj}(X_{oj1}, X_{oj2}) + \\ & + \sum_{j=1}^q \frac{\partial C_{oj}(X_{oj1}, X_{oj2})}{\partial X_{oj1}} (X_{j1} - X_{oj1}) + \sum_{j=1}^q \frac{\partial C_{oj}(X_{oj1}, X_{oj2}, X_{oj3})}{\partial X_{oj3}} (X_{j1} - X_{oj1}) + \\ & + \sum_{j=1}^q \frac{\partial C_{oj}(X_{oj1}, X_{oj2}, X_{oj3})}{\partial X_{oj2}} (X_{j2} - X_{oj2}) + \sum_{j=1}^r \frac{\partial C_{oj}(X_{oj1}, X_{oj2}, X_{oj3})}{\partial X_{oj3}} (X_{j3} - X_{oj3}) + \dots, \end{aligned} \quad (4.14)$$

де p , $2q$, $3r$ – число змінних C_k , C_l і C_m , тобто

$$p + 2q + 3r = n_1.$$

У виразі (4.14) можна згрупувати й перенумерувати постійні члени й члени з незалежними змінними:

$$\sum_{j=1}^{n_1+q_1+2r_1} C_{oi} + \sum_{j=1}^{n_1} C'_{oj} (X_j - X_{oj}).$$

Аналогічно

$$\sum_{i=1}^{n_1-q_1-2r_1} C_{oi} + \sum_{i=1}^{n_2} |C'_{oj}| (X_{oi} - X_i),$$

де $C_{oi} < 0$ і

$$\begin{aligned} n_1 &= p + 2q + 3r_1, & n_1 - q - 2r &= p + q + r; \\ n_2 &= p_1 + 2q_1 + 3r_1, & n_2 - q - 2r &= p_1 + q_1 + r_1. \end{aligned}$$

Оскільки статистика вартості визначається за технічними параметрами, її необхідно перераховувати для фазових параметрів. Розкладання в ряд (лінеаризація) виконується в межах такої точки (X_{o1}, \dots, X_{oN}) , що лежить усередині часто використовуваної ділянки. Такий вибір вектора фазових параметрів X_0 відіграє роль апіорних даних про стаціонарну точку X_{opt} , оскільки завдяки інтуїтивній, але оптимізуючій ролі інженера, її розташування в околі використовуваних параметрів однієї із сучасних систем того ж класу найбільш ймовірно.

Таким чином, завдяки лінеаризації обмежень, задача спрощується і з'являється можливість одержати розв'язання в замкнутому вигляді. Потрібно знайти

$$\min D \left[\frac{\Delta V_r}{V_r} \right] = \min_{\left\{ \begin{matrix} X_i \\ X_j \end{matrix} \right\}} \left(\frac{const}{\prod_{j=1}^{n_1} X_j} + \sum_{i=1}^{n_2} X_i^2 \right) + D_c, \quad (4.15)$$

при обмеженнях на параметри, отриманих з (4.14):

$$\sum_{k=1}^{n_1+n_2} C_{ok} + \sum_{j=1}^{n_1} C'_{oj} (X_j - X_{oj}) + \sum_{i=1}^{n_2} |C'_{oi}| (X_{oi} - X_i) = C, \quad (4.16)$$

$$X_{j \text{ nред}} \geq X_j \geq 0, \quad 1 \gg X_i < X_{i \text{ nред}},$$

де $N_1 = n_1 + n_2 - (q + 2r) - (q_1 + 2r_1)$.

Хоча розв'язання задачі (4.15), (4.16) опуклого програмування можна знайти методом покоординатного спуску або одним із градієнтних методів, розв'язати методом виродженого динамічного або блокового програмування виявляється простіше й дозволяє при наявності обмежень визначити загальний підхід до подібних сепарабельних задач.

Таким чином, при постановці розглянутої задачі про перерозподіл зусиль блоків системи в процесі вимірювань радіальної швидкості були зроблені такі припущення й обмеження:

- відношення сигнал/шум $q \gg 1$, через що типову структуру реального вимірювача, що використовує ШШС, можна вважати квазіоптимальною, що забезпечує одержання ефективної оцінки;

- враховані основні параметри системи, що характеризують якість блоків і беруть участь в оптимізації. Інші параметри, наприклад, надійність, вага, фіксовані й дорівнюють часто застосовуваним значенням. Відповідно статистика набирається з вибірок при тих же фіксованих параметрах;

- вимірювання здійснюються в умовах стійкого автосупроводу за кутами і за частотою;

- тому схеми пошуку сигналу повинні оптимізуватись за іншими критеріями;

- у всіх наближеннях (3.1) - (3.35) і (4.15) залишаються лише складові першого порядку малості;

- лінійні обмеження (4.16) справедливі лише на деякій ділянці навколо вектора очікуваної оптимальності, апріорного \bar{X}_0 . Довжина ділянки пропорційна радіусу кривизни $C_i(X_i)$;

- дисперсія похибки вимірювання швидкості служить цільовою функцією для оптимізації РТС на максимальній дальності.

У результаті можна зробити нижчеподані висновки:

1. При формалізації задачі отримані опукла цільова функція і лінеаризовані обмеження на параметрі.

2. Цільова функція придатна для оцінки першого наближення точності проєктованих реальних вимірювальних систем того ж класу.

3. При деяких спрощеннях, пов'язаних з використанням ШСС, за цільовою функцією можна оцінити точність систем, що працюють із простими сигналами, якщо взяти до уваги ефекти, пов'язані з втратами енергії або потужності сигналу при обробці.

Нормально постановка задачі параметричного сигналу справедлива для будь-яких вимірювальних та інформаційних систем, у тому числі сумісних, оптимізація яких викладена в розд. 5.

4.5. Визначення мінімуму цільової функції

Скористаємося методом динамічного програмування для розв'язання задачі (4.15), (4.16). Визначимо мінімум першого доданка в припущенні, що обмежень на «фазові» параметри X_j немає, крім $X_j > 0$:

$$F_{n_1}(C_1) = \min_{\{X_j\}} \frac{1}{\prod_{j=1}^{n_1} X_j}$$

при

$$\sum_{j=1}^{n_1} C'_{oj} X_j = C_{\partial 1},$$

$$\text{де } \Delta C_{n_1} = C_1 - \sum_{j=1}^{n_1 - q_1 - 2r_1} C_{oj}; \quad C_{\partial 1} = \Delta C_{n_1} + \sum_{j=1}^{n_1} C'_{oj} X_{oj};$$

$$X_j = 0.$$

Якщо скористатися результатом оптимізації для логарифма $F_{n_1}(C)$, для якого через його монотонність результати не змінюються, то можна одержати

$$F_{n_1}(C) = \frac{\prod_{j=1}^n C'_{oj}}{\left(\frac{C_{e1}}{n_1}\right)^{n_1}}$$

при оптимальному поводженні

$$X_{n_1} = \frac{C_{el}}{n_1 C'_{on_1}}. \quad (4.17)$$

Доведемо справедливість (4.17).

Зробимо перший крок оптимізації.

Тоді

$$F_1(C_{el}) = \frac{C'_{o1}}{C_{el}}$$

при єдиному поводженні $X_{1min} = \frac{C_{el}}{C'_{o1}}$.

Для другого кроку

$$F_2(C_e) = \min \frac{C'_{o1}}{C_{el} - C'_{o2} X_2}$$

при

$$X_{2min} = \frac{C_{el}}{2C'_{o2}}.$$

Неважко побачити, що

$$F_2(C_e) = \min_{\{X_2\}} \frac{1}{X_2} F_1(C_e - C'_{o2} X_2).$$

Звідси можна скласти рекурентне рівняння

$$F_N(C_e) = \min_{\{X_N\}} \frac{1}{X_N} F_{n-1}(C_e - C'_{oN} X_N). \quad (4.18)$$

Доведення (4.18) проведемо покоординатним способом знаходження мінімуму.

Доведемо справедливість розв'язати методом математичної індукції. Розв'язання справедливе для $N=1,2$. Виходить, якщо

воно справедливе для $N=n_1$, то воно повинне бути справедливе й для $N=n_1+1$.

Дійсно, з (4.18) одержимо

$$F_{n_1+1}(C_{e1}) = \min_{\{X_{n_1+1}\}} \frac{1}{X_{n_1+1}} \frac{\prod_{j=1}^{n_1} C'_{oj}}{\left(\frac{C_{e1} - C'_{o(n_1+1)} X_{n_1+1}}{N} \right)^N} = \frac{\prod_{j=1}^{n_1+1} C'_{oj}}{\left(\frac{C_{e1}}{n_1+1} \right)^{n_1+1}}$$

при

$$X_{n_1+1} = \frac{C_{e1}}{(n_1+1)C'_{o(n_1+1)}}.$$

Оскільки розв'язання справедливе й для $N=n_1+1$, то воно справедливе й для довільного N .

Визначимо мінімум суми доданків у припущенні лише $X_i \geq 0$:

$$F_{n_2}(\Delta C_2) = \min \sum_{i=1}^{n_2} X_i^2$$

при

$$\sum_{i=1}^{n_2} [C'_{oi}(X_{oi} - X_i) + C_{oi}] \leq C_2$$

або

$$\sum_{i=1}^{n_2} C'_{oi} X_i \geq \sum_{i=1}^{n_2} C'_{oi} X_{oi} - \Delta C_{n_2} = C_{e2},$$

$$X_i \geq 0,$$

$$\text{де } C_{e2} = \sum_{i=1}^{n_2 - q_2 - 2r_2} C_{oi} + \sum_{i=1}^{n_2} C'_{oi} X_{oi} - C_2, \quad \Delta C_{n_2} = C_2 - \sum_{i=1}^{n_2 - q_2 - 2r_2} C_{oi}.$$

На першому кроці одержимо

$$F_1(C_{e2}) = \left(\frac{C_{e2}}{C_{o1}} \right)^2$$

при

$$X_{min1} = \frac{C_{e2}}{C'_{o1}}.$$

На другому кроці одержимо рекурентне рівняння

$$F_2(C_{e2}) = \min_{\{X_2\}} \left[X_r^2 + \left(\frac{C_{e2} - C'_{o2} X_2}{C'_{o1}} \right)^2 \right] = \min_{\{X_2\}} [X_r^2 + F_1(C_{e2} - C'_{o2} X_2)]$$

при

$$C'_{o1} X_1 + C'_{o2} X_2 = C_{e2}.$$

Методом індукції можна довести справедливість цього розв'язання для будь-якого n_2 :

$$F_{n_2}(C_{e2}) = \min [X_{n_2}^2 + F_{n_2+1}(C_{e2} - C'_{on_2} X_{n_2})].$$

Тому для другого кроку одержимо

$$F_{n_2}(C_{e2}) = \frac{C_{e2}^2}{(C'_{o1})^2 + (C'_{o2})^2}$$

при

$$X_{2min} = \frac{C_{e2} C'_{o2}}{(C'_{o1})^2 + (C'_{o2})^2},$$

а для третього – відповідно

$$F_3(C_{e2}) = \frac{C_{e2}^2}{(C'_{on1})^2 + (C'_{on2})^2 + (C'_{on3})^2}$$

при

$$X_{3min} = \frac{C_{e2} C'_{o3}}{(C'_{o1})^2 + (C'_{o2})^2 + (C'_{o3})^2}.$$

З огляду на співвідношення F_2 і F_3 , припустимо, а потім і доведемо за індукцією, аналогічно доведенню (4.17), що

$$F_{n_2}(C_{e2}) = \frac{C_{e2}^2}{\sum_{i=1}^{n_2} (C'_{oi})^2} \quad (4.19)$$

при

$$X_{n_2 \min} = \frac{C_{e2} C'_{on_2}}{\sum_{i=1}^{n_2} (C'_{oi})^2}.$$

Потім визначимо цільову функцію (4.15) у вигляді:

$$F(\Delta C) = \min_{\left\{ \begin{array}{l} \Delta C_1 \\ \Delta C_2 \end{array} \right\}} [F_{n_1}(\Delta C_{n_1}) + F_{n_2}(\Delta C_{n_2})]$$

при

$$\Delta C_{n_1} + \Delta C_{n_2} \leq \Delta C.$$

Відповідно до отриманих розв'язань (4.17) і (4.19) вираз для F запишемо у вигляді

$$F(\Delta C) = \min_{\left\{ \begin{array}{l} \Delta C_1 \\ \Delta C_2 \end{array} \right\}} \left[\frac{A}{(\Delta C_{n_1} + B)^{n_1}} + \frac{(D - \Delta C_{n_2})^2}{\varepsilon} \right] + D_c$$

При

$$\Delta C_{n_1} + \Delta C_{n_2} \leq \Delta C,$$

де $A = \text{const } n_1^{n_1} \prod_{j=1}^{n_1} C'_{oj}$; $B = \sum_{j=1}^{n_1} C'_{oj} X_{oj}$.

Методом Ньютона – Рафсона (дотичних) [15] оптимальне поводження на останньому кроці $\Delta C_{2 \text{opt}}$ визначимо з рівняння

$$\frac{A \varepsilon^{\frac{n_1}{2}}}{[B + \Delta C - \Delta C_{n_2 \text{opt}}]^{n_1+1}} = D - \Delta C_{n_2 \text{opt}}. \quad (4.20)$$

Для $\Delta C_{n_2 \text{opt}}$ з формули (4.20) одержимо рекурентну формулу

$$\Delta C_{n_2(k)} = D - \frac{[\Delta C + B - \Delta C_{n_2(k-1)}] A \varepsilon \frac{n_1}{2}}{\frac{n_1}{2} A \varepsilon (n+1) + [\Delta C + B - \Delta C_{n_2(k-1)}]^{n_1+2}}, \quad (4.21)$$

при

$$\Delta C_{n_2(1)} \in [0, D],$$

де k – номер кроку ітерації.

На рис. 4.3 показана процедура послідовного визначення $\Delta C_{n_2 opt}$.

Звичайно

$$(\Delta C + B) \gg \Delta C \quad (\Delta C + B) \gg \frac{n_1}{2} A \varepsilon (n_1 + 1).$$

Тоді

$$\Delta C_{n_2 opt} \approx D - \frac{A \varepsilon \frac{n_1}{2}}{(\Delta C + B)^{n_1+1}}, \quad (4.22)$$

$$F(\Delta C) \approx \frac{A}{(\Delta C + B - D)^{n_1}} + \varepsilon \frac{\left(A \frac{n_1}{2}\right)^2}{(\Delta C + B)^{2(n_1+1)}} + D_c$$

і

$$\Delta C_{n_1 opt} = \Delta C - \Delta C_{n_2 opt}.$$

Параметри $X_{j opt}$ і $X_{i opt}$ визначаються з (4.17) і (4.19), а оптимальні технічні параметри системи – з рівняння

$$y_{j opt} = y_{j opt}(X_{j opt}).$$

Оптимальні вартості блоків знаходяться з рівнянь одно- або багатовимірних лінійних обмежень.

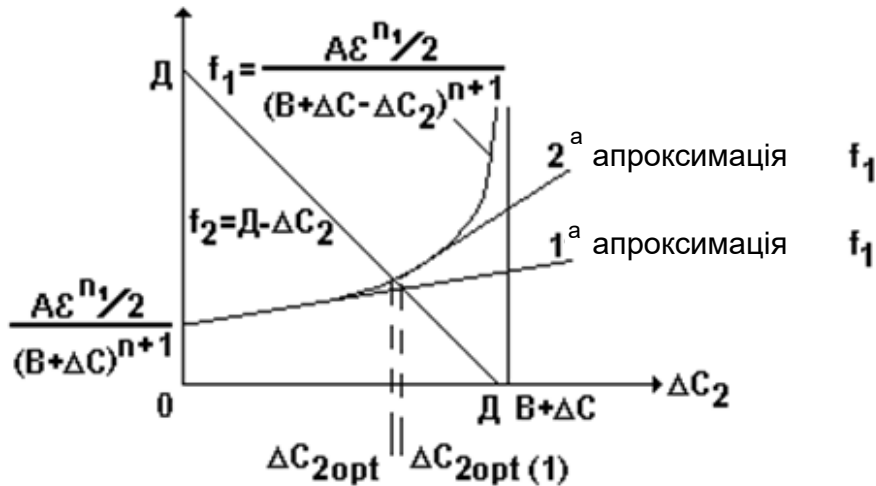


Рис. 4.3. Структура оптимального розв'язання

Таким чином, для області хорошої апроксимації кривих регресії лінійними функціями отриманий оптимальний розподіл вартостей блоків і оптимальні параметри, при яких система забезпечує мінімум дисперсії похибки вимірювань радіальної швидкості при заданій загальній структурі системи (рис. 4.2) на максимальній дальності й при обмежених асигнуваннях.

З розвитком відповідної техніки й виробництва вартість блоків зменшується, а в області граничних значень параметрів y_{imax} і y_{imin} зменшується незначно. Тому зменшення крутості обмежень C_{oi}^l при незмінних параметрах, близьких до граничних, як і самі величини y_{imax} і y_{imin} , відбивають природу обмежень й економіко-технічний стан даного виробництва.

Зі співвідношення для цільової функції F видно, що між параметрами A і E , U і D існують деякі протиріччя, що відбиваються на величині оптимальної точності.

З постановки задачі (4.22) можна побачити шляхи подальшого вдосконалювання даної апаратури і технології виробництва.

Крім того, з постановки задачі (4.22) випливає, що оскільки межа вимірювань радіальної швидкості визначається обмеженням знанням швидкості поширення радіохвиль D_c , то існує межа і для асигнувань. Якщо потрібно одержати дисперсію $F(\Delta C)$ у α більшу, ніж D_c , то з (4.22) одержимо значення максимальних асигнувань на оптимальну систему

$$\Delta C_{\max} = B + \sqrt{\frac{A}{(\alpha-1)D_c}} + \frac{\frac{n_1^2}{4} \varepsilon A^{\frac{2n_1-1}{n_1}}}{[(\alpha-1)D_c]^{\frac{n_1+1}{n_1}} n_1 (D + \sqrt{\frac{A}{(\alpha-1)D_c}})^{2(n_1+1)}}. \quad (4.23)$$

Як і результати (4.17), (4.19) формула (4.23) справедлива для оптимального розв'язання задачі (4.22), досить близького до вектора X_0 .

4.6. Оптимальні параметри систем

З огляду на отримані результати, запишемо в прийнятих позначеннях оптимальні параметри й вартості системи у вигляді, зручному для розрахунків:

$$\Delta C_{2(k)} = D - \frac{[\Delta C + B - \Delta C_{2(k-1)}] A \varepsilon^{\frac{n_1}{2}}}{\frac{n_1}{2} A \varepsilon (n_1 + 1) + [\Delta C + B - \Delta C_{2(k-1)}]^{n_1+2}}$$

при $(\Delta C + B) \gg \Delta_{,,2}$

$$(\Delta_{,,} + B) \gg \frac{n_1}{2} A \varepsilon (n_1 + 1);$$

$$\Delta C_{2opt} \cong D - \frac{A \varepsilon^{\frac{n_1}{2}}}{(\Delta C + B)^{n_1+1}}, \Delta C_{1opt} = \Delta C - \Delta C_{2opt}.$$

Згідно з (4.17) і (4.19) визначимо «фазові» параметри X_j

$$X_{kopt} = \frac{C_1 - \sum_j C_{0j} + \sum_j C'_{0j} X_{0j}}{nC'_{0k}}.$$

Звідси можна одержати також усі технічні параметри

$$y_{jopt} = y_{jopt}(X_{jopt});$$

а) відношення α_j ширини діаграми спрямованості антени за потужністю в площині θ_x до величини дисперсії відповідної похибки кутового супроводу

$$\alpha_{jopt} = \frac{2X_{jopt}^2}{1 - X_{jopt}^2};$$

б) тривалість сигналу T_{opt} , яка дорівнює часу інтегрування

$$T_{opt} = X_{jopt}^{\frac{1}{3}};$$

в) відношення a_{opt} смуги пропускання високочастотного тракту по половині рівня максимуму до ширини спектра ШШС «по нулях», що впливає на криві відлику $a_{opt}(X_{jopt})$ (рис. 3.3).

$$a_{opt} = \frac{\ln 2 + B}{1 - X_{jopt}},$$

де $B = 1 - \rho$;

г) смуга системи фазового автопідстроювання частоти

$$n_{\text{ТМЄ} \cdot opt} = \frac{\ln \frac{1}{1 - X_{jopt}}}{T_0},$$

де T_0 – фіксований час інтегрування, що дорівнює очікуваному значенню $T = T_0$;

д) абсолютне значення результуючої похибки $|\delta t|_{opt}$ каналу синхронізації при відомому законі руху

$$|\delta t|_{opt} = \tau_e (1 - X_{jopt});$$

е) середньоквадратична похибка σ_{aopt} каналу синхронізації

$$\sigma_{\delta opt} = \sqrt{\frac{2\pi}{2}} \tau_e (1 - X_{jopt});$$

ж) коефіцієнт шуму $Ш_{yвч1opt}$ підсилувача високої частоти наземного приймача

$$Ш_{yвч1opt} = \frac{1}{X_{jopt}},$$

з) сумарна шумова температура $T_{0ачopt}$ наземної приймальної антени

$$T_{0ач opt} = 1/X_{jopt},$$

и) інші параметри, що беруть участь в оптимізації

$$y_{jopt} = X_{jopt}.$$

«Фазові» параметри X_{jopt} визначаються як

$$X_{jopt} = \frac{C'_{0i}}{\sum_{i=1}^{n2} (C'_{0i})^2} \left(\sum_i C_{0i} + \sum_{i=1}^{n2} C'_{0i} X_{0i} - C \right).$$

При цьому

$$\frac{\Delta f_{iopt}}{f_0} = 2 \sqrt{\frac{m}{l}} X_{iopt}.$$

Оптимальні вартості системи слід визначити з виразу

$$C_k = C_{0k} + C^l_{0k} (X_{kopt} - X_{k0})$$

або

$$C_k = C_{0k} + C^l_{0k} (X_{kopt} - X_{k01}) + C^l_{0j2} (X_{k2opt} - X_{k02}) \quad \text{і т. д.}$$

Визначимо область значень «фазових» параметрів, для яких розв'язання (4.17) – (4.19) вважаються оптимальними. Ця область

$$X_{i1} \cong X_{i0} + \sqrt{\frac{2h}{C''_{i0n}}} \frac{1}{[1+(C'_{in})^2]^{\frac{1}{4}}}. \quad (4.24)$$

Якщо характеризувати близькість $C_{in}(X_i)$ і лінійних обмежень параметром h , то формула (4.24) дає припустиму границю «фазового» параметра X_{i1} , у межах якої по обидві сторони від X_{i0} лінійну апроксимацію можна вважати задовільною й, отже, розв'язання практично оптимальним.

Можна користуватись також методом дотичних Ньютона.

Однак на практиці набагато зручніше оцінювати близькість функції C_{in} і її апроксимації (4.25) параметром $\Delta C_{an} = AC$, який, як і h , той самий для всіх i та j

$$X_{i1} \cong X_{oi} + \sqrt{\frac{2\Delta C_{an}}{C''_{oi}}}. \quad (4.25)$$

Таким чином, якщо розв'язання (багатовимірний вектор параметрів X_{opt}) попадає в інтервал $[2X_{i0}-X_{i1}, X_{i1}]$, де X_{i1} знаходимо з (4.24) або (4.25), то розв'язання є досить точним і визначається формулами (4.17) і (4.10). Якщо вихідні криві регресії лінійні без лінеаризації, то розв'язки точні завжди.

Однак умови фізичної реалізації параметрів накладають обмеження на параметри й на оптимальне розв'язання. Тому якщо розв'язок попадає в інтервал хорошої апроксимації лінійними функціями, але перевищує граничне значення, то треба при цьому обмежитися граничним значенням параметра.

4.7. Приклад оптимізації системи за трьома параметрами, для яких відомі обмеження за вартістю

Уже відомі деякі функції вартості блоків системи залежно від відповідних параметрів [18]

$$C^{[1]} = K_1 \sqrt{X^{[1]}};$$

$$C^{[2]} = K_2 + K_3 (X^{[2]})^2;$$

$$C^{[3]} = \frac{K_4}{(X^{[3]})^{K_5}},$$

де $C^{[1]}$ і $X^{[1]}$, $C^{[2]}$ і $X^{[2]}$, $C^{[3]}$ і $X^{[3]}$ – відповідно вартість і потужність передавача, вартість і КСД антени, вартість і відносна короточасна нестабільність стандартів частоти;

K_1, \dots, K_5 – числові коефіцієнти, що залежать від рівня розвитку техніки й виробництва відповідних блоків.

Припустимо, що оптимізується вимірювач радіальної швидкості об'єкта. Вимірювач має відому структуру (рис. 4.2) і відому цільову функцію (4.11). Відомими також вважаються обмеження на параметри X_i і X_j за умовами їх фізичної реалізованості

$$0 < X_i < X_{i\text{пред}}, X_{j\text{пред}} < X_j < X_{j\text{max}}.$$

Нехай, наприклад, відомі коефіцієнти $K_1=12$; $K_2 = 5$; $K_3=6 \cdot 10^{-4}$; $K_4=0,062$, $K_5=1/2$.

Первинні значення параметрів (потужності передавача $X_0^{[1]}$, КСД антени $X_0^{[2]}$ і відносна нестабільність частоти $X_0^{[3]}$) обрані так, що або відомі для системи такого ж класу, або кращі для застосування, на думку експертів, тощо.

$$X_0^{[1]}=100, X_0^{[2]}=400, X_0^{[3]}=10^{-6},$$

а вартість всієї проектованої частини системи $C_0=120$ (грош. од.). У цьому випадку $n_1=2$, $n_2=1$.

Користуючись формулами (4.17), (4.19), (4.21), (4.23) і дотримуючись алгоритмів та формул підрозд.4.6., можна визначити $X_{opt}^{[i,j]}$, $C_{opt}^{[i,j]}$, A , E , B_1 , B_2 , C_1 , C_2 ,.

Результати розрахунку оптимальних параметрів зведені в табл. 4.1.

Таблиця 4.1

X, C K	1	2	3	4	5
$X^{[1]}$	100	1463	2568	4248	4500
$X^{[2]}$	400	320	240	207	196
$X^{[3]}$	10^{-4}	$6,2 \cdot 10^{-6}$	$8,08 \cdot 10^{-7}$	$3,3 \cdot 10^{-7}$	$1,5 \cdot 10^{-6}$
$C^{[1]}$	14,1	54,5	71,5	92	95
$C^{[2]}$	101	66,2	39,8	30,8	28,1
$C^{[3]}$	6,2	0,4	1,19	0,832	0,567
$C^{*[1]}$	$7,05 \cdot 10^{-2}$	$1,83 \cdot 10^{-3}$	$1,39 \cdot 10^{-2}$	$1,08 \cdot 10^{-2}$	$1,05 \cdot 10^{-2}$
$C^{*[2]}$	0,48	0,384	0,295	0,248	0,235
$C^{*[3]}$	$1,58 \cdot 10^5$	$1,61 \cdot 10^3$	$3,7 \cdot 10^6$	$6,3 \cdot 10^5$	$5 \cdot 10^5$

З таблиці можна побачити, що вплив $X^{[3]}$ малий, оскільки маємо лише невеликі зміни вартості. Тому $X_{opt}^{[3]}$ доцільно брати на два порядки краще первинного значення, тобто $X_{opt}^{[3]} = 10^{-8}$, не змінивши істотно вартість системи.

Неважко побачити з таблиці, що отримані оптимальні параметри можуть істотно підвищити точність вимірювача, що проектується, при тих же асигнуваннях $C_0 = 120$.

Дійсно, похибка вимірювань зменшиться при цьому в E_ϕ разів, де

$$E_\phi = \frac{D_{\omega \partial 0}}{D \omega \partial opt} = \frac{\frac{const}{X_o^{[1]} X_o^{[2]}} + X_o^{[3]}}{\frac{const}{X_{opt}^{[1]} X_{opt}^{[2]}} + X_{opt}^{[3]}}$$

З огляду на те, що $X_{opt}^{[3]} \ll D_{\omega \partial opt}$, визначимо ефективність E_ϕ оптимізації $E_\phi = 41,2$.

Таким чином, похибку вимірювання параметра сигналу можна зменшити в 41 раз при проектуванні системи в порівнянні із системами, у яких були параметри $X_{o,i,j}$. Варто очікувати, що для більшого числа відомих залежностей вартості від параметрів ефективність оптимізації буде більшою.

З розд. 4 випливає, що для широкого класу задач можливі постановка й розв'язання задач у цілому в аналітичному вигляді. Рішення справедливе для будь-якого числа технічних параметрів і будь-яких техніко-економічних залежностей. Основна перевага даної методики оптимізації полягає в тому, що вона дозволяє узагальнити задачу для довільного числа критеріїв.

5. ОПТИМІЗАЦІЯ ТЕХНІЧНИХ ПАРАМЕТРІВ СУМІСНИХ ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ РАДІОТЕХНІЧНИХ СИСТЕМ

У розділі наведені ідеї постановки й розв'язання задач оптимізації складних сумісних РТС за умовними критеріями якості з урахуванням усіх даних про параметри й техніко-економічних, наприклад маркетингових, даних.

Завдання на прочитання

1. Переконатися, що для сумісних систем постановка загальних задач оптимізації опирається на вже викладену постановку й розв'язання однокритерійної задачі, що виконує лише одну функцію.

Результати такого розв'язання для підсистем, або для інформаційних і вимірювальних каналів, використовуються як блоки в сепарабельному програмуванні, тобто при функціонально розподілених змінних.

2. Розібратися з поняттям «суміщення». Навести відомі вам приклади суміщення апаратури й сигналів.

3. Розібратися з принципом побудови РТС із фазоманіпульованими сигналами.

4. Розібратися з принципом розв'язання задач блокового програмування.

5. Ознайомитися з новою задачею оптимізації самого суміщення.

6. Ознайомитися з методом геометричного програмування.

У розд. 5 поставлена й вирішена в загальному вигляді задача оптимізації технічних параметрів за умовними критеріями якості системи, що вимірює один параметр руху цілі або що має один інформаційний канал. Але для дистанційного керування об'єктами потрібний комплекс систем, що мають різні цілі функціонування: передачу команд, програм, вимірювання різних параметрів руху й ін. На практиці виявлено, що всі зазначені цілі функціонування можуть бути досягнуті в одній смузі частот і в одній (багатоцільовій) системі. **Сигнал, що використовується для багатьох застосувань, називається сумісним.**

Багатоцільове застосування апаратури називається апаратурним суміщенням. Сигнальне суміщення має на меті економію часу одержання різної інформації або (і) економію смуги частот, призначеної для її передачі. **Апаратурне суміщення має на меті економію асигнувань** на комплекс або на апаратурно суміщену систему. Найкращі показники якості, включаючи економічні й оперативні, – в апаратурно- і сигнально-суміщених систем.

5.1. Принципи суміщення інформаційно-вимірювальних систем

Сигнальне суміщення може бути послідовним у часі в одній смузі частот (наприклад послідовні складові частини сигналу) або паралельним, одночасним, з поділом за будь-яким параметром селекції (за частотою, затримкою, фазою, часовою структурою, поляризацією і т.п.).

Оскільки вимірювачами оцінюються радіолокаційні параметри одного сигналу, то одне сигнальне поле несе всю необхідну інформацію про цілі. Тому суміщувати потрібно вимірювальний сигнал з інформаційним. Інформаційний сигнал повинен бути цифровим. Кращим інформативним параметром визнано насамперед фазу (для протилежних сигналів), потім структуру ПВП (для ортогональних сигналів) та ін.

Для вимірювачів будь-якого параметра сигналу оптимальна обробка припускає формування його автокореляційної функції, оскільки забезпечується максимум відношення сигнал/шум на виході приймача при заданій апертурі дискримінатора.

Тому основними принципами суміщення є: використання високочастотної частини системи й формувача функції автокореляції (ФАК), використання єдиного формувача суміщеного сигналу, використання оптимальності суміщення, тобто визначення оптимального перетину у системі для поділу каналів (рис. 5.1). Розрахунку оптимального суміщення повинен передувати вибір оптимальної вимірювальної структури кожного каналу, оптимізація технічних параметрів повинна здійснюватися за умовним критерієм якості разом з оптимізацією параметрів суміщення.

5.2. Постановка задачі

Існуючі і перспективні РТС звичайно будуються таким чином, що більшість каналів обробки сигналів після високочастотної частини системи можна вважати незалежними. Виняток можуть становити суміщені РТС, що використовують широкосмугові шумоподібні сигнали. Однак при малих необхідних похибках синхронізації й відстеження частоти системою ФАПЧ вимірвальні канали й канали обробки інформації також можна вважати незалежними або описуваними окремими критеріями.

При постановці задачі використовуються припущення про те, що система є високоякісною, і застосовують послідовний складений сигнал. Це припущення є досить загальним, оскільки завжди можливе спрощення задачі для випадків простих сигналів. Наявність у структурі повного або часткового складу блоків визначається лише в результаті параметричного синтезу, що дає остаточну відповідь доцільності застосування кожного блока і його параметрів. Щодо цього параметричний синтез є тісно пов'язаним зі структурним. Використання в РТС послідовних оцінок сигналів припускає найзагальнішу структуру, що включає пристрої формування й обробки складних сигналів. При цьому характеристики розстроювань, неідеальностей і збурювань у цих пристроях входять до виразів для дисперсій похибок формально так само, як і у системах із простими сигналами. Тому формалізована задача може бути застосована для широкого класу сигналів. Взаємозв'язок параметричного синтезу РТС із синтезом сигналів проявляється на етапі розрахунків оптимальних технічних параметрів за загальними формулами, у які входять також залежності якісного показника від сигнальних параметрів.

Суміщена РТС із укрупненою структурою (рис. 5.1) має один високочастотний радіоканал, у якому транслюється один суміщений радіосигнал, що несе різну інформацію для всіх каналів його обробки. Можливі й допоміжні радіоканали, які також є загальними для інформації різного призначення. Тому структура й зв'язки якісних показників істотно не зміняться.

На рис. 5.1 подана можлива типова укрупнена структура оптимізованої системи однієї сторони радіолінії, що використовує кореляційний метод приймання ансамблю послідовних складених сигналів. Можна показати також, що наведена далі методика параметричного синтезу справедлива принципово для будь-якої структури з незалежними або слабко залежними каналами обробки інформації.

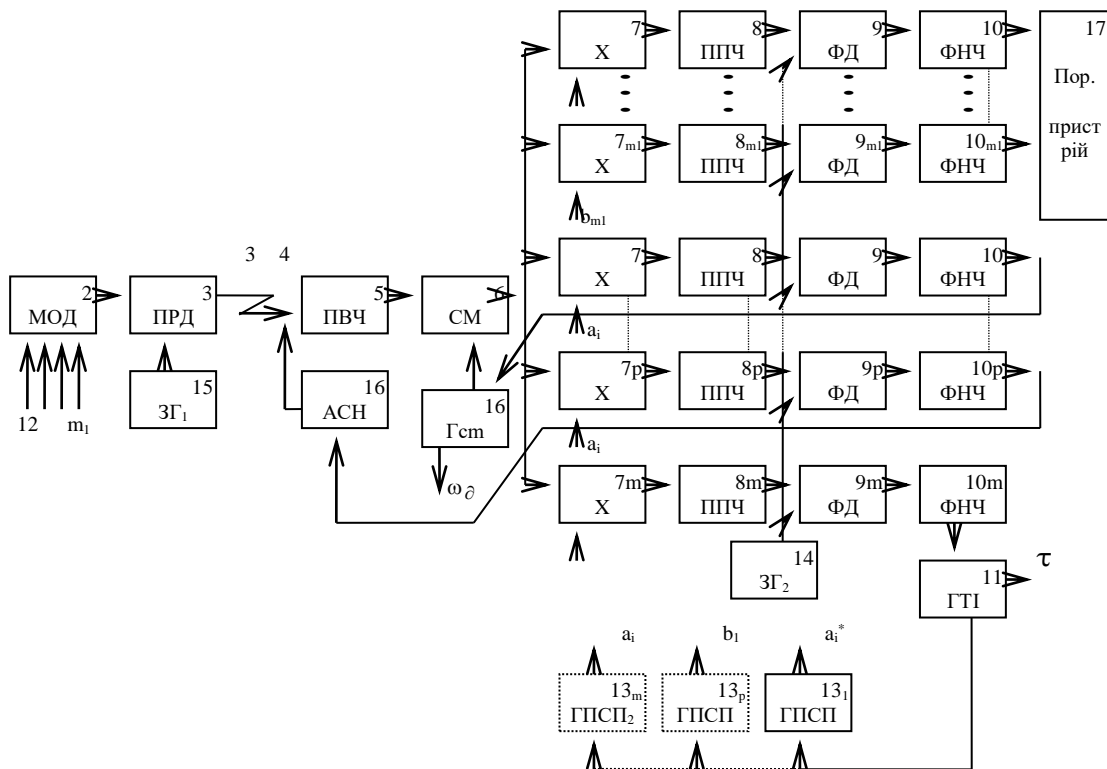


Рис. 5.1

На рис. 5.1 прийняті такі позначення:

1 - модулятор передавача для передачі послідовних складених сигналів; 2 - передавач; 3 - передавальна антена; 4 - приймальна антена; 5 - підсилювач радіочастоти; 6 - змішувач;

7₁...7_n - помножувачі у n каналах; 8₁...8_n - вузькосмугові підсилювачі проміжної частоти; 9₁...9_n - амплітудні та фазові детектори (АД, ФД); 10₁...10_n - фільтри нижніх частот; 11 - генератор тактових імпульсів; Г - гетеродин;

13₁...13_n - генератори опорних послідовностей; 14 - керований гетеродин; 15 - задавальні генератори 1 і 2; 16 - автоматична

система наведення за кутами; 17 – порівняльний пристрій, блок прийняття рішень.

На схемі зазначені укрупнені блоки вимірників параметрів сигналу за простором, частотою й затримкою, оптимальна структура яких буде визначена у відповідному завданні на наступному етапі. Структура суміщеної системи потрібна не тільки для формалізації критеріїв і для постановки й розв’язання задач параметричного синтезу, але й для наступного етапу сигнального синтезу, коли потрібні знання про конкретні схеми обробки сигналу. Система оптимізується таким чином, щоб знайти оптимальний розподіл асигнувань по блоках системи або оптимальні параметри з погляду мінімуму асигнувань на всю систему при обмеженнях на якісні показники системи й за умовами фізичної реалізованості параметрів.

Технічні параметри кількісно описують деяку якість відповідних систем блоків. Тому знайдені в результаті розв’язання задачі параметри повинні бути вихідними даними для детального розроблення принципової схеми й конструкції відповідних блоків або для вибору стандартних блоків з уніфікованого ряду при проектуванні. Кількість блоків описується такими технічними параметрами:

- антени – коефіцієнтами спрямованої й корисної дії, коефіцієнтами використання площі антени;
- фідерні пристрої – коефіцієнтами хвилі, що біжить;
- підсилювачі високої частоти – шириною смуги пропускання, коефіцієнтом підсилення і коефіцієнтом шуму;
- змішувачі – коефіцієнтом перетворення й коефіцієнтом шуму;
- помножувачі й інші пасивні елементи в ланцюзі – коефіцієнтами передачі й коефіцієнтами шуму;
- підсилювачі проміжної частоти – смугою пропускання, коефіцієнтом підсилення;
- фазові детектори – втратами післядетекторної обробки;
- фільтр нижніх частот – смугою пропускання (частотою зрізу);
- задавальні генератори і гетеродини – нестабільністю частоти й потужністю;

- передавач – потужністю, смугою пропускання й внесеною нестабільністю частоти й т.д.

Дисперсія похибки D_p вимірювання якого-небудь параметра λ обернено пропорційна відношенню сигнал/шум на виході приймача (на виході дискримінатора)

$$D_\lambda = \frac{\sigma_{a\lambda}^2}{q(X)}, \quad (5.1)$$

де $\sigma_{a\lambda}^2$ – дисперсія апіорного розподілу;

λ – оцінка параметра;

q – відношення сигнал/шум;

X – технічний параметр, що впливає на дисперсію похибки вимірювання p -го параметра сигналу.

У розд. 4 було показано, що відношення q енергії сигналу до спектральної щільності шуму на виході відповідного каналу приймача можна подати у вигляді добутку монотонних функцій технічних параметрів системи, розстроювань і неідеальностей приймача, а також різних збурювань

$$D_\lambda = \frac{A_\lambda}{n_p \prod_{i=1} X_{\lambda i}}, \quad (5.2)$$

де $X_{\lambda i}$ – монотонна функція розстроювань і неідеальностей приймача, а також різних збурювань;

A_λ – постійний коефіцієнт, що включає в себе всі параметри, що не оптимізуються;

n_λ – число параметрів, що оптимізуються у каналі.

Прямий внесок у похибку вимірювання параметра дають нестабільності еталонів, методичні адитивні похибки, невраховані ефекти, випадкові параметри за умовами роботи апаратури (наприклад, збої лічильника, параметрична

нестабільність) і ряд інших факторів. Тому для більшості вимірювальних каналів з урахуванням зазначених факторів

$$D_p = \frac{A_p}{\prod_{i=1}^{n_p} X_{\lambda i}} + \sum_{j=1}^{n_{p2}} X_{\lambda j}^2 + X_c^2, \quad (5.3)$$

де X_c^2 – дисперсія похибок за рахунок незнання умов поширення радіохвиль;

$X_{\lambda j}$ – середньоквадратична похибка за рахунок нестабільності еталонів та інших факторів.

Покажемо, що такий критерій, як імовірність безпомилкової передачі дискретної інформації, що передана в РТС, також можна подати у вигляді (5.2). У системі можливе використання ансамблю кодів з алфавітом m . Будемо припускати, і звичайно це завжди виконується, що коди, які застосовуються в системі, практично ортогональні. Якщо сигнали, наприклад когерентні, то можна використати таку формулу для середньої ймовірності похибки передачі інформації:

$$P_{ошcp} = \sqrt{m_1 - 1} \exp\left(-\frac{q_n}{2} - 1,4\right), \quad (5.4)$$

де q_n – відношення сигнал/шум на виході приймача.

Вважаючи середню ймовірність похибки передачі інформації менше припустимої, маємо обмеження на відношення сигнал/шум

$$\frac{1}{qn} \leq \frac{1}{2 \ln \frac{\sqrt{m_1 - 1}}{P_{ошдон}} - 2,8}. \quad (5.5)$$

Відповідно до того, що відношення сигнал/шум можна записати у факторизованому вигляді (у вигляді добутку функцій

технічних параметрів, параметрів розстроювань, збурювань і неідеальностей), ліву частину обмеження (5.5) можна подати у вигляді (5.2). Тоді можна використати для показника (5.5) позначення $D_{p\partial on}$

$$\frac{D_{p\partial on}}{P_{oш\partial on}} \leq \frac{1}{2 \ln \frac{\sqrt{m_1 - 1}}{2,8}}. \quad (5.6)$$

Аналогічні співвідношення можна одержати для будь-якої схеми обробки сигналів. Таким чином, для всіх каналів наведена нижче система нерівностей описує обмеження на технічні параметри й параметри сигналів та умов роботи з боку припустимих якісних показників. Крім того, існують обмеження за умовами, чи можна фізично виконати блоки, які описуються відповідними параметрами

$$D_p(\vec{X}_i, \vec{X}_j) \leq D_{p\partial on}, \quad (5.7)$$

$$0 < X_{pi} < X_{pi\text{пред}}, X_{pj\text{пред}} \leq X_{pj} \ll 1. \quad (5.7a)$$

Якщо відома техніко-економічна статистика у вигляді реалізації виробів із заданою вартістю системи і її частин, то лінії багатовимірної середньоквадратичної регресії вартості на зазначені параметри можуть описувати зв'язок вартості системи з її технічними, сигнальними й узагалі системними параметрами. Якщо статистика задана у вигляді дискретного, нехай навіть уніфікованого, стандартного ряду елементів, при великому числі параметрів, що оптимізуються, не обійтися без інтерполяції і функціонального подання залежності вартість-параметр. Такий підхід дозволить уникнути складності, нерозв'язності стохастичного програмування і дасть можливість одержати загальне розв'язання в аналітичному вигляді. Для цього подамо багатовимірну середньоквадратичну регресію вартості на параметри в найзагальнішому вигляді і знайдемо, що, незважаючи на адитивність вартості блоків системи, ряд блоків

має вартість, що не є адитивною стосовно своїх параметрів. Наприклад, якщо для антени враховується ряд таких параметрів, як КСД, ККД, габарити, вага, коефіцієнт використання площі антени, коефіцієнт хвилі, що біжить, середньоквадратична похибка виготовлення дзеркала або припустимий рівень бічних пелюсток і т.п., то вартість антени являє собою складну залежність від зазначених параметрів. Таке подання загальної вартості системи не перешкодить вирішенню задачі в загальному вигляді, хоча конкретні залежності вартості від параметрів в основному ще не є об'єктом спостереження й систематизації. Таким чином,

$$C(X_1, \dots, X_N) = \sum_k C_k(X_k) + \sum_l C_l(X_{l1}, X_{l2}) + \sum_t C_t(X_{t1}, X_{t2}, X_{t3}) + \dots, \quad (5.8)$$

де k – індекс підсумовування вартості блоків, що описуються одним технічним параметром;

l – індекс підсумовування вартості блоків, що описуються двома технічними параметрами;

t – індекс підсумовування вартості блоків, що описуються трьома технічними параметрами, й т.п.

У загальному вигляді задача параметричного синтезу суміщених систем може бути записана таким чином:

$$\min C(X_{11}, \dots, X_{p1}, \dots, X_{m1}); \quad (5.9)$$

$$D_l(X_{l1}, \dots, X_{lnl}) \leq D_{l\text{ДБ}}; \quad (5.9a)$$

$$D_p(X_{p1}, \dots, X_{pn_p}) \leq D_{p\text{ДБ}};$$

$$D_m(X_{m1}, \dots, X_{mn_m}) \leq D_{m\text{ДБ}};$$

$$X_{pi\min} \leq X_{pi} \leq X_{pi\max}; \quad (5.9b)$$

$$\forall i_p \in [1, n_p]; \forall p \in [1, m].$$

Такий підхід хороший тим, що за допомогою математичної індукції розв'язання можна поширити на довільне число параметрів і каналів, і що результати оптимізації системи вродіб придатні для використання в більш складних цільових

функціях. Задачу глобального параметричного синтезу РТС (5.9)-(5.9б) іноді називають виродженим динамічним програмуванням, а через те, що використовується поділ змінних – сепарабельним програмуванням. Розв’язок (або оптимальний вектор параметрів $\vec{X}_{pi}(\vec{Y}_{pi})$) може бути знайдений стосовно заданої сукупності показників якості [4, 5], якщо за цільову функцію взяте кожне з обмежень (5.9)-(5.9б). Сукупністю застосовуваних показників якості \vec{D}_p , тобто безумовним критерієм переваги, визначається множина не гірших систем.

З обчислювальної точки зору і для зручності узагальнень доцільно як критерій оптимальності брати мінімум вартості системи:

$$F(\vec{D}_p) = \min_{\{\vec{X}\}} C(\vec{X}). \quad (5.10)$$

Одиничність та існування $F(\vec{D}_p)$ або $\inf C(X)$ зумовлюється опуклістю функцій (5.2), (5.3) і монотонністю $\vec{X}(\vec{Y})$ і $C(\vec{X})$. Мінімум буде тим глибше, чим більше параметрів системи бере участь в оптимізації, тобто чим більше є інформації $C_{pi(j)}(X_{pi(j)}), (n_{p1} \gg 1, n_{p2} \gg 1)$ про стан техніки й виробництва.

Цільову функцію (5.9) доцільно лінеаризувати з нижчеподаних причин. По-перше, все одно прогнозована крива $C_p(X_{ij})$ після згладжування може мати помітну дисперсію хоча б по одному параметру X_{ij} на якому-небудь інтервалі. По-друге, якщо невідома хоча б одна залежність $C_p(X_{pi})$ або її вигляд відрізняється від інших, що частіше й буває, то розв’язання задачі в аналітичному вигляді складне або одержати його неможливо. По-третє, і це найголовніше, розв’язок отримуємо: 1) в аналітичному вигляді; 2) для довільного числа каналів й параметрів, що оптимізуються; 3) для довільних нелінійних $C_p(X_{ij})$, якщо застосовувати отримані рішення як ітераційні.

Лінеаризацію (5.9) доцільно здійснити у межах такого вектора параметрів \vec{X}_o , що найчастіше використовується в

системах того ж класу. Такий вибір точки розкладання, власне кажучи, використовує всю апіорну інформацію про оптимальний розв'язок, що на практиці звичайно добувається від серії до серії за допомогою інженерної індукції методом проб і помилок.

$$C(X_{p1}, \dots, X_{pN}) = \sum_{p=1}^m [\sum_i C_{0pi} + \sum_i C'_{0pi} (X_{pi} - X_{0pi}) + \sum_j C_{0pj} + \sum_j C'_{0pj} (X_{0pj} - X_{pj})],$$

де n_{p1} і n_{p2} – відповідна кількість прямо і зворотно визначених параметрів X_{pi} і X_{pj} , у каналі p ;
 m – число каналів.

Лінійні обмеження дозволяють одержати більш прості рішення задач оптимізації.

5.3. Розв'язання задачі параметричного синтезу суміщених радіотехнічних систем

Задачу, сформульовану у вигляді (5.9) - (5.9б), простіше і результативніше розв'язувати поетапно (кроками) методом виродженого динамічного програмування [1]. Для довільної i -ї групи параметрів p -го каналу, які описують дисперсію похибки вимірювань параметра сигналу або інший якісний показник, таке розв'язання неважко визначити. Для цього, використовуючи подвійність задачі, визначимо мінімум вартості F групи параметра каналу або критерію D_p

$$F_{pi}(D_{p1}) = \min [\sum_i C_{0pi} + \sum_i C'_{0pi} (X_{pi} - X_{0pi})] = \\ = B_{p1} + n_p \sqrt{ \frac{A_p \prod_{i=1}^{n_p} C'_{0pi}}{D_{p1}} }$$

при

$$\frac{A_p}{\prod_{i=1}^{n_p} X_{pi}} \leq D_{p1}, X_{n_p opt} = \frac{1}{n_p C'_{0pn_p}} \sqrt{\frac{A_p \prod_{i=1}^{n_p} C'_{0pi}}{D_{p1}}}, \quad (5.11)$$

де $B_{p1} = \sum_i C'_{0pi} X_{0pi}$.

Аналогічно знаходиться мінімум для групи j будь-якого каналу p

$$\begin{aligned} F_{pi}(D_{p1}) &= \min \left[\sum_i C_{opi} + \sum_{i=1}^{n_{pi}} C'_{opi} (X_{pi} - X_{opi}) \right] = \\ &= B_{p1} + n_p \sqrt{\frac{A_p \prod_{i=1}^{n_p} C'_{opi}}{D_{p1}}}, \end{aligned}$$

де $\{X_{pi}\}$ – множина параметрів p -го каналу,

$$B_{p1} = \sum_i C'_{opi} X_{opi}$$

Аналогічно знаходиться мінімум для груп параметрів j будь-якого каналу p

$$F_{pj}(D_{p2}) = B_{p2} + \min \sum_{j=1}^{n_{p2}} C'_{opj} (-X_{pj}) = B_{p2} - \sqrt{\sum_{j=1}^{n_{p2}} C'^2_{opj}} \sqrt{D_{p2}} \quad (5.12)$$

при

$$\sum_{j=1}^{n_{p2}} X_{pj}^2 \leq D_{p2}, X_{npj opt} = C'_{opj} \sqrt{\frac{D_{p2}}{\sum_{j=1}^{n_{p2}} C'^2_{opj}}},$$

$$\text{де } B_{p2} = \sum_l C_{opl} + \sum_{j=1}^{n_{p2}} C'_{opj} X_{opj}.$$

Для несуміщеної системи, через те, що канали незалежні й відсутні загальні блоки й елементи системи, мінімальна вартість системи визначається як сума знайдених мінімумів.

$$F = F(D_{p1}, \dots, D_p, \dots, D_m) = \min_{\{X_{pi}, D_{pi}\}} \sum_{p=1}^m [F_i(D_{p1}) + F_j(D_{p2})],$$

при

$$D_{p1} + D_{p2} = D_p,$$

$$F = \sum_{p=1}^m \min [F_i(D_{p1}) + F_j(D_{p2})] = \sum_{p=1}^m [F_i(D_{p1opt}) + F_j(D_p - D_{p1opt})], \quad (5.13)$$

$$\{X_{pi}, D_{pi}\}.$$

$$D_{p1} + D_{p2} < D,$$

$$\text{де } D_{p1} \cong \frac{\sqrt{1 + 4D_p \frac{n_p^4 \left(\sqrt[n_p]{A_p \prod_{i=1}^{n_{p1}} C'_{opi}} \right)^2}{2 \sum_{j=1}^{n_p} C'_{opj}{}^2}} - 1}{\left(\sqrt[n_p]{A_p \prod_{i=1}^{n_{p1}} C'_{opi}} \right)^2 \sum_{j=1}^{n_{p2}} C'_{oj}{}^2}, \quad (5.14)$$

$n_p \gg 1$.

У загальному випадку D_{p1opt} легко визначити для будь-якого n_p методом Ньютона-Рафсона за такою ітераційною формулою:

$$D_{pI(k)} = \frac{D_p - D_{pI(k-1)}^{2b} \frac{(3n_{p1} + 2)n_{p1}Q_{n_{p1}}^2}{2Q_{n_{p2}}^2}}{1 + bD_{pI(k-1)}^{2b} n_p^4 \frac{Q_{n_{p1}}^2}{Q_{n_{p2}}^2}}, \quad (5.15)$$

$$\text{де } b = \frac{n_{p1} + 1}{n_{p1}}, \quad (5.16)$$

$$Q_{n_{p1}} = n_{p1}^{n_{p1}} \sqrt{A_p \prod_{i=1}^{n_{p1}} C'_{opi}}, \quad (5.16a)$$

$$Q_{n_{p2}} = \sqrt{\sum_{j=1}^{n_{p2}} C'_{opj}{}^2}. \quad (5.16b)$$

Таким чином, якщо $\alpha = \frac{Q_{n_{p1}}^2}{Q_{n_{p2}}^2}$, то розв'язок спроститься

$$D_{pIopt} \cong \frac{\sqrt{1 + 4\alpha D_p} - 1}{\alpha}. \quad (5.17)$$

З формул (5.15)- (5.16b) одержимо

$$D_{pI(k)} = \frac{D_p - \alpha D_{pI(k-1)}^{2b} \frac{3n_{p1} + 2}{2n_{p1}}}{2\alpha b D_{pI(k-1)}^{2b}}. \quad (5.18)$$

Для суміщених систем, коли хоча б один передавач є загальним, оптимальний розв'язок (5.11) - (5.16b) змінює свій вигляд. Для зручності обидві частини виразів (5.2) для будь-якого p можна поділити на найкраще значення якісного показника D_{pIpred} і нормувати вираз таким чином:

$$\frac{A_p \prod_{i=1}^{n_{p1}} X_{p i n_{i n p}}}{A_p \prod_{i=1}^{n_{p1}} X_{p i}} \leq \frac{D_{p1}}{D_{p1 n_{p e d}}}$$

або

$$\frac{D_{p1 n_{p e d}}}{\prod_{i=1}^{n_{p1}} X_{p i}^0} \leq D_{p1},$$

де $X_{p i n_{i n p}}$ – значення функції для найкращого технічного параметра;

$D_{p1 n_{p e d}}$ – найкраща (потенційна) при даному розвитку техніки дисперсія похибки вимірювань або краще значення показника якості.

Тоді

$$X_{p i}^0 = \frac{X_{p i}}{X_{p i n_{i n p}}}.$$

Тут і далі під $X_{p i}$ будемо розуміти відносні величини «фазових» параметрів $X_{p i}^0$, хоча не завжди нормування (5.2) є доцільним, наприклад при малому числі параметрів, що оптимізуються. Результати (5.11) - (5.16б) при цьому не зміняться, потрібно лише при нормуванні замінити A_p величиною $D_{p1 n_{p e d}}$. Тоді, з огляду на нормування F і $X_{p j}$, задачу можна записати у вигляді:

$$F = \min \sum_{p=1}^m [B_{p1} + B_{p2} + \sum_{i=1}^{n_{p1}} C'_{op i} X_{p i} - \sum_{j=1}^{n_{p2}} C'_{op j} X_{p j}] \quad (5.19)$$

при

$$\frac{D_{p1 n_{p e d}}}{\prod_{i=1}^{n_{p1}} X_{p i}} \leq D_{p1}, \quad 0 \leq X_{p i} \leq 1,$$

$$\sum_{j=1}^{n_{p2}} X_{pj}^2 \leq D_{p2}, \quad X_{pjn_{p2}} \leq X_{pj} \ll 1,$$

$$D_{p1} + D_{p2} \leq D_p \quad \forall p \in [1, m].$$

Нехай серед усіх параметрів групи $i \in K$, а серед групи $j \in l$ таких, які описують суміщені блоки (наприклад блоки 1-6 на рис. 5.1). При цьому передбачається або сигнальне суміщення, або його відсутність при послідовному в часі використанні каналів. На формалізації задачі ця обставина не позначається.

Оскільки K параметрів групи i є загальними в добутку $\prod_{i=1}^{n_{p1}} X_{pi}$, позначимо їхній добуток буквою K_i і назовемо його параметром суміщення групи i :

$$K_i = \left(\prod_{i=1}^k X_{pi} \right)^{-1}. \quad (5.20)$$

Оскільки l параметрів групи j також є загальними, їхня сума позначається буквою L_j . Назвемо її параметром суміщення групи j :

$$\sum_{j=1}^l X_{pj}^2 \leq L_j. \quad (5.21)$$

З огляду на позначення (5.20), (5.21), обмеження до задачі (5.19) можна записати з розділеними змінними, що дозволяє скористатися тим же методом розв'язання

$$F = \min \sum_{p=1}^m [B_{p1} + B_{p2} + \sum_{i=1}^{n_{p1}} C'_{opi} X_{pi} - \sum_{j=1}^{n_{p2}} C'_{opj} X_{pj}] ,$$

$$\frac{D_{p1nped}}{\prod_{i=1}^{n_{p1}} X_{pi}} \leq \frac{D_{p1}}{K_i}, \quad 0 \leq X_{pi} \leq 1, \quad (5.22)$$

$$\sum_{j=1}^{n_{p2}-l} X_{pj} \leq D_{p2} - L_j, \quad X_{pjn_{p2}} \leq X_{pj} \ll 1,$$

$$\forall i_p \in [1, n_{p1}], \forall j_p \in [1, n_{p2}],$$

$\frac{1}{\prod_{i=1}^k X_i} \leq K_i$ - для параметрів суміщеної частини систем

$$\sum_{j=1}^l X_j^2 \leq L_j.$$

Виконавши перші кроки оптимізації (5.11), (5.12) і перенумерувавши доданки, одержимо

$$F = B_{p1} + B_{p2} + \min \left[\sum_{p=1}^m (Q_{n_{p1}-k} K_i^{\frac{1}{n_{p1}-k}} - Q_{n_{p2}-l} \sqrt{D_{p2}-L} + Q_k K_i^{\frac{1}{k}} - Q_l \sqrt{L}) \right], \quad (5.23)$$

де $Q_{n_{p1}-k} = (n_{p1} - k)^{n_{p1}-k} \sqrt{\frac{D_{p1n_{p1}-k} \prod_{i=1}^{n_{p1}-k} C'_{opi}}{D_{p1}}}$,

$$Q_{n_{p2}-l} = \sqrt{\sum_{j=1}^{n_{p2}-l} C'_{opj}{}^2},$$

$$Q_k = k^k \sqrt{\prod_{i=1}^k C'_{oi}}, Q_l = \sqrt{\sum_{i=1}^l C'_{oj}{}^2}.$$

Припустимо, що кількість параметрів n_p , що оптимізуються, у каналах однакова, тобто $n_{p1}=n$. Тоді

$$F = B_{p1} + B_{p2} + \min \left(K_i^{\frac{1}{n-k}} \sum_{p=1}^m Q_{n_{p1}-k} + K_i^{\frac{1}{k}} Q_k - \sum_{p=1}^m Q_{n_{p1}-l} \sqrt{D_{p2}-L} - Q_l \sqrt{L} \right) \quad (5.24)$$

Мінімум по K_i легко знаходиться з виразу (5.24)

$$K_i = \left(\frac{n-k}{k} \frac{Q_k}{\sum_{p=1}^m Q_{n_{p1}-k}} \right)^{\frac{(n-k)}{n}}. \quad (5.25)$$

Шукаючи L_{jopt} , варто врахувати ту обставину, що звичайно на практиці $L < D_{p2}$, оскільки сумісний блок як еталон для вимірювань являє собою задаючий генератор або стандарт частоти й т.п. Таким чином,

$$L_{jopt} \cong \frac{Q_l^2}{\left(\sum_{p=1}^m \frac{Q_{n_{p2-l}}}{\sqrt{D_{p2}}} \right)^2}. \quad (5.26)$$

Точніше L_{iopt} можна визначити з кубічного рівняння

$$(\sqrt{L})^3 \sum_{p=1}^m \frac{Q_{n_{p2-l}}}{D_{p2} \sqrt{D_{p2}}} + (\sqrt{L}) \sum_{p=1}^m \frac{Q_{n_{p2-l}}}{\sqrt{D_{p2}}} - Q_l = 0,$$

користуючись розв'язком Кардано. Мінімальні асигнування на суміщену систему визначаються з (5.24) з урахуванням (5.25), (5.26)

$$F = B_{p1} + B_{p2} + \frac{Q_k^d (\sum Q_{n_{p2-l}})^{1-d}}{d^d (1-d)^{1-d}} - \frac{\sum Q_{n_{p2-l}} \sqrt{D_{p2} \left(\sum_{p=1}^m \frac{Q_{n_{p2-l}}}{\sqrt{D_{pi}}} \right)^2 - Q_l^2 + Q_l^2}}{\sum_{p=1}^m \frac{Q_{n_{p2-l}}}{\sqrt{D_{p2}}}}. \quad (5.27)$$

Оптимальні технічні параметри знаходяться у нижченаведеному порядку.

Спочатку визначаються оптимальні параметри суміщення K_i і L_j , а потім параметри суміщеної частини системи за формулами (5.11), (5.12). Визначивши відношення $\frac{D_{p2}}{K_i}$, за тими ж формулами легко визначити інші технічні параметри групи i , а

знаючи різницю $D_{p2} - L_j$, можна визначити інші технічні параметри групи j . За формулами (5.24) - (5.26) визначаються також оптимальні асигнування на блоки системи. Якщо оптимальні параметри вийшли за межі обмежень (5.22), то вони фіксуються й задача розв'язується знову для інших параметрів. Для випадку, коли $n_n = n_{2i} = n_{p1}$, оптимальну величину параметра K_i суміщення групи i можна визначити за ітераційною формулою

$$K_{i[q+1]} = \frac{1}{\frac{K}{Q_k} \sum_{p=1}^m \frac{Q_{n_{p2}-k}}{n_{p2}-k} K_{i[q]} \frac{n_{p1}}{(n_{p1}-k)}}, \quad (5.28)$$

де q -крок ітерації,

$$K_{i[1]} = \frac{(n_{p1}-k)Q_k}{kQ_{n_{p1}-k}}.$$

Для кращої збіжності при першому кроці доцільно нумерувати канали в порядку спадання

$$\frac{Q_{n_{p2}-k}}{n_{p2}-k}.$$

Далі можна визначити оптимум F для p каналів, вводячи обмеження

$$D_{p1} + D_{p2} \leq D_p.$$

Черговими кроками нескладно знайти оптимум F також для всіх параметрів, що раніше не оптимізувались, закладених у коефіцієнти A_p .

5.4. Обговорення результатів

Як видно з наведених виразів, параметричний синтез суміщених систем можливий для довільного числа незалежних вимірюваних параметрів, каналів і обмежень. Отримані оптимальні рішення можуть дозволити оцінювати оптимальні

технічні параметри й мінімальні асигнування на частину системи, що оптимізується, у деякій області $\vec{X}_{p_{i,j}}$, якщо цільова функція в (5.9) близька до гіперплощини. При використанні розв'язок ітераційних формул, якщо регулювати крок наближення залежно від кривизни обмежень $C_{p_{i,j}}(\vec{X}_{p_i}, \vec{X}_{p_j})$, оптимум визначається послідовно для довільних опуклих обмежень, а також для опукло-увігнутих й увігнутих функцій, що задовольняють умову гіперболічної залежності цільової функції від параметрів. Середнє геометричне кутових коефіцієнтів ліній регресії вартості на параметри суміщених блоків позначимо C'_{oc} , а несуміщених блоків – $C'_{орн}$, то

$$C'_{oc} = \sqrt[k]{\prod_{i=1}^k C'_{oi}}, \quad C'_{op} = \sqrt[n-k]{\prod_{i=1}^{n-k} C'_{opi}},$$

тоді неважко помітити згідно з (5.24), що

$$F_1 = n \exp[\alpha \ln C'_{oc} + (1 - \alpha) \ln C'_{ора}] \left(\sum_{p=1}^m \varepsilon_p^{\frac{1}{n(1-\alpha)}} \right)^{1-\alpha}, \quad (5.29)$$

де $Z'_{ора}$ – середньоарифметичне значення $C'_{орн}$ по всіх p каналах

$$\alpha = \frac{k}{n},$$

$$\varepsilon_p = \frac{D_{пред}}{D_{p1}} \left(1 - \frac{C'_{орн} - C'_{ора}}{C'_{ора}} \right).$$

При збільшенні α або числа параметрів суміщених блоків k , асигнування F_1 , як і слід було сподіватися, зменшуються, оскільки $\varepsilon_p < 1$ і експонентний множник також зменшується при $C'_{oc} \leq C'_{ора}$. Якщо ж $C'_{oc} > C'_{ора}$, то експонентний множник при збільшенні α збільшується. У цьому випадку може існувати оптимальне число α або число параметрів суміщених блоків k . Оптимальний параметр α можна визначити з рівняння

$$A \sum_{p=1}^m \varepsilon_p^x + x \sum_{p=1}^m \varepsilon_p^x \ln \varepsilon_p^x = 0, \quad (5.30)$$

$$\text{де } x = \frac{1}{n(1-\alpha)}, \quad A = \ln \frac{C'_{oc}}{C'_{opn}}.$$

Розкладаючи ε_p в ряд Тейлора в межах $\delta_p = 1$ для високоякісних систем і залишаючи лише члени другого порядку малості, одержимо

$$x_{opt} = \frac{A_m}{B(1+A)}, \quad \alpha_{opt} = 1 - \frac{B(1+A)}{A_{mn}}, \quad (5.31)$$

$$\text{де } B = \sum_{p=1}^m \ln \frac{1}{\varepsilon_p}.$$

Оптимум існує для

$$A \geq \frac{B}{mn - B}.$$

Таким чином, для дослідження питань оптимальності сумісних систем велике значення має метод аналізу ефективності суміщення. Застосований у роботі метод визначення оптимальності суміщення має ту перевагу, що він органічно пов'язаний з параметричним синтезом системи й може дати практичні рекомендації з більш загальних питань оптимізації суміщених систем.

У результаті рішення загальної задачі оптимізації сумісної системи можна зробити такі висновки:

- параметричний синтез сумісної системи можна робити для типових багатокритерійних систем з довільним числом технічних параметрів і незалежних каналів вимірювання параметрів руху й передачі інформації;
- для кожної задачі параметричного синтезу, що відрізняється хоча б одним обмеженням або критерієм, як чисельно, так і якісно, тобто складом обмежень, існують свої

оптимальні значення параметрів або оптимальні розв'язання, які перетворюють цільову функцію (вартість) у мінімум;

- у випадку, коли середнє геометричне похідних від вартості за параметром для сумісних блоків менше або дорівнює середньому геометричному таких же похідних для несумісних блоків, число сумісних блоків доцільно по можливості збільшити, оскільки вартість системи при цьому монотонно падає. У протилежному випадку може спостерігатися зазначений вище оптимум числа параметрів сумісних блоків;

- у випадку, коли число параметрів, що оптимізуються, у всіх каналах однаково, задача має канонічне розв'язання в аналітичному вигляді. У протилежному випадку розв'язання для параметрів суміщення знаходяться числовими методами за отриманими ітераційними формулами;

- отримані оптимальні розв'язання дозволяють виконувати інженерні розрахунки технічних параметрів системи при мінімально необхідних асигнуваннях на системи. Розв'язання правильні для достатньої наближеності лінійних обмежень реальним лініям регресії вартості на параметри;

- результати розв'язання можна використати для довільних обмежень, якщо використовувати застосований у розд. 4 метод послідовних наближень для функцій регресії;

- розв'язання задачі аналогічне для оптимізації сумісних двобічних систем, що використовують прості й складні сигнали, або для систем іншого класу, якщо показники якості систем формалізуються в такий же спосіб;

- задача взагалі допускає поетапне визначення оптимальних параметрів.

5.5. Оптимізація сумісних РТС за основними ТТВ методом геометричного програмування при позиноміальних залежностях вартості від параметрів

Розглянемо питання оптимізації РТС на множині критеріїв точності вимірювань параметрів руху літальних апаратів або похибок передачі інформації при істотному обмеженні на вартість системи і при позиноміальних залежностях вартості елементів від їхніх параметрів. Позином – це поліном з

позитивними коефіцієнтами. Зазначений клас залежностей є досить широким для практичних застосувань.

Розв'язання задач таким методом особливо ефективно при невеликому числі параметрів, що оптимізуються. Хоча геометричне програмування й спрощує рішення, подаючи задачу у подвійному вигляді і зводячи розв'язання до системи лінійних рівнянь, збільшення числа змінних приводить до росту розмірності матриці ваги, яку необхідно обертати.

Для задач підвищеної складності (по Зенеру: «Метод геометричного програмування») розв'язання може бути отримане лише приблизно. Але принципово технічна задача завжди має розв'язок і його одиничність для опуклих обмежень показана. Крім того, розв'язок може бути отриманий лише в числовому вигляді. Таке положення нас би влаштовувало, якби повністю виключалася можливість одержання розв'язання в загальному вигляді. Однак така можливість є навіть для геометричного програмування. Для цього також необхідна лінеаризація обмежень за вартістю в межах оптимального рішення, що настільки спрощує задачу (зводячи її до нульового ступеня труднощів), що розв'язок також може бути отриманий в загальному вигляді.

Отже, проілюструємо рішення розглянутої задачі сепарабельного програмування методом геометричного програмування, що дозволить нам не тільки переконатися в справедливості отриманого рішення, але й глянути на проблему з іншого боку, яка може мати свої можливості у вивченні задач синтезу. Лінеаризація обмежень хоча й обмежує клас задач функціоналами спеціального вигляду, але це більш широкий клас, ніж опукле програмування. Крім того, лінеаризація дає закінчений результат у вигляді формул і програм з усіма відомими параметрами.

Отже, як і раніше (підрозд. 5.2), припускаємо, що обмеження на асигнування існує і має сенс, оскільки є висока кореляція вартості РТС і її технічних параметрів, від яких залежать якісні показники РТС, у цьому випадку вектор дисперсій \vec{D}_p похибок оцінки параметрів і ймовірностей похибок

передачі інформації. Розглянемо спочатку однокритерійну задачу. У цьому випадку відповідно до розд. 4

$$\min \Delta C = \min \sum_{i=1}^{n_l} C'_i X_i \quad (5.32)$$

при

$$1 \geq \frac{A_l}{D_l} \prod_{i=1}^{n_l} X_i^{-1}, X_i > 0, \quad (5.33)$$

де $\Delta C = C_o - \sum_{i=1}^{n_l} C'_{oi} X_{oi}$ – збільшення асигнувань на розглянуту частину системи;
 C – асигнування на РТС, що відповідають вектору технічних параметрів X_o ;

$$C'_{oi} = \frac{\partial C(\vec{X})}{\partial X_i} /_{X_i=X_{oi}}; \quad (5.34)$$

n_l – кількість технічних параметрів \vec{X}_o ;

A_l, D_l – постійний коефіцієнт і дисперсія оцінки вимірюваного параметра.

Формули (5.32) – (5.34) описують пряму задачу (5.32), вимушені (5.33) і природні обмеження (5.34). Передподвійна функція $\Delta_{,,\varnothing}$ буде мати вигляд

$$\Delta C \geq \Delta C_{\varnothing} = \prod_{i=1}^{n_l} \left(\frac{C'_i}{\delta_i} \right)^{\delta_i} \frac{A_l}{D_l} \prod_{i=1}^{n_l} X_i^{-1} \prod_{i=1}^{n_l} X_i^{\delta_i} \lambda_o^{\lambda_0} \lambda_l^{\lambda_l}, \quad (5.35)$$

де $\vec{\delta}$ – векторна змінна, або вага

$$\lambda_0 = 1, \lambda_l(\delta) = \delta_{n_l+1}.$$

Умови ортогональності визначаються з (5.35) векторним рівнянням

$$D = \underline{a\delta}, \quad (5.36)$$

Розв'язання (5.41) придатне для опуклих функцій $C(\bar{X})$, якщо формулу (5.41) використати як ітераційну, а також для увігнутих функцій, що задовольняють умову гіперболічної залежності показника від параметрів.

Розглянемо задачу оптимізації сумісної РТС за критеріями точності і за критеріями ймовірності похибок передачі інформації.

Задача формалізується аналогічно до попередньої.

$$\Delta C = \sum_{k=1}^p \sum_{i=1}^{n_k} C'_{ik} X_{ik}, \quad (5.42)$$

$$I \geq \frac{A_k}{D_k} \prod_{i=1}^{n_k} X_{ik}^{-1}; \quad k = \overline{1, p}; \quad (5.43)$$

$$X_{ik} > 0. \quad (5.44)$$

Якщо немає спільних (загальних) технічних параметрів хоча б для двох каналів, то задача розпадається на p більш простих, уже розв'язаних задач (5.32) - (5.34).

Нехай число спільних параметрів – n_0 . Запишемо вимушені обмеження (5.43) у вигляді

$$I \geq \frac{A_k}{D_k} \prod_{i=1}^{n_0} X_{oi}^{-1} \prod_{l=1}^{n_k - n_0} X_{ll}^{-1}. \quad (5.45)$$

Передподвійна функція має вигляд

$$\begin{aligned} \Delta C \geq & \prod_{k=1}^p \left(\frac{A_k}{D_k \delta_{l+k}} \right)^{\delta_{l+k}} \prod_{i=1}^{n_0} \left(\frac{C_i}{\delta_i} \right)^{\delta_i} \prod_{i=n_0+1}^{n_0+m_1} \left(\frac{C_i}{\delta_i} \right)^{\delta_i} \prod_{i=n_0+m_1+1}^{n_0+m_2} \left(\frac{C_i}{\delta_i} \right)^{\delta_i} \\ & \prod_{i=n_0+m_{p-1}+1}^{n_0+m_p} \left(\frac{C_i}{\chi_i} \right)^{\delta_i} \prod_{i=1}^l X_i^{\delta_i} \prod_{t=1}^p \prod_{i=1}^{n_0} X_i^{-\delta_{l+t}} \prod_{i=n_0+1}^{n_0+m_1} X_i^{\delta_i} \dots \\ & \prod_{i=n_0+m_p+1}^{n_0+m_p} X_i^{\delta_i} \prod_{t=1}^p \lambda_t^{\lambda_t}, \end{aligned} \quad (5.46)$$

$$l = n_0 + \sum_{k=1}^p m_k, \quad l = n_0 + \sum_{k=1}^p m_k, \quad (5.47)$$

$$n_0 \delta_i + m \sum_{k=1}^P \delta_{\ell+k} = (n_0 + m) \delta_i = 1,$$

тобто

$$\delta_i = \frac{1}{n_0 + m} \quad \text{для} \quad i = \overline{1, n_0}. \quad (5.51)$$

Визначимо оптимальні значення базисних змінних $\delta_{\ell+j}$ ($j = \overline{1, P}$), при яких $\Delta C = \max \Delta C_{\partial}$, тобто існує строга рівність (5.46) (сильний екстремум). Оскільки існує співвідношення (5.49) і (5.51), тобто

$$\sum_{j=1}^P \delta_{\ell+j} = \delta_i = \frac{1}{n_0 + m}, \quad (5.52)$$

то

$$\Delta C_{\partial} = \max \prod_{k=1}^P \left(\frac{A_k}{D_k \delta_{\ell+k}} \right)^{\delta_{\ell+k}} \lambda_k^{\lambda_k} \quad \text{при} \quad \sum_{j=1}^P \delta_{\ell+j} = \frac{1}{n_0 + m}.$$

Максимум інваріантний стосовно базисних рішень і досягається, коли

$$\delta_{(\ell+j)opt} = \delta_{(\ell+j-1)opt} = \frac{1}{P(n_0 + m)}.$$

Тоді

$$\begin{aligned} \lambda^{\lambda} &= B_{\lambda} = [p(n_0 + m)]_{n_0+m}^{\frac{1}{p}}; \\ \delta_{\ell+j} &= \delta_{\ell+j-1} = \frac{1}{p(n_0 + m)}; \\ \delta_i &= \delta_{\ell+j} = \frac{1}{p(n_0 + m)} \quad \text{для} \quad n_0 \leq i \leq \ell. \end{aligned} \quad (5.53)$$

Наведемо перевірку правильності отриманих подвійних змінних δ_i , для чого підставимо ці значення (5.53), (5.51) в умову нормалізації (5.48):

$$\sum_{i=1}^{\ell} \delta_i = \frac{n_0}{n_0 + m} + \frac{\ell - n_0}{p(n_0 + m)} \equiv 1.$$

тому що $\ell = n_0 + mp$.

Обчислимо подвійну функцію:

$$\Delta C_{min} = \Delta C_{\partial max} = (n_0 + m)p^{\frac{m}{n_0+m}} \left[\prod_{k=1}^p \left(\frac{A_k}{D_k} \right)^{\frac{1}{p}} \prod_{i=1}^{n_0} C'_i \prod_{i=n_0+1}^{\ell} C'_i{}^{\frac{1}{p}} \right]^{\frac{1}{n_0+m}}. \quad (5.54)$$

Оптимальний вектор технічних параметрів можна визначити з умови інваріантності членів прямого позінома від індексу i :

$$u_i = \Delta C_{\partial}(\bar{\delta}) \delta_i, \text{ т.е.} \quad \frac{U_i}{\delta_i} = \Delta C_{\partial}(\bar{\delta}).$$

Звідси

$$X_{i opt} = \frac{\Delta C_{min} \delta_i}{C'_{x_i}} = \frac{\Delta C_{min}}{C'_i(n_0 + m)} \quad \text{для} \quad i \leq n_0. \quad (5.55)$$

Для визначення несумісних технічних параметрів X_i , де $I > n_0$, використовуємо розрахункове співвідношення

$$u_i = \frac{\delta_i}{\lambda_j(\bar{\delta})} \quad \text{для} \quad i > n_0.$$

Тоді

$$\frac{A_k}{D_k} \prod_{i=1}^{n_0} X_i \prod_{i=n_0+(k-1)m}^{n_0+m} X_i = \frac{\delta_{\ell} + j}{\lambda_j} = 1;$$

$$\prod_{i=n_0+(k-1)m}^{n_0+km} X_i = \frac{D_k}{A_k} \left[\frac{C'_i(n_0 + m)}{\Delta C_{min}} \right]^{np} \quad \text{для} \quad \forall_i > n_0, \quad i \quad k = \overline{1, p}.$$

Звідси

$$X_{i \text{ opt}} = \left(\frac{D_k}{A_k} \right)^{\frac{1}{m}} \left[\frac{C_i(n_0 + m)}{\Delta C_{\min}} \right]^{\frac{n_0}{m}} \quad (5.56)$$

для $n_0 + m_k \geq i \geq n_0 + m_{k-1} + 1$.

У більш загальному випадку, коли $m_k \neq m_{k-1}$, передподвійну функцію прямої задачі (5.46) можна подати у вигляді

$$\Delta C_{\delta} = \prod_{k=1}^P \left(\frac{A_k}{D_k \delta_{\ell+k}} \right)^{\delta_{\ell+k}} \lambda_k^{\lambda_k} \prod_{i=1}^{\ell} \left(\frac{C'_i}{\delta_i} \right)^{\delta_i} \prod_{\ell+1}^{\ell} X_i^{\delta_i} \prod_{k=1}^P \left(\prod_{i=1}^{n_0} X_i \prod_{i=n_0 m_{k-1}}^{n_0 + m_k} X_i \right)^{-\delta_{\ell+k}}. \quad (5.57)$$

Умови нормалізації й ортогональності мають вигляд

$$\sum_{i=1}^{\ell} \delta_i = 1. \quad (5.58)$$

З умов ортогональності випливає, що

$$\begin{aligned} X_1 \quad & 1 \cdot \delta_1 + \dots + 0 \cdot \delta_{n_p} + \dots + 0 \cdot \delta_{\ell} - 1 \cdot \delta_{\ell+1} - 1 \cdot \delta_{\ell+2} - \dots - 1 \cdot \delta_{\ell+p} = 0; \\ & \dots \dots \dots \\ X_{n_0} \quad & 0 \cdot \delta_1 + \dots + 1 \cdot \delta_{n_0} + \dots + 0 \cdot \delta_{\ell} - 1 \cdot \delta_{\ell+1} - 1 \cdot \delta_{\ell+2} - \dots - 1 \cdot \delta_{\ell+p} = 0; \\ X_{n_0+1} \quad & 0 \cdot \delta_1 + \dots + 1 \cdot \delta_{n_0} + 1 \cdot \delta_{n_0+1} + \dots + 0 \cdot \delta_{\ell} - 1 \cdot \delta_{\ell+1} - 0 \cdot \delta_{\ell+2} - \dots - 0 \cdot \delta_{\ell+p} = 0; \\ & \dots \dots \dots \\ X_{n_0+m_1} \quad & 0 \cdot \delta_1 + \dots + 0 \cdot \delta_{n_0} + 1 \cdot \delta_{n_0+m_1} + \dots + 0 \cdot \delta_{\ell} - 1 \cdot \delta_{\ell+1} - 0 \cdot \delta_{\ell+2} - \dots - 0 \cdot \delta_{\ell+p} = 0; \\ X_{n_0+m_1+1} \quad & 0 \cdot \delta_1 + \dots + 0 \cdot \delta_{n_0} + \dots + 1 \cdot \delta_{n_0+m_2} + \dots + 0 \cdot \delta_{\ell} - 0 \cdot \delta_{\ell+1} - 1 \cdot \delta_{\ell+2} - \dots - 0 \cdot \delta_{\ell+p} = 0; \\ & \dots \dots \dots \\ X_{n_0+m_2} \quad & 0 \cdot \delta_1 + \dots + 0 \cdot \delta_{n_0} + \dots + 1 \cdot \delta_{n_0+m_2} + \dots + 0 \cdot \delta_{\ell} - 0 \cdot \delta_{\ell+1} - 1 \cdot \delta_{\ell+2} - \dots - 0 \cdot \delta_{\ell+p} = 0; \\ & \dots \dots \dots \\ X_{\ell} \quad & 0 \cdot \delta_1 + \dots + 0 \cdot \delta_{n_0} + \dots + 0 \cdot \delta_{\ell-1} + \dots + 1 \cdot \delta_{\ell} - 0 \cdot \delta_{\ell+1} - 0 \cdot \delta_{\ell+2} - \dots - 1 \cdot \delta_{\ell+p} = 0. \end{aligned} \quad (5.59)$$

$$\delta_i = \delta_{i-1} = \sum_{k=1}^P \delta_{\ell+k} \quad \text{для} \quad i \leq n_0; \quad (5.60)$$

$$\delta_i = \delta_{i-1} = \delta_{\ell+k} \quad \text{для} \quad m_{k-1} + n_0 \leq i \leq n_0 + m_k. \quad (5.61)$$

Відповідно до (5.60), (5.61) умова нормалізації набуває вигляду

$$n_0 \sum_{k=1}^p \delta_{\ell+k} + \sum_{k=1}^p \left(\sum_{i=n_0+m_{k-1}}^{n_0+m_k} \delta_i \right) = 1.$$

Далі

$$\sum_{k=1}^p \delta_{\ell+k} [n_0 + (m_k - m_{k-1})] = 1;$$

або

$$(\delta_{\ell+k}, n_0 + (m_k - m_{k-1})) = 1. \quad (5.62)$$

Звідси випливає, що при виконанні умови $n_0 \gg m_k - m_{k-1}$ і при $m_k - m_{k-1} \approx \text{const}(k)$, $m_k = m_{k-1}$. У цьому випадку

$$\sum_{k=1}^p \delta_{\ell+k} = \frac{1}{n_0 + m_k + m_{k-1}}. \quad (5.63)$$

Тоді

$$\delta_{\ell+k} = \frac{1}{p(n_0 + m_k + m_{k-1})}. \quad (5.64)$$

Якщо $m_k - m_{k-1} = f(k)$, то зміною нумерації змінних $f(k)$ завжди можна зробити обвідну опуклою або увігнутою функцією $f(k)$, після чого для визначення $\delta_{\ell+k}$ використовувати відповідні розв'язання (5.62) інтегрального рівняння

$$(\delta_{(k)}(y), n_0 + f(y)) = \delta_{(k)}(y).$$

Далі визначаємо значення δ_i для $i > n_0$ й $i \leq n_0$ і оптимальні значення технічних параметрів із системи функціональних рівнянь

$$u_i = \Delta C_{\delta} \delta_i \quad \text{для } i \leq n_0. \quad (5.65)$$

і

$$u_i = \frac{\delta_i}{\lambda_k} \quad \text{для } i \geq n_0. \quad (5.66)$$

Логарифмуванням лівої й правої частин система (5.57) зводиться до лінійної системи, що відповідає (5.36), (5.37)

$$\sum_{i=1}^{n_0} a_{ij} \ln X_i = \ln(\Delta C_{\delta} \cdot \delta_j) - \ln C_j. \quad (5.67)$$

З викладеного можна зробити нижченаведені висновки.

Висновки

1. Методика оптимізації радіотехнічних систем може бути заснована на методі геометричного програмування, якщо статистичні ансамблі вартостей і параметрів блоків системи згладжуються позіномами. Розв'язання найбільш доцільне, якщо кількість технічних параметрів, що беруть участь в оптимізації, не занадто велика. У протилежному випадку доцільно користуватися лінійною апроксимацією цих залежностей. Лінійне згладжування приводить до точного розв'язання задачі оптимізації РТС.

2. Використання методу геометричного програмування дозволяє приблизно, але простіше розв'язувати задачу для випадків, коли динамічна програма (покрокова) досить складна $m_k \neq m_{k-1}$.

3. Є можливість для практично цікавих випадків виконати геометричну програму до кінця в найзагальнішому вигляді, коли лінійні члени позіномів згладжування статистичних даних є найістотнішими.

4. Наведена програма не обмежена кількістю вимушених обмежень і кількістю технічних параметрів у каналах.

6. ЗАДАЧІ ЛІНІЙНОГО БУЛЕВОГО ПРОГРАМУВАННЯ В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ І МЕРЕЖАХ

6.1. Методи розв'язання задач дискретної оптимізації і їх класифікація

Задачі дискретної оптимізації і їх постановка

З метою єдиного уявлення про понятійний апарат теорії алгоритмів сформулюємо визначення основних термінів.

Під *масовою задачею* (або просто *задачею*) будемо розуміти деяке загальне питання, на яке варто дати відповідь. Звичайно задача містить кілька параметрів, або вільних змінних, конкретні значення яких не визначені. Задача P визначається такою інформацією:

- загальним списком усіх її параметрів;
- формулюванням тих властивостей, які повинні задовольняти відповідь або, інакше кажучи, розв'язок задачі.

Індивідуальна задача I виходить з масової P , якщо всім параметрам задачі P присвоїти конкретні значення. Інакше кажучи, індивідуальна задача оптимізації – це пари (F, C) , де F – довільна множина, область припустимих точок, а C – функція вартості, що здійснює відображення $C:F \rightarrow R$. Потрібно знайти точку $f \in F$, для якої $C(f) < C(y)$ для всіх $y \in F$. Така точка називається *глобальним оптимальним розв'язком*.

Задачі математичного програмування, у яких для всіх (для частини) компонентів припустимих параметрів висуваються вимоги цілочисельності, називаються задачами *цілочисельного (частково цілочисельного) програмування*.

Задачі цілочисельного програмування, у яких припустимі лише значення компонентів векторних альтернатив 0 або 1, називаються *задачами цілочисельного програмування з булевими змінними*, або *задачами булевого (бівалентного) програмування*.

Задачі цілочисельного програмування, у яких припустимі лише значення компонентів векторних альтернатив $(\overline{0, k-1})$, називаються *задачами програмування з k -значними змінними* або *k -значного програмування* (у загальному випадку число k різне для різних компонентів).

Задачі оптимізації на кінцевих множинах об'єктів, що будуються за правилами комбінаторики, називаються *задачами комбінаторного програмування*.

Усі перераховані задачі оптимізації зважуються на відповідних дискретних множинах (у більшості випадків кінцевих). У загальному випадку під *дискретною множиною* розуміється множина, усі точки якої є ізольованими точками даної множини.

Кожна задача комбінаторного програмування в принципі може бути подана як задача *цілочисельного* (частково *цілочисельного*) програмування. Однак таке подання не завжди є доцільним.

Таким чином, *задача оптимізації* – це множина I індивідуальних задач оптимізації. В індивідуальній задачі подані вихідні дані і є досить інформації для одержання розв'язання, у той час як задача оптимізації це набір індивідуальних задач, породжуваних однаковою способом. Так, якщо ми говоримо про індивідуальну задачу комівояжера, то матриця відстаней задана, якщо ж ми говоримо про задачу комівояжера в цілому, то мається на увазі набір усіх індивідуальних задач, що відповідають усім матрицям відстаней.

Оскільки задачі оптимізації будуть розглядатися в обчислювальному аспекті, то необхідно позначити подання індивідуальної комбінаторної задачі оптимізації на вході обчислювальної машини. Можна, звичайно, виписати всі припустимі розв'язання й для кожного вказати значення C . Однак для більшості задач деякі індивідуальні задачі будуть мати непропорційно велику кількість припустимих рішень, як це може бути, наприклад, у задачі комівояжера. Надалі будемо припускати, що F і C задані неявно за допомогою двох алгоритмів A^* і A^{**} . Алгоритм A^* по даному комбінаторному об'єкту f і множині параметрів S вирішує, чи є f множиною рішень F , обумовленою даними параметрами. У свою чергу A^{**} за даним припустимим рішенням f та іншою множиною параметрів Q видає значення $C(f)$. Тоді індивідуальну комбінаторну задачу I можна визначити, як подання параметрів S і Q з використанням фіксованого кінцевого алфавіту й деякого стандартного розумного кодування.

Отже, комбінаторна оптимізаційна задача P – це обчислювальна задача, що складається з трьох частин:

- 1) множини D_n індивідуальних задач;
- 2) для кожної $I \in D_n$ кінцевої множини $S_n(I)$ припустимих рішень індивідуальної задачі I ;
- 3) функції m_n , що протиставляє кожній індивідуальній задачі $I \in D_n$ і кожному припустимому рішенню $\sigma \in S_n(I)$ деяке позитивне ціле число, що називається величиною розв'язання σ .

Інакше кажучи, за даним поданням параметрів S і Q для алгоритмів A^* і A^{**} знаходження оптимального припустимого розв'язання називають *оптимізаційним варіантом* даної задачі.

Якщо необхідно за даними S і Q знайти вартість оптимального розв'язання, то такий варіант називають *обчислювальним варіантом* комбінаторної задачі оптимізації. Якщо вартість C просто обчислити, то обчислювальний варіант комбінаторної оптимізації не може бути набагато складніше, ніж оптимізаційний варіант.

Особливо важливий при вивченні обчислювальної складності задачі третій варіант комбінаторної задачі оптимізації. Цей варіант, що називається *варіантом розпізнавання*, має такий вигляд. Для даної індивідуальної задачі I , тобто подання S і Q і цілого числа L , визначити, чи існує таке припустиме розв'язання f , що $C(f) < L$.

На відміну від двох раніше розглянутих варіантів, варіант розпізнавання являє собою питання, на яке можна відповісти «так» чи «ні». Очевидно, відповісти на це питання не набагато складніше, ніж розв'язати відповідну обчислювальну задачу, тому що після її розв'язання залишається тільки порівняти оптимальну вартість $C(f)$ з L і видати відповідь «так» у тому випадку, якщо $C(f) < L$.

Таким чином, використовуючи тільки припущення про те, що C легко обчислити, ми встановили, що кожний з варіантів – оптимізаційний, обчислювальний і розпізнавання – не складніший, ніж попередній. Виникає питання, чи не мають ці варіанти однаково складність. Іншими словами, чи не можна розв'язати обчислювальний варіант ефективно використовуючи гіпотетичний алгоритм, що вирішує задачу у варіанті

розпізнавання й чи не можна зробити то ж саме для оптимізаційного й обчислювального варіантів?

Якщо вартість оптимального розв'язання є цілим числом, алгоритм якого обмежений поліномом від розміру входу, то щораз, коли варіант розпізнавання може бути вирішений ефективно, то ж саме справедливо й для обчислювального варіанта. Слід зазначити, що справедливість припущення про то, що $\log C(f)$ обмежений поліномом від розміру входу, треба, щоб для розв'язання обчислювальної задачі можна ефективно використовувати будь-який алгоритм, що вирішує задачу розпізнавання. Загальний метод для розв'язання оптимізаційного варіанта задачі з використанням алгоритму для обчислювального варіанта не відомий.

Багато задач оптимізації допускають графову постановку й можуть бути ефективно вирішені за допомогою методів теорії графів. Ці ж задачі добре описуються мовою дискретного програмування, у зв'язку з чим їх можна вважати частинними задачами цього напрямку. Облік специфіки задач, а також формулювання їх мовою теорії графів дозволяє одержувати більш прості методи розв'язку задачі на порядок більше, ніж розміри цих же задач у загальній постановці дискретного програмування. З іншого боку, інтерпретація на основі графів забезпечує наочність і допустимість постановки досить складних прикладних задач і методів їхнього розв'язання.

Значення цілочисельних і комбінаторних задач оптимізації швидко зростає. Це пояснюється необхідністю пошуку ефективних рішень усе більш ускладнених технічних, технологічних, організаційних і соціальних проблем, які в багатьох випадках після відповідної формалізації зводяться до задач вибору на кінцевих множинах або до змішаних задач вибору на кінцевих і континуальних множинах.

Перелік прикладних задач цілочисельного й комбінаторного програмування досить різноманітний.

До числа основних типів цих задач, що одержали найбільш широку популярність, належать задачі:

а) про ранець (про рюкзак) – про визначення оптимального набору неподільних об'єктів, що задовольняє обмеження за обсягом або (i) за вагою, або (i) енергоспоживанням й т.д.;

б) про мінімальні покриття в застосуванні до синтезу технічних пристроїв;

в) про максимальне інцидентне сполучення в застосуванні до формування організаційних систем;

г) про призначення – про оптимальний розподіл завдань між виконавцями й у більш широкій постановці про оптимізацію складу засобів і розподілу задач по ланках організаційно-технічного комплексу;

д) транспортного типу;

е) про розміщення – про оптимальне розташування й потужності джерела в мережі;

ж) пошуку оптимальних шляхів у мережах, зокрема пошуку оптимального маршруту обходу n пунктів (задача комівояжера);

з) оптимального складання розкладу при організації технологічних процесів, процесів обслуговування, навчання й ін.

Будь-яка математична модель оптимізації при введенні в обмеження умов цілочисельності стає математичною моделлю цілочисельної оптимізації. Узагальнену модель цілочисельного програмування можна подати в такому вигляді

$$\left. \begin{array}{l} f(\bar{x}) \longrightarrow \min \\ \bar{x} \in \Delta_\beta \\ \Delta_\beta = \{ \bar{x} \mid g_i(\bar{x}) = 0 (i = 1, 2, \dots, l); \\ \bar{x} = (x_1, x_2, \dots, x_n), x_j \in \Psi_1^j (j = 1, 2, \dots, h); \end{array} \right\} \text{ або } \left. \begin{array}{l} f(\bar{x}) \longrightarrow \max \\ \bar{x} \in \Delta_\beta \\ g_i(\bar{x}) \leq 0 (i = l+1, \dots, m); \\ x_j \in \Psi_2^j (j = h+1, \dots, n) \end{array} \right\} \quad (6.1)$$

У випадку, якщо в (6.1)

$$\Psi_1^1 = \Psi_1^2 = \dots = \Psi_1^h = \Psi_1 = \Psi_2 = N = \{0, 1, \dots, k-1\},$$

маємо модель цілочисельного програмування без обмежень на значення цілих чисел.

Якщо $\Psi_1^j = \{0, 1, \dots, k-1\} (j = \overline{1, h})$, $\Psi_2 = \{0, 1, \dots, k-1\}$, то маємо модель k -значного програмування.

Якщо $\Psi_1^1 = \Psi_1^2 = \dots = \Psi_1^h = \Psi_1 = \Psi_2 = \{0, 1\}$, то маємо модель програмування з булевими змінними: якщо $\Psi_1^j = (j = \overline{1, h})$ – модель цілочисельного програмування, а якщо $\Psi_2 = R^l$ – модель частково цілочисельного програмування.

Далі в розд. 6 будуть розглянуті варіанти формалізованого опису відомих практичних задач, які знайшли своє застосування в обчислювальних системах.

6.2. Класифікація методів розв'язання задач дискретної оптимізації

Наявність значних труднощів і специфічних особливостей у розв'язанні задач цілочисельного програмування з булевими змінними (БЗ) породило велику кількість методів і алгоритмів. Огляди з даного напрямку є у вигляді окремих статей і містяться в ряді відомих монографій. Однак найбільш повними є огляди в роботі [86]. З всіх існуючих методів і алгоритмів розв'язання комбінаторних задач виділимо тільки ті, які застосовні до класу задач цілочисельного лінійного програмування (ЦЛП) з БП (рис. 6.1).

КОМБІНАТОРНІ МЕТОДИ. Найбільш простими серед комбінаторних методів є *алгоритми повного перебору*, які виконують перевірку всіх можливих $x = (x_1, \dots, x_n) \in 2^n$, одержуючи тим самим точне розв'язання задачі ЦЛП з БП. Однак перебір всіх 2^n при $n > 50$ неприйнятний навіть на швидкодіючих ЕОМ. Якщо ж n становить кілька сотень, то число потрібних обчислювальних операцій може перевищити поріг Бремермана-Ешбі 10^{100} , що унеможливорює практичне використання даних алгоритмів навіть досить у віддаленому майбутньому.

Адитивні алгоритми, застосовувані для розв'язання задач ЦЛП з БЗ, є найбільш простими й вимагають для своєї реалізації виконання операцій додавання й віднімання (виключаються операції множення й ділення). Існують різні модифікації адитивних алгоритмів. Однак основна ідея побудови таких алгоритмів полягає в проведенні такого зондування підмножин альтернатив, при якому досягаються результати, еквівалентні повному перебору, але без здійснення останнього й за істотно менший час. З цією метою використовується ряд прийомів і відповідні їм адитивні числові процедури, описані в наступному підрозділі. Адитивним алгоритмам властиві всі недоліки методів повного перебору, а також, якщо в оптимальному розв'язанні 80 – 90 % компонентів вектора $x = (x_1, \dots, x_n)$ набувають значення

0 (або 1), то таке розв'язання буде знайдено швидко. Найгіршим випадком для таких алгоритмів є наявність 50 – 60 % нулів (або одиниць), що чергуються між собою. У цьому випадку алгоритм робить майже повний перебір усіх рішень.

Найбільш широко використовуваним у цей час є *метод гілок і меж*, уперше сформульований в алгоритмі Ленда й Дойга [24-27]. До переваг такого методу належать:

- відносна легкість у застосуванні, тому що він не вимагає для своєї реалізації великого об'єму машинної пам'яті;
- простота й видимість обчислювальної схеми;
- кінцевість обчислень, що не має потреби в доведенні, оскільки впливає безпосередньо з самої побудови його обчислювальної схеми.

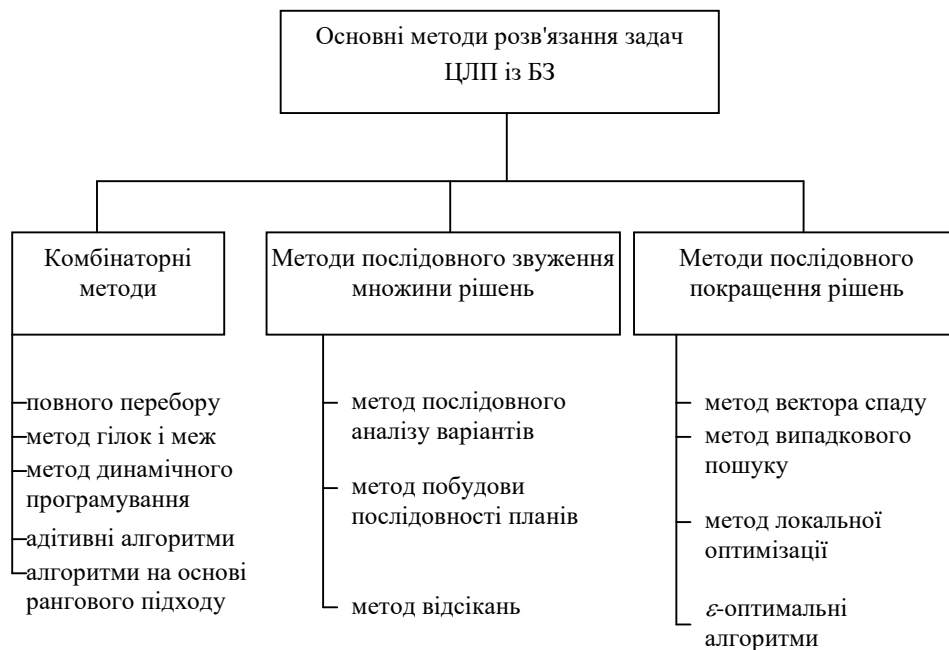


Рис. 6.1. Класифікація методів розв'язання завдання ЦЛП з БЗ

Ефективність розв'язання алгоритмами на основі методу гілок і меж визначається обраними стратегіями розгалуження й підходами до оцінки нижньої (верхньої) границь. Однак, як показують проведені дослідження [22, 23], для точного розв'язання задач ЦЛП виявляються патологічні задачі, обсяг обчислень яких досить близький до повного перебору. У зв'язку з необхідністю розроблення паралельних обчислювальних

алгоритмів для розв'язання задач ЦЛП з БЗ була зроблена спроба реалізації алгоритмів, заснованих на методі гілок і меж, на паралельних обчислювальних структурах (ПВС). При цьому з'ясувалася принципова неможливість такої реалізації, тому що зі збільшенням числа n процесорних елементів виграш за часом можна одержати лише для $n < n^*$, а при $n > n^*$ час розв'язання починає різко збільшуватися через збільшення обсягів міжпроцесорного обміну. Тому ставиться задача знаходження оптимального числа процесорних елементів n^* при розв'язанні методом гілок і меж.

Метод динамічного програмування (ДП) ґрунтується на принципі оптимальності Р. Беллмана: «оптимальна стратегія має таку властивість: яким би не був початковий стан і початкова стратегія, наступні стратегії (розв'язання) повинні бути оптимальні стосовно поточного стану системи». Сутність динамічного підходу полягає в заміні розв'язання даної n -крокової задачі послідовністю задач: однокрокової, двокрокової й т.д. Оптимальне керування системою на кожному кроці роботи алгоритму не залежить від передісторії процесу, тобто як система перейшла в поточний стан, а визначається тільки самим цим станом. Ці багатокрокові процедури й одержали назву «динамічне програмування» [23].

Важливою перевагою методу ДП є його передбачуваність, тобто можливість теоретично оцінити обсяг обчислювальної роботи і ємність заповненої пам'яті. Однак необхідність запам'ятовування на кожному кроці m -компонентного вектора приводить до потреби у великій ємності пам'яті. Ця обставина, що образно називається «прокляттям розмірності» [23], становить основні труднощі в застосуванні апарата динамічного програмування.

Останніми в списку комбінаторних методів (рис. 6.1) подані **алгоритми на основі рангового підходу** до розв'язання задач ЦЛП з БЗ на графах [32-35, 40]. Ранговий підхід до розв'язку задачі (0,1) - рюкзак заснований на розбиванні n -вимірного одиничного куба на n рангів, кожний вектор якого має однакове число одиниць і нулів. До переваг такого методу, безсумнівно, можна віднести: простоту й наочність подання розв'язання задачі; його можливість реалізації на ПВС.

МЕТОДИ ПОСЛІДОВНОГО ЗВУЖЕННЯ МНОЖИНИ РІШЕНЬ. Оптимізаційна схема академіка В.С. Міхайлевича, уведена за назвою *методу послідовного аналізу варіантів* (ПАВ) [35, 36], причому незалежно й практично одночасно з появою методу гілок і меж, зводиться до повторення таких послідовностей перетворень:

- 1) розбиття множини варіантів розв'язання задачі на кілька підмножин, кожна з яких має додаткові специфічні властивість;
- 2) використання специфічних властивостей підмножин розв'язання задач для пошуку логічних протиріч в описі окремих підмножин;
- 3) виключення з подальшого розгляду тих підмножин варіантів розв'язання, в описі яких є логічні протиріччя.

У цілому метод ПАВ для задачі ЦЛП з БЗ близький до адитивних алгоритмів, тому що він заснований на аналізі системи обмежень задачі. Однак сам аналіз здійснюється різними шляхами.

У загальному випадку метод ПАВ включає (як окремі випадки) схеми методів ДП і гілок і меж. Тому всі недоліки, властиві цим методам, у загальному випадку не усуваються в методі ПАВ.

Помітне застосування в дискретній оптимізації знайшов *метод побудови послідовності планів*, загальний формалзм якого розроблений В.А. Ємелічевим. Ідея цього методу полягає в заміні вихідної екстремальної задачі більш простою, допоміжною задачею й побудові послідовності планів цієї задачі в порядку «погіршення» значень оцінної функції (міноранти або мажоранти).

Серед точних методів розв'язання задач дискретної оптимізації значне поширення одержав *метод відсікання*. Ідея методу полягає ось у чому. Спочатку розв'язується задача ЦЛП при відкиданні умови цілочисельності змінних. Одержуване в такий спосіб розв'язання є оптимальним, якщо воно цілочисельне. Якщо ж оптимальне розв'язання не цілочисельне, то до неї додатково додається нове обмеження, що полягає в такому: множина припустимих рішень знову отриманої задачі містить будь-яке припустиме розв'язання попередньої задачі й не містить оптимального розв'язання попередньої задачі.

Обмеження, що додаються в даному процесі, називають відсіканнями, тому що вони відтинають певну частину множини припустимих рішень задачі ЦЛП.

До недоліків методу відсікань належить непередбачуваність його поведіння, необхідність обліку все більшого числа обмежень, слабка пристосованість до розв'язання частково цілочисельних задач, чутливість до похибок округлення, а також слабке застосування для розв'язання задач ЦЛП з БЗ через існування більш ефективних методів точного їхнього розв'язання.

МЕТОДИ ПОСЛІДОВНОГО ПОКРАЩЕННЯ РІШЕНЬ.

Основною ідеєю методів і алгоритмів послідовного покращення рішень є така організація пошуку на множині альтернатив, що не звужуються і при якій поступово виділялися б найкращі припустимі розв'язання. Основу цього напрямку становлять наближені методи: метод локальної оптимізації, метод вектори спаду, метод випадкового пошуку.

Метод локальної оптимізації спрямований на побудову ітеративних алгоритмів послідовного поліпшення рішень і розроблений в [36-39]. Істотний вплив зробили роботи з дискретної математики [23-30], що привело до побудови ефективних наближених алгоритмів широкого кола дискретних задач, які з самого початку орієнтуються на існування багато екстремальних рішень. Алгоритми цього методу є вузькоспеціалізовані, і тому не знайшли широкого універсального застосування.

До наближених методів належить *метод вектори спаду*, запропонований І.В. Сергієнко. Застосування цього методу до задач ЦЛП з БЗ дало істотний виграш у порівнянні з адитивними алгоритмами й алгоритмами, заснованими на методі ПАВ. Недоліком цього методу є істотна залежність точності й збіжності від обраного радіуса наближень.

У рішенні задач ЦЛП з БЗ подальший розвиток одержали *методи випадкового пошуку*. Алгоритми цих методів характеризуються тим, що поступове поліпшення досягається шляхом відбору придатних альтернатив з послідовності альтернатив, виділюваних з використанням деякого ймовірнісного механізму вибору на множині альтернатив. При

цьому розрізняють алгоритми керованого й некерованого випадкового пошуку. Застосування цих алгоритмів найбільш доцільно використовувати при побудові початкового наближення для роботи інших алгоритмів. Для задачі ЦЛП з БЗ ефективний алгоритм випадкового пошуку побудований у [24], що дає можливість у процесі її розв'язання змінювати функції розподілу ймовірностей і в підсумку зменшувати час одержання прийнятного розв'язання. Недоліком такого алгоритму є утруднення визначення часу виконання в «гіршому» випадку, що не дозволяє його використовувати в масштабі реального часу.

У роботах [30-32, 34] показано, що для задач ЦЛП з БП існують поліноміальні асимптотичні ϵ -оптимальні алгоритми випадкового пошуку, які дозволяють за кінцеве число кроків знаходити наближене розв'язання для одновимірної задачі. Недоліком методу випадкового пошуку, як і для методу гілок і меж, є принципові труднощі реалізації цих алгоритмів на n -процесорних ПОС. Найбільш повна класифікація сучасних методів розв'язання задач дискретної оптимізації наведена в роботі [23].

Таким чином, з погляду розпаралелювання найкращими є методи, засновані на ранговому підході, яким надалі й буде приділена увага.

Показники ефективності послідовних і паралельних алгоритмів

Основні поняття й визначення, пов'язані з поданням алгоритмів у вигляді графів, дані в [27, 28]. Для коректного порівняння розроблюваних алгоритмів з відомими необхідно визначити основні показники їхньої ефективності, що дозволяють проводити порівняльний аналіз.

Найбільш широко застосовуваним показником ефективності роботи алгоритму є відрізок часу, затрачений алгоритмом для розв'язання поставленої задачі. Однак такий підхід не однозначно визначає ефективність алгоритмів, тому що цей показник залежить від типу ЕОМ, на якому виконується алгоритм, від вибору мов програмування й т.д. При аналізі алгоритмів час їхньої роботи буде виражатися в термінах числа

елементарних кроків (арифметичних операцій, порівнянь і т.п.), необхідних для виконання цього алгоритму на гіпотетичній обчислювальній машині. Інакше кажучи, виконання будь-якої операції вимагає одну одиницю часу.

Однак число кроків, виконуваних алгоритмом, не однакове для різного розміру входу вихідних даних. Для усунення такої неоднозначності в поводженні алгоритму при переході від одного входу до іншого будемо розглядати всі входи даного розміру n разом і визначимо складність алгоритму для входу як число кроків алгоритму в найгіршому разі по всіх входах. У комбінаторних задачах оптимізації входом є комбінаторний об'єкт: граф, множина цілих чисел, сімейство кінцевих множин і т.д. Щоб увести цей вхід в обчислювальній системі для розв'язання, необхідно яким-небудь чином закодувати або подати у вигляді послідовності символів над деяким фіксованим алфавітом, таким як двійковий алфавіт. Кодування комбінаторних об'єктів можна зробити кожним з багатьох відомих способів [26-28]. Оскільки вхід алгоритму подається у вигляді послідовності або ланцюжка символів, визначимо розмір входу як довжину цієї послідовності, тобто число символів у ній. При аналізі алгоритмів звичайно цікавить швидкість росту складності алгоритму [27].

Визначення. Нехай $f(n)$ і $g(n)$ – функції, які визначені на множині цілих позитивних чисел і набувають позитивні дійсні значення.

1. Позначимо

$$f(n)=O(g(n)), \quad (6.2)$$

якщо існує така константа $c > 0$, що $f(n) \leq Cg(n)$ для досить великих n .

2. Позначимо

$$f(n)=O(g(n)), \quad (6.3)$$

якщо існує така константа $c > 0$, що $f(n) \geq Cg(n)$ для досить великих n .

3. Позначимо

$$f(n) = \Theta(g(n)), \quad (6.4)$$

якщо існують такі константи $c, c' > 0$, що $Cg(n) \leq f(n) \leq C'g(n)$ для досить великих n .

Замість (6.4) можна записати

$$f(n) \cup \cap g(n). \quad (6.5)$$

Відношення (6.5) є відношенням еквівалентності. Клас еквівалентності цього відношення, що містить $f(n)$ (тобто множина всіх функцій $g(n)$, таких, що $f(n) = \Theta(g(n))$), називають швидкістю росту $f(n)$ [27]. Завдяки уведеному поняттю, швидкість росту складності алгоритму, як показано у [27], можна оцінити зверху, використовуючи вираз типу «вимагає часу $O(n^3)$ ».

Таким чином, тимчасова складність алгоритму відбиває витрати часу, що вимагаються для нього. Функція, яка вхідній довжині n ставить у відповідність максимальний час, не буде повністю визначена доти, поки не зафіксована схема кодування, обрано обчислювальний пристрій, що визначає час роботи. Той самий алгоритм A можна виконувати за різний час на різних структурах. Тому при переході від однієї структури до іншої часто користуються поняттям ємнісної складності обчислювальної системи, що характеризує ріст апаратурних витрат при такому переході. Апаратурні витрати можуть задаватися в числі елементарних процесорних елементів, кількості зв'язків у системі, ємності пам'яті й т.д., при цьому використовується вираз «потрібно $O(n^3)$ комірок пам'яті» [27]. При аналізі складності алгоритмів оцінка в найгіршому випадку не завжди об'єктивно несе інформацію про час, затрачуваний алгоритмом на розв'язання задачі, оскільки гірші випадки, при рішенні конкретної задачі, можуть бути рідкими. Тому в роботі для кожного алгоритму теоретично визначена складність у найгіршому випадку, а при експериментальному моделюванні роботи алгоритмів проводяться оцінки їхньої складності в

середньому, що дозволяє більш об'єктивно оцінити можливості алгоритму.

Як основні характеристики наближених алгоритмів використовують абсолютну Δf і відносну δf похибки отриманого наближеного розв'язання, які визначаються зі співвідношень:

$$\Delta f = |f(x) - f(x^*)|;$$
$$\delta f = \frac{|f(x) - f(x^*)|}{f(x^*)}, \quad (6.6)$$

де f – цільова функція, визначена на деякій множині M ;
 $f(x)$ – припустиме розв'язання задачі, що є наближенням;
 $f(x^*)$ – оптимальне розв'язання задачі.

Таким чином, як основні критерії ефективності для точних алгоритмів розв'язку задачі ЦЛП з БП будемо використовувати функцію тимчасових і апаратних витрат, для наближених алгоритмів – відносну δf похибку. Показники якості, пов'язані з ефективністю розпаралелювання розроблюваних алгоритмів, будуть уведені безпосередньо при їхньому порівнянні з існуючими.

6.3. Ранговий підхід до розв'язку задачі (0, 1) – рюкзак

6.3.1 Формальна модель n -вимірною одиничного куба в ранговому підході

Розглянемо сутність рангового підходу до задачі цілочисельного лінійного програмування (ЦЛП) з булевими змінними (БЗ) на прикладі задачі про рюкзак. Загальна постановка цієї задачі формулюється в такий спосіб. Необхідно знайти вектор \vec{x} , що доставляє максимум функції

$$f(\vec{x}) = \sum_{j=1}^n c_j \cdot x_j, \quad (6.7)$$

при виконанні умов

$$\sum_{j=1}^n a_{ij} \cdot x_j \leq b_i, \quad (6.8)$$

$$x_j \in \{0,1\}, \quad i = (\overline{1,m}); \quad j = (\overline{1,n}). \quad (6.9)$$

Для спрощення викладу математичної моделі розглянемо одновимірну задачу, тобто максимізуємо функціонал

$$f(\vec{x}) = c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n, \quad (6.10)$$

при обмеженнях

$$\sum_{j=1}^n a_{1j} \cdot x_j \leq b, \quad c_1 \geq c_2 \geq \dots \geq c_n; \quad a_{ij} > 0; \quad c_j > 0; \quad j = (\overline{1,n}). \quad (6.11)$$

Поставимо у відповідність (6.10), (6.11) граф G (рис.6.2), що являє собою бінарне дерево всіх розв'язків, число яких дорівнює 2^n .

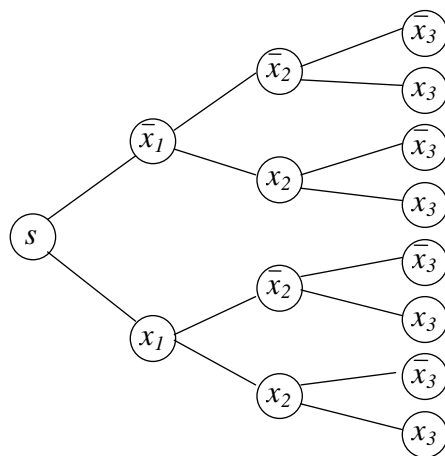


Рис. 6.2. Граф G

Множина $X = \{x_j\}$ всіх векторів розмірності n , усі компоненти яких $x_j \in \{0,1\}$ становлять множини можливих значень. Деяка її підмножина V , усі вектори якої задовольняють (6.11), утворить множини припустимих рішень.

Множина $H \subset V$ є множиною оптимальних рішень вихідної задачі, якщо для будь-яких векторів $x \in H$ функціонал (6.5) досягає свого екстремального значення.

Усю множину можливих рішень можна розбити на групи векторів, що містять: один компонент $x_j = 1$, $j = \overline{1, n}$ і всі інші рівні 0; два компоненти $x_j = 1$ і всі можливі їхні сполучення по 2, а інші – рівні 0; три компоненти $x_j = 1$ і всі можливі їхні сполучення й т.д. n -компонент $x_j = 1$. Якщо позначити підмножини векторів цих груп через m^r $r = \overline{1, n}$, тоді множину всіх можливих рішень можна записати як об'єднання підмножин m^r

$$X = \bigcup_{j=1}^n m^r. \quad (6.12)$$

Як показано у [32], за графом G можна побудувати граф G' (рис.6.3), у якому множина шляхів рангу r (ранг шляху – число ребер, що утворюють шлях) відповідає групам підмножин, описуваних співвідношенням (6.12). Для цього вершину s з'єднуємо спрямованими ребрами з вершинами $1, 2, \dots, n$ і т.д.; вершину i з'єднуємо з вершинами $i+1, \dots, n$. В останню вершину n входять ребра, спрямовані з усіх вершин, і жодне ребро з цієї вершини не виходить.

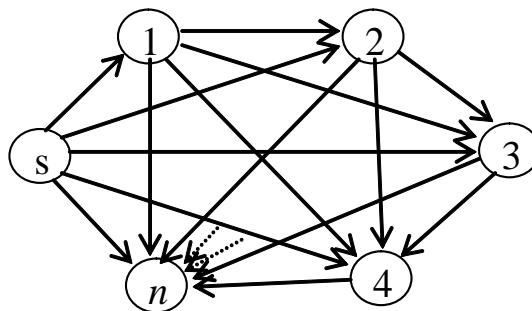


Рис. 6.3. Граф G'

Дерево шляхів DA графа G' з вершини s будується таким чином [32]: на нульовому ярусі ($r=0$) розташуємо вершину s . На першому ярусі розмістимо всі вершини графа G' , що мають зв'язок з вершиною s , і з'єднаємо їх з s (при цьому утворилася

підмножина шляхів рангу $r=1$). У другому ярусі розмістяться всі вершини, що мають зв'язок з вершинами першого ярусу, без вершини з номером 1 і з'єднаємо їх з вершинами першого ярусу (утворені всі шляхи рангу $r=2$) і т.д. доти, поки в останньому не залишиться одна вершина n . На рис.6.4 $n = 4$.

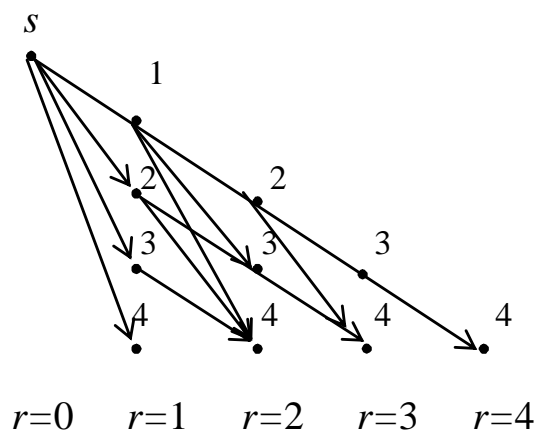


Рис. 6.4. Граф $D\Delta$

Геометрично вершина k графа $D\Delta$ рангу r – це множина векторів \vec{x} $(x_1, x_2, \dots, x_k, \dots, x_n)$, у яких $x_k = 1$, а на позиціях від 1 до k знаходиться r одиниць (рис.6.5). Ребру, що входить у вершину k графа $D\Delta$, відповідає одиничний вектор \vec{e}_k $(0, 0, \dots, 0, 1, 0, \dots, 0)$ n -вимірного одиничного куба B^n з одиницею в k -й позиції μ_{sj}^r , рангу r у графі $D\Delta$ відповідає вектор \vec{x} , що дорівнює сумі одиничних векторів ребер, за якими він досяг вершини j рангу r , починаючи з вершини s . Наприклад, шляху $\mu_{s24}^{r=2}$ відповідає вектор $\overline{x_{s24}}$, що утворюється сумою нульового вектори $\vec{0}$ $\{0000\}$ і одиничних векторів $\vec{e}_2 = \{0100\}$, $\vec{e}_4 = \{0001\}$, тобто

$$\overline{x_{s24}} = \vec{0} \{0000\} + \vec{e}_2 \{0100\} + \vec{e}_4 \{0001\} = \{0101\}$$

Нехай у графі $D\Delta$ кожному ребру, що входить у вершину j , $j = (\overline{1, n})$ відповідає дві ваги: вага c_j , що дорівнює коефіцієнту при x_j у функціоналі (6.10), і вага a_{1j} , що дорівнює коефіцієнту при x_j в обмеженні (6.11). Тоді шлях μ_{sj}^r у графі $D\Delta$ з вершини s у

вершину j характеризується двома довжинами: $d_c(\mu_{sj}^r)$ – довжиною за вагами функціонала й $d_a(\mu_{sj}^r)_1$ – довжиною за вагами обмежень.

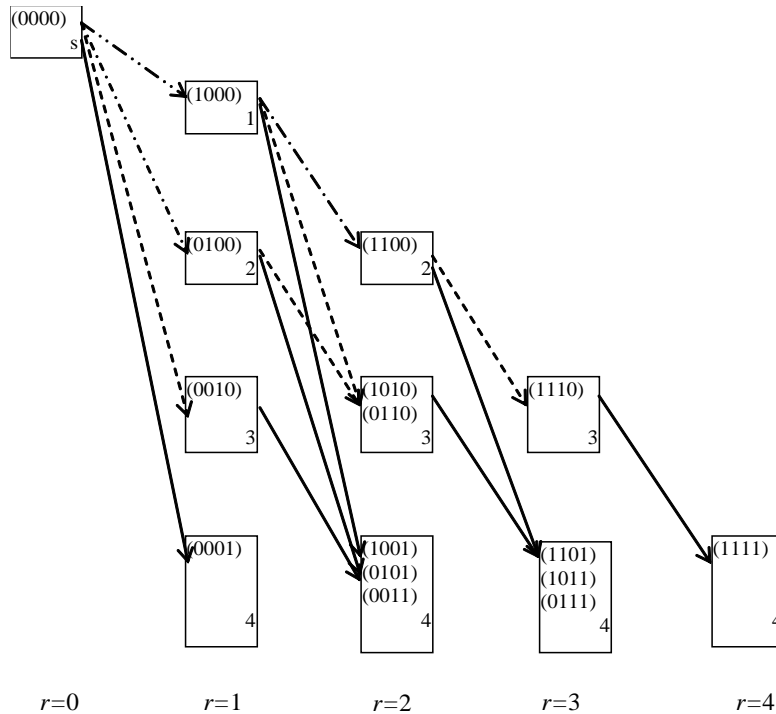


Рис. 6.5. Геометрична інтерпретація графа $D\Delta$

Множину шляхів $m_s^r(j)$ у графі $D\Delta$ до вершин j , розташованих на ярусах $r = (\overline{1, n})$ від вершини s , можна подати у вигляді

$$m_s^r(j) = m_{sj}^{r=1} \cup m_{sj}^{r=2} \cup \dots \cup m_{sj}^{r=n}, \quad j = (\overline{1, n}), \quad (6.13)$$

де m_{sj}^r – множина шляхів у графі $D\Delta$ від вершини s до вершин j , розташованих на r -х ярусах графа $D\Delta$ (ранг шляху $\mu_{sj}^r \in m_{sj}^r$ визначається числом ребер, що утворюють цей шлях). Слід мати на увазі, що множина шляхів $m_{sj}^{r=k}$ у графі $D\Delta$ відповідає множині векторів $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_v\}$, що містять k одиниць. Отже, $|m_{sj}^r| = C_n^{r=k}$, тобто кожному шляху в множині $m_{sj}^{r=k}$ відповідає деякий вектор (x_1, x_2, \dots, x_n) . З (6.13) випливає

$$|m_s^r(j)| = C_n^{r=1} + C_n^{r=2} + \dots + C_n^{r=n} = 2^n - 1. \quad (6.14)$$

Таким чином, граф $D\Delta$ являє собою впорядкований за рангами еквівалент n -вимірному одиничного куба B^n , у якому шляхи $\mu_{sj}^r \in m_{sj}^r$ відповідають вершинам B^n . Довжина кожного шляху за вагою функціонала визначає значення функціонала (6.10) у вершинах одиничного куба B^n . Довжина за вагами обмежень визначає, чи відповідає дана вершина B^n обмеженням (6.11), тобто належить вершина n -вимірному одиничного куба B^n гіперплощини (6.3.5). Якщо $d_a(\mu_{sj}^r)_1 \leq b$, то вершина належить гіперплощині (6.11) і будемо говорити, що шлях μ_{sj}^r задовольняє властивість v . Якщо $d_a(\mu_{sj}^r)_1 > b$, то вершина n -вимірному куба, що відповідає шляху μ_{sj}^r , не належить гіперплощині (6.11), а шлях μ_{sj}^r вважаємо таким, що не задовольняє властивість v .

Оптимальному рішенню задачі (6.10),(6.11) в D (відповідає найдовший шлях за вагами функціонала, що задовольняє властивість v).

У випадку m -вимірної задачі (6.7) - (6.9) ребрам, що входять у вершини графа $D\Delta$, крім ваги c_j функціонала, відповідає m ваг a_{ij} обмежень, а шлях μ_{sj}^r характеризується довжинами: $d_c(\mu_{sj}^r)$ – довжиною за вагами функціонала і $d_a(\mu_{sj}^r)_i$, $i = \overline{(1, m)}$ – довжинами за вагами m обмежень.

6.3.2. Принцип оптимізації за напрямком в n -вимірному одиничному кубі й узагальненій процедурі розв'язання ЦЛП з БП на основі ідеї рангового підходу

На основі розглянутої математичної моделі рангового підходу для побудови алгоритмів розв'язання задач ЦЛП з БП покладений принцип оптимізації за напрямком у дискретному просторі станів, заданим графом $D\Delta$ [32 - 35]. Подання n -вимірному одиничного куба у вигляді графа $D\Delta$ дозволяє розбити множину всіх шляхів графа $D\Delta$ з нульової вершини s

на Ω локальних областей, де $|\Omega|$ не перевищує величину $\frac{n^2}{2}$, оскільки число вершин у графі $D\Delta$ визначається сумою чисел натурального ряду

$$\Omega = n + (n-1) + \dots + 1 = \frac{n \cdot (n+1)}{2} \approx \frac{n^2}{2}, \quad (6.15)$$

причому Ω -області в графі $D\Delta$, упорядковані за рангами і шляхом наступного рангу, можуть бути отримані на основі шляхів попереднього рангу за рахунок приєднання до них ребра (j, p) у графі $D\Delta$

$$m_{sp}^{r=r+1} = \{(\forall(\mu_{sj}^r \in m_{sj}^r)) \cup (j, p)\}$$

Нехай задані деякі правила відсікань $\{L_w\}$ шляхів μ_{sj}^r у множинах m_{sj}^r . Тоді, якщо в множинах містяться шляхи, що задовольняють властивість v і правила $\{L_w\}$, то під *оптимізацією за напрямком* у графі $D\Delta$ до вершини p будемо розуміти формування множин $m_{sp}^{r=r+1}$ наступного рангу, які виходять за рахунок виділення в m_{sj}^r шляхів, приєднання до яких ребра (j, p) дозволить у множині $m_{sp}^{r=r+1}$ одержати шляхи, що задовольняють правила $\{L_w\}$ на основі такого рекурентного співвідношення:

$$\overline{\forall(\mu_{sj}^r \in m_{sj}^r)} \left[\mu_{sp}^{r=r+1} = L_w \{ \mu_{sj}^r \cup (j, p) \} \right]; \quad (6.16)$$

$$p = (\overline{r+1}, n); j = (\overline{r}, n),$$

де $\mu_{sj}^r \cup (j, p)$ – шлях з вершини s графа $D\Delta$ у вершину p , що проходить через проміжну вершину j і задовольняє правила $\{L_w\}$, тобто одержуваний за рахунок приєднання до шляху μ_{sj}^r ребра (j, p) , якщо таке з'єднання не суперечить правилам $\{L_w\}$. Надалі для спрощення викладу, якщо шлях $\mu_{sp}^{r=r+1} = \mu_{sj}^r \cup (j, p)$ задовольняє правила $\{L_w\}$, то будемо говорити, що він задовольняє й властивість v .

Таким чином, для розв'язку задачі (6.7), (6.8), використовуючи правила $\{L_w\}$ і оптимізацію за напрямком (6.16), побудуємо деяку узагальнену процедуру A_0 , що дозволяє формувати множини локальних екстремумів Ω і виділяти серед них глобальний.

Уведемо узагальнену процедуру A_0 , що дозволяє на основі обраного правила відсікань $\{L_w\}$ розв'язувати задачу (6.7) – (6.9).

Узагальнена процедура A_0

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = \overline{(1, n)}$, що задовольняють властивість ν . Виділяються шляхи $\mu_{sj}^{* r=1}$, що визначають локальні екстремуми областей Ω_j .

КРОК 2. Формуються множини шляхів $m_{sp}^{r=r+1}$ $p = \overline{(r+1, n)}$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу відповідно до рекурентних співвідношень (6.16). В утворених множинах $m_{sp}^{r=r+1}$ здійснюється відсікання шляхів відповідно до обраного правила відсікань $\{L_w\}$ і виділяються шляхи $\mu_{sp}^{* r=r+1}$, що визначають локальні екстремуми областей Ω_p .

КРОК 3. Перевіряємо, чи $m_{sp}^{r=r+1}$ всі множини наступного рангу порожні. Якщо це так, то переходимо до кроку 4, якщо ні, то перевіряємо $r=(n-1)$. У випадку виконання рівності переходимо до кроку 4, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 4. Виділяємо серед множин локальних екстремумів Ω_j $j = \overline{(1, n^2 / 2)}$ глобальний і процедура A_0 закінчує роботу.

Узагальнена процедура A_0 дозволяє визначити локальні екстремуми в Ω -областях графа $D\Delta$ щораз на кроці 2 і потім на кроці 4 виділяти глобальний екстремум з $n^{2/2}$ локальних, отриманих на основі принципу оптимізації за напрямком (6.16) з використанням правил відсікань, що вводяться, $\{L_w\}$ шляхів у m_{sj}^r множинах .

Таким чином, з поданої математичної моделі n -вимірного одиничного куба B^n у вигляді графа $D\Delta$ і сформульованого принципу оптимізації за напрямком на основі рангового підходу впливають нижченаведені задачі.

1. Визначення стратегій відсікань $\{L_w\}$ безперспективних шляхів у множинах m_{sj}^r , що приводять до наближених і точних рішень задачі ЦЛП з БЗ (6.7) - (6.9).

2. Побудови наближених і точних алгоритмів на основі обраних правил відсікань $\{L_w\}$ для розв'язання одновимірних і багатовимірних задач ЦЛП з БЗ.

3. Створення паралельних обчислювальних структур як спеціалізованих пристроїв для розв'язання даного класу задач дискретної оптимізації.

6.3.3. Правила відсікання безперспективних варіантів рішень на основі узагальненої процедури рангового підходу

В основу методу відсікання безперспективних варіантів для задачі (0,1) - рюкзак покладемо множину стратегій $\{L_w\}$, застосування яких до узагальненої процедури A_0 приведе до побудови алгоритмів розв'язання цієї задачі цілочисельного лінійного програмування (ЦЛП) з булевими змінними (БЗ).

Найбільш простою стратегією відсікань L_1 при формуванні $r=r+1$ шляхів наступного рангу в множинах $m_{sp}^{r=r+1}$ на основі процедури A_0 є виділення в m_{sj}^r шляхів максимальної довжини за вагами функціонала c_j , довжини яких за вагами обмежень не перевищують величини b_i , тобто $d_a(\mu_{sj}^r)_i \leq b_i$. У цьому випадку шляхами, що задовольняють властивість v , будемо називати шляхи μ_{sj}^r , довжини яких за вагами обмежень не перевищують величину b_i . Тоді рекурентне співвідношення (6.16), що відповідає стратегії L_1 , набуває виду

$$\mu_{sp}^{r=r+1} = \max_{c_j} \{ \mu_{sj}^r \cup (j, p) \}; \quad p = (\overline{r+1, n}); \quad j = (\overline{r, n}). \quad (6.17)$$

Приклад 6.1. Потрібно максимізувати функціонал

$$f(\vec{x}) = 20x_1 + 18x_2 + 18x_3 + 15x_4 + 10x_5 + 4x_6 + 1x_7$$

при обмеженні

$$7x_1 + 2x_2 + 5x_3 + 2x_4 + 3x_5 + 1x_6 + 4x_7 \leq 12$$

у відповідності до стратегії L_1 .

Розв'язання. Побудуємо граф $D\Delta$ (вершину s , що відповідає тривіальному рішенняю задачі, показувати не будемо). Шлях μ_{sj}^r , $j = (\overline{r, n})$ задається ідентифікатором, що визначає через які вершини попередніх рангів він проходить (рис. 6.6). Наприклад, ідентифікатор $s146$ указує на шлях $\mu_{s146}^{r=3}$, що проходить через вершини $s \rightarrow 1 \rightarrow 4 \rightarrow 6$ у графі $D\Delta$ (рис. 6.4). У дужках на першому місці стоїть довжина шляху за вагами функціонала, а далі m ваг за першим, другим, і т.д. m -м обмеженнями. Так, запис "s146(39,10)" відповідає шляху $\mu_{s146}^{r=3}$, що має $d_c(\mu_{s146}^{r=3}) = 39$, а $d_a(\mu_{sj}^r)_1 = 10$. Шлях μ_{sj}^r , що належить оптимальному рішенняю, позначений "*".

Ітерація 1. Будуємо шляхи рангу $r = 1$, що задовольняють обмеження задачі. Таких шляхів у графі $D\Delta$ – сім: $\mu_{s1}^{r=1} - \mu_{s7}^{r=1}$. Серед них вибираємо шлях з найбільшою довжиною за вагами функціонала: $\mu_{s1}^{r=1}$. Позначимо його μ_{s1}^* .

Ітерація 2. Будуємо шляхи рангу $r = 2$ відповідно до принципу оптимізації за напрямком (6.16) і стратегією L_1 (6.17). Наприклад, сформуємо шлях $\mu_{s13}^{r=2}$ на основі шляху $\mu_{s1}^{r=1}$ додаючи ребро (1,3). При цьому довжина шляху буде: за вагами функціонала $d_c(\mu_{s13}^{r=2}) = 38$, а за вагами обмеження $d_a(\mu_{s13}^{r=2})_1 = 12$, тобто шлях $\mu_{s13}^{r=2}$ задовольняє обмеження задачі й властивість v . Аналогічно будуються інші шляхи рангу $r = 2$. Потім серед усіх шляхів рангу 2 вибираємо шлях з найбільшою довжиною $d_c(\mu_{sj}^{r=2})$. Це шляхи: μ_{s12}^* і μ_{s13}^* .

Ітерація 3. Будуємо шляхи рангу $r = 3$ на основі шляхів рангу $r=2$ і (6.16), (6.17). При додаванні до шляху $\mu_{s_{12}}^{r=2}$ ребра (2,3) утвориться шлях $\mu_{s_{123}}^{r=3}$, довжина якого за вагами обмежень $d_a(\mu_{s_{123}}^{r=3})_1 = 14$, що не задовольняє умову задачі. Тому такий шлях не показаний. Якщо в множині $m_{s_j}^{r=2}$ кілька шляхів, більший з яких при приєднанні якого-небудь ребра не задовольняє властивість ν , то згідно з (6.16) вибираємо наступний шлях, що належить цій множині, але менший за вагами функціонала. На $r=2$ (рис. 6.6) це можна спостерігати при побудові шляху з множини $m_{s_3}^{r=2}$ в. $m_{s_3}^{r=3}$. Серед усіх шляхів рангу 3 вибираємо шлях з найбільшою довжиною $d_c(\mu_{s_j}^{r=3})$. Це шлях: $\mu_{s_{124}}^{r=3}$.

s1(20,7)							
1							
s2(18,2)* 2	s12(38,9)	2					
s3(18,5) 3	s13(38,12)						
	s23(36,7)* 3		3				
s4(15,2)	s14(35,9)	s124(53,11)					
	s24(33,4)	s234(51,9)* 4					
4	s34(33,7)		4				
s5(10,3)	s15(30,10)	s125(48,12)	s2345(61,12)*				
	s25(28,5)	s235(46,10)					
5	s35(28,8)	s145(45,12)		5	5		
	s45(25,5)						
s6(4,1)	s16(24,8)	s126(42,10)	s1246(57,12)				
	s26(22,3)	s236(40,8)	s2356(50,11)				
	s36(22,6)	s146(39,10)					
	s46(19,3)	s156(34,11)					6
6	s56(14,4)			6	6		
s7(1,4)	s17(21,11)	s237(37,11)	s2367(41,12)				
	s27(19,6)	s247(34,8)					
	s37(19,9)	s257(29,9)					
	s47(16,6)	s167(25,12)					
	s57(11,7)						
7	s67(5,5)			7	7	7	7
r=1	r=2	r=3	r=4	r=5	r=6	r=7	

$\mu_{s_{2345}}^{r=4} \rightarrow \bar{x}^{no} = (0111100)$
 $f(\bar{x}^{no}) = 61.$

Рис.6.6. Ілюстрація роботи стратегії L_1

Ітерація 4. Будуємо шляхи рангу $r = 4$ аналогічно описаним в ітерації 3. Серед усіх шляхів рангу 4 також вибираємо шлях з найбільшою довжиною $d_c(\mu_{s_j}^{r=4})$. Це шлях: $\mu_{s_{2345}}^{r=4}$.

Ітерація 5. Будуємо шляхи рангу $r = 5$. Як видно з рис. 6.6, шляхів, що задовольняють умову задачі рангу $r = 5$, немає. Тоді з

усіх шляхів μ_{sj}^{*r} , $r = (\overline{1,4})$, $j = (\overline{r,n})$ вибираємо найбільший:
 $\mu_{s_{2345}}^{*r=4}$, а вектор $\bar{x}^{opt} = (0111100)$, йому відповідний, є оптимальним розв'язанням задачі.

Використання стратегії L_1 не завжди дає оптимальне розв'язання задачі. Виникнення ситуації відсікання шляхів за (6.17), що відповідають оптимальному рішенням задачі, обумовлене тим, що шляхи з більшим значенням за функціоналом можуть набирати й більшу довжину за вагами обмежень. Отже, на деякому ранзі виникає ситуація, коли через обмеження шляхи наступних рангів у графі $D\Delta$ побудувати неможливо, а шляхи з меншим значенням довжини за вагами функціонала, що мають і меншу довжину за вагами обмежень, відкинуті за стратегією L_1 . На їхній основі могли б бути отримані шляхи більшого рангу й за рахунок цього й більшої величини за вагами функціонала.

Позначимо через $\mu_{sj=\gamma}^{**r=q}$ шлях рангу $r=q$ з найбільшою довжиною $d_c \left(\mu_{sj}^{**r} \right)$ серед шляхів $\left\{ \mu_{sj}^{*r<q} \right\}$ рангу $r < q$, що відповідають локальним екстремумам, обумовленим процедурою A_0 з використанням стратегії L_1 . Тоді сформулюємо правило відсікання L_2 для одновимірної задачі на основі теореми 6.1.

Теорема 6.1. У графі $D\Delta$ не існує шляху, що задовольняє властивість v рангу $r < q$, для якого виконується нерівність

$$d_c \left(\mu_{sj}^{*r<q} \right) < d_c \left(\mu_{sj=\gamma}^{**r=q} \right),$$

Доведення. На кожному етапі формування множин шляхів $m_{sp}^{r=r+1}$ $p = (\overline{r,n})$ процедурою A_0 , що використовує правило L_1 , у цих множинах будуються шляхи $\mu_{sp}^{r=r+1}$, що задовольняють властивість v , відповідно до (6.17). Припустимо, що в $D\Delta$ існує

ШЛЯХ $\mu_{sj=k}^{-r<q}$, ДОВЖИНА ЯКОГО $d_c\left(\mu_{sj=k}^{-r<q}\right) < d_c\left(\mu_{sj=\gamma}^{**r=q}\right)$, І ВІН ЗАДОВОЛЬНЯЄ ВЛАСТИВІСТЬ ν . Існування такого шляху можливе, якщо існують шляхи $\left\{\mu_{sj}^{-r<q}\right\}$, що задовольняють властивість ν , довжини яких $d_c\left(\mu_{sj}^{-r<q}\right)$ перевищують довжини шляхів $d_c\left(\mu_{sj}^{*r<q}\right)$, відповідних локальним екстремумам, виділеним процедурою A_0 з використанням стратегії L_1 . Покажемо, що дане припущення неправильне. Для цього доведемо нижченаведене твердження.

Твердження 6.1. Процедура A_0 з використанням стратегії L_1 дозволяє визначати в графі $D\Delta$ шляхи рангу $r < q$ максимальної довжини за вагами функціонала, що задовольняють властивість ν , від вершини s до всіх інших вершин графа $D\Delta$.

Якщо правильне твердження 6.1, то, значить, не існує шляхів $\mu_{sj}^{-r<q}$, що задовольняють нерівність $d_c\left(\mu_{sj}^{-r<q}\right) > d_c\left(\mu_{sj}^{*r<q}\right)$,

й, отже, припущення про існування шляху $\mu_{sj=k}$, для якого виконується нерівність $d_c\left(\mu_{sj=k}^{r<q}\right) > d_c\left(\mu_{sj}^{**r=q}\right)$, тож неправильне й теорема доведена.

Правомірність твердження 6.1 для множин $m_{sj}^{r=1}$, $m_{sj}^{r=2}$ очевидна, оскільки множини $m_{sj}^{r=1}$ містять по одному шляху рангу $r=1$, а в множинах $m_{sj}^{r=2}$ містяться всі шляхи $\mu_{sj}^{r=2}$ рангу $r=2$ графа $D\Delta$. Нехай на основі множин $m_{sj}^{r=2}$, $j = (\overline{r, n})$ процедурою A_0 з використанням стратегії L_1 сформовані множини $m_{sj}^{r=3}$.

Виділимо в $m_{sj}^{r=3}$ шляхи μ_{sj} максимальної довжини за вагами функціонала. Припустимо, що в графі $D\Delta$ існує шлях μ_{sj} , що задовольняє властивість ν , але довший, ніж μ_{sj} . Відповідно до рекурентних співвідношень (6.17) останнє можливе, якщо в множинах $m_{sj}^{r=2}$ є шляхи μ_{sj} довщі, ніж шляхи μ_{sj} . Але це суперечить раніше встановленому факту, що шляхи μ_{sj} найдовші шляхи рангу $r=2$ у графі $D\Delta$. Отже, припущення про існування шляху μ_{sj} неправильне, а значить, твердження 6.1 справедливе й для множин $m_{sj}^{r=3}$.

Припустимо, що воно виконується й для $m_{sj}^{r=k+1}$, і доведемо, що воно правильне й для множин $m_{sj}^{r=k}$. Нехай на базі множин шляхів $m_{sj}^{r=k}$ процедура A_0 , що використовує стратегію L_1 , побудувала множини шляхів $m_{sj}^{r=k+1}$. Виділимо в множинах $m_{sj}^{r=k+1}$ шляхи максимальної довжини μ_{sj} й припустимо, що в графі $D\Delta$ існує шлях μ_{sj} , що задовольняє властивість ν , але довший μ_{sj} . Останнє можливо, якщо існує шлях μ_{sj} , довший, ніж μ_{sj} , але який задовольняє властивість ν . Але це суперечить первісному припущенню про те, що шляхи μ_{sj} – найдовші шляхи рангу $r = k$ у графі $D\Delta$, і, отже, припущення про існування шляху μ_{sj} в $D\Delta$ $r = k + 1$ неправильне, тобто твердження 6.1 справедливе й для множин $m_{sj}^{r=k+1}$.

Отже, ми показали, що твердження 6.1 правильне для множин $m_{sj}^{r=1}$, $m_{sj}^{r=2}$, і довели, що воно виконується для $m_{sj}^{r=3}$. Далі було припущено, що воно поширюється й на множини $m_{sj}^{r=k}$, й доведено, що у випадку правильності припущення, твердження 6.1 справедливе й для множин рангу $r=k+1$. Отже, на основі принципу повної математичної індукції твердження 6.1 справедливе й для довільних множин $m_{sj}^{r<q}$, тому що при $r > q$ множини шляхів $m_{sj}^{r<q}$ порожні й, отже, теорема 6.1 правильна.

З доведеної теореми 6.1 впливає стратегія відсікань L_2 , що дозволяє значення функціонала шляху $\mu_{sj}^{** r=q}$ використовувати як верхню оцінку для відсікання шляхів у множинах $m_{sj}^{r<q}$ графа $D\Delta$. Стратегія L_2 використовується при побудові багато етапних алгоритмів, тому приклад, що пояснює роботу стратегії L_2 , буде розглянутий пізніше.

Не порушуючи стратегій L_1 і L_2 , можна ввести додаткове відсікання в множинах шляхів m_{sj}^r , що являє собою стратегію L_3 , засновану на властивість графа $D\Delta$ і міститься ось у чому. Зі структури графа $D\Delta$ видно, що для кожної його вершини j , що відповідає області Ω_j , вага γ_j дорівнює сумі коефіцієнтів c_j у функціоналі (6.10), обумовлений за правилом

$$\gamma_j = c_{j+1} + c_{j+2} + \dots + c_n, \gamma_n = 0; j = (\overline{1, n-1}), \quad (6.18)$$

являє собою верхню оцінку приросту величини значення локального екстремуму в області Ω_j на всіх наступних рангах. Скорочення числа формованих шляхів μ_{sj}^r у множинах m_{sj}^r може бути досягнуто, якщо виконуються умови, обумовлені досить очевидним твердженням 6.2.

Твердження 6.2. Якщо сума довжини шляху $d_c(\mu_{sj}^r)$ з підмножини $r m_{sj}^r$ у вершину w і ваги γ_w цієї вершини менше вже отриманої максимальної довжини $d_c(\mu_{sj=p}^r)$, то шляхи наступного

рангу, побудовані на основі μ_{sj}^r , не можуть визначати оптимальне розв'язання задачі.

Справедливість твердження 6.2 випливає з того, що процедура A_0 до стратегії L_1 вибирає глобальний екстремум з усіх локальних, а найбільша довжина за рахунок продовження шляху μ_{sj}^r вже менше одного з локальних екстремумів. Таким чином, перевірка умови

$$d_c(\mu_{sp}^r) + \gamma_p < \max_{\{c_j\}} \left\{ d_c \left(\begin{matrix} * \\ \mu_{sp}^r \end{matrix} \right) \right\}, \quad (6.19)$$

де $d_c(\mu_{sp}^r)$ – довжина шляху μ_{sp}^r до вершини p рангу r за вагами c_j ; дозволяє виключити цей шлях з подальшого аналізу як безперспективний, якщо умова виконана. На цій перевірці заснована стратегія L_3 .

Приклад 6.2. Потрібно максимізувати функціонал

$$f(\bar{x}) = 20x_1 + 18x_2 + 18x_3 + 15x_4 + 10x_5 + 4x_6 + 1x_7$$

при обмеженні

$$7x_1 + 2x_2 + 5x_3 + 2x_4 + 3x_5 + 1x_6 + 4x_7 \leq 12,$$

у відповідності до стратегій L_1 і L_3 .

Розв'язання. Ілюстрація розв'язання наводиться на рис. 6.7. Тут і далі підкреслення шляху означає то, що тільки на основі цього шляху відбувається побудова множин наступного рангу, інакше шлях виключається з подальшого аналізу. Ліворуч у вигляді стовпця проставлене значення γ_j , $j = (\overline{1,7})$ (6.18), обумовлене для кожної вершини графа $D\Delta$. Так, наприклад, $\gamma_3 = c_4 + c_5 + c_6 + c_7 = 15 + 10 + 4 + 1 = 30$.

Відмінною рисою від прикладу 6.2 є наявність порожніх множин m_{sj}^r . Це обумовлено застосуванням відсікань за (6.19). Наприклад, побудуємо шлях $\mu_{s15}^{r=2}$. До цього часу $\max \left\{ d_c \left(\mu_{s12}^{* r=2} \right) \right\} = 38$, а $d_c \left(\mu_{s12}^{r=2} \right) = 30$. Тоді $d_c \left(\mu_{s15}^{r=2} \right) + \gamma_5 = 35$. Отже, на основі шляху $\mu_{s15}^{r=2}$, згідно з твердженням 6.2, неможливо побудувати шлях, що перевищує за вагами функціонала значення $\mu_{s12}^{* r=2}$ вже існуючого локального екстремуму (шлях μ_{s12}).

В іншому розв'язання не змінилося в порівнянні з прикладом 6.1.

γ_j								
66	<u>s1(20,7)</u> 1							
48	<u>s2(18,2)*</u> 2	<u>s12(38,9)</u> 2						
30	<u>s3(18,5)</u>	<u>s13(38,12)</u>						
	3	<u>s23(36,7)*</u> 3		3				
15	<u>s4(15,2)</u>	<u>s14(35,9)</u>	<u>s124(53,11)</u>					
	4	<u>s24(33,4)</u>	<u>s234(51,9)*</u>					
		<u>s34(33,7)</u> 4		4				
5		5	<u>s125(48,12)</u> 5	<u>s2345(61,12)*</u> 5	5	5		
1		6	6	6	6	6	6	
0		7	7	7	7	7	7	7
	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	

$* r=4$
 $\mu_{s2345} \rightarrow \bar{x}^{no} = (0111100)$
 $f(\bar{x}^{no}) = 61$

Рис.6.7.Ілюстрація роботи стратегії L_1 і L_3

Як видно з прикладу 6.2, спільне застосування стратегій L_1 і L_3 дозволяє в 3 рази зменшити число векторів, що будуються в множинах m_{sj}^r , у порівнянні з застосуванням однієї стратегії L_1 .

У результаті роботи процедури A_0 зі стратегіями L_1 , L_3 залишаються множини m_{sj}^r , для яких $r > q$ (у прикладах 6.1 і 6.2 вони є порожніми), а для ефективної фільтрації безперспективних шляхів необхідно мати оцінки значення функціонала в областях

$\Omega_j, j = (\overline{1, n^2/2})$ при $r > q$. Для цього введемо стратегію вибору шляхів з множин L_4 , що ґрунтується на процедурі, яка дозволяє одержати шляхи максимально можливого рангу r у графі $D\Delta$. Її можна реалізувати, якщо в $D\Delta$ визначити найкоротші шляхи за вагами обмежень на основі процедури A_0 між вершиною s і всіма іншими вершинами графа $D\Delta$, при цьому рекурентне співвідношення (6.16) для одновимірної задачі (6.10), (6.11) набуде такого виду:

$$\mu_{sp}^{r=r+1} = \min_{a_{ij}} \{ \mu_{sj}^r \cup (j, p) \}; \quad p = (\overline{r+1, n}); \quad j = (\overline{r, n}), \quad i = 1. \quad (6.20)$$

Неважко бачити, що при цій стратегії формування шляхів у множинах m_{sj}^r залишається справедливим твердження 6.2 і для відсівання шляхів у множинах m_{sj}^r можна використовувати умову (6.19).

Приклад 6.3. Потрібно максимізувати функціонал

$$f(\bar{x}) = 47x_1 + 45x_2 + 38x_3 + 32x_4 + 22x_5 + 22x_6 + 20x_7 + 18x_8$$

при обмеженні

$$10x_1 + 4x_2 + 10x_3 + 8x_4 + 9x_5 + 5x_6 + 1x_7 + 1x_8 \leq 30$$

у відповідності до стратегій L_3 і L_4 (рис. 6.8).

Розв'язання. З кожної множини m_{sj}^r для подальшого аналізу вибираються шляхи $\mu_{sj}^r \in m_{sj}^r$, що мають мінімальне значення за вагою $d_a(\mu_{sj}^r)$. Такі шляхи на рис. 6.9 підкреслені. Інші – виключаються з подальшого аналізу за (6.20). На ранзі $r = 4$ у множині $m_{s6}^{r=4}$ шлях $\mu_{s1246}^{r=4}$, що лежить в основі оптимального, втрачається через то, що існує шлях $\mu_{s2456}^{r=4}$, менший за вагами функціонала й менший за вагами обмеження. Слід зазначити

також, що для даного прикладу застосування стратегій L_3 і L_4 не дає оптимального розв'язання задачі. Так, отриманий шлях $\mu_{s245678}$

не є оптимальним, тому що останній був відкинтий на ранзі $r = 4$ у множині $m_{s6}^{r=4}$ з зазначеної вище причини.

η_j							
197	$s1(47,10)^*$						
152	$s2(45,4)$	2	$s12(92,14)$	2			
114	$s3(38,10)$		$s13(85,20)$		$s123(130,24)$		
		3	$s23(83,14)$	3			
82	$s4(32,8)$		$s14(79,18)$		$s124(124,22)^*$		
			$s24(77,12)$		$s234(115,22)$		
		4	$s34(70,18)$	4			
60	$s5(22,9)$		$s15(69,19)$		$s125(114,23)$		
			$s25(67,12)$		$s235(105,23)$		
			$s35(60,19)$		$s245(99,21)$		
		5	$s45(54,17)$	5			
38	$s6(22,5)$		$s16(69,15)$		$s126(114,19)$		
			$s26(67,9)$		$s236(105,19)$		
			$s36(60,15)$		$s246(99,17)$		
		6	$s46(54,13)$	6			
18		7		7	$s127(112,15)$	7	
0		8		8		8	$s245678(141,27)$
							8
							$s245678(159,28)$
							8
							8

$$\mu_{s124678}^{r=6} \rightarrow \bar{x}^{r=6} = (11010111)$$

$$f(\bar{x}^{r=6}) = 184, k = 3.$$

Рис.6.8.Ілюстрація роботи стратегії L_3 і L_4

Реалізація процедури A_0 з урахуванням стратегій L_3, L_4 дозволяє сформулювати додаткову стратегію L_5 відсікання безперспективних шляхів у множинах m_{sj}^r , обумовлену теоремою 6.2.

Теорема 6.2. Якщо з множини $m_{sj}^r, j = (\overline{r_{max}, n})$ максимального рангу r_{max} , побудованого процедурою A_0 зі стратегіями L_3, L_4 , вибрати шлях μ_{sw}^r з найбільшим значенням за вагами функціонала c_j , то у випадку повторного розв'язання вихідної задачі досить формувати шляхи першого рангу $r=1$ у множини

$$m_{sp}^r, p = (\overline{1, k}), \tag{6.21}$$

де $k = w - r_{max} + 1$,

а w – номер множини m_{sw}^r з шляхом максимальної довжини за вагами функціонала, обумовлений при першому розв'язанні задачі.

Доведення. Припустимо, що існує шлях μ_{sp}^{*r} ($p > k$), який досягає значення максимального рангу r_{max} . Тому що $p > k$, то шлях μ_{sp}^{*r} у графі $D\Delta$ буде проходити по нижніх вершинах. Але ваги c_j , згідно з (6.11), відсортовані в порядку спадання й вершинам, що знаходяться на більш низьких горизонтальних лінійках, в $D\Delta$ відповідають і менші ваги за функціоналом. Тому на кожному ранзі шлях μ_{sp}^{*r} зможе набрати не більш, ніж шлях μ_{sw}^r , а, отже, він і в сумі набере за вагами $\{c_j\}$ довжину не більшу, ніж μ_{sw}^r , що було потрібно довести.

Відповідно до стратегії відсікань L_5 , у випадку повторного розв'язку задачі (приклад 6.3) алгоритмом будувати вектори множин з $m_{s4}^{r=1}$ по $m_{s8}^{r=1}$ не потрібно, тому що кінцевою вершиною шляху максимального рангу $\mu_{s256788}^{r=6}$ є вершина $w = 8$. Ранг цього шляху дорівнює $r_{max} = 6$, по (6.21) $k = 3$, а отже, будувати шляхи першого рангу необхідно тільки в $m_{s1}^{r=1}$, $m_{s2}^{r=1}$, $m_{s3}^{r=1}$.

Найбільш важливою стратегією L_6 є стратегія, заснована на понятті виділення коридора в множині m_{sj}^r , що дозволить будувати точні алгоритми розв'язку задачі (6.10), (6.11). Оскільки нумерація змінних і вершин у графі $D\Delta$ відповідає порядку спадання коефіцієнтів c_j у функціоналі, а процедура A_0 на кожному етапі формує множини m_{sj}^r , починаючи з індексів $j = (\overline{r, n})$, то шляхи $\mu_{sj}^r \in m_{sj}^r$ виявляться завжди відсортованими в порядку спадання довжин за вагами функціонала.

У вигляді теореми 6.3 сформулюємо правило, що дозволяє в множині m_{sj}^r виділити шлях $\bar{\mu}_{sj}^r$ з довжиною $d_c(\bar{\mu}_{sj}^r)$, стосовно якого всі шляхи з меншим значенням довжини за вагами функціонала можуть бути виключені з аналізу як неперспективні.

Теорема 6.3. Якщо відсортувати в множині $m_{sj}^r, j = (\overline{r, n})$ рангу r вектори в порядку спадання за вагами функціонала c_j , то шляхи, у яких довжина $d_c(\mu_{sj}^r)$ менше довжини $d_c\left(\begin{matrix} \min \mu_{sj}^r \\ a_{ij} \end{matrix}\right)$, не можуть визначати оптимальне розв'язання задачі.

Доведення. Припустимо, що на основі шляху μ_{sj}^{*r} з множини m_{sj}^r , у якої довжина $d_c\left(\begin{matrix} *r \\ \mu_{sj} \end{matrix}\right)$ менша, ніж довжина $d_c\left(\begin{matrix} \min \mu_{sj}^r \\ a_{ij} \end{matrix}\right)$, вдалося побудувати оптимальне розв'язання задачі (6.13) - (6.16). Тоді в множини $m_{sp}^{r=r+1}, p = (\overline{r, n})$ наступного рангу повинен потрапити шлях i , мінімальний за вагами обмежень, адже він тим більше буде задовольняти властивість v . Однак довжина за вагами функціонала в такому шляху буде більша, ніж у μ_{sj}^{*r} , і, отже, припущення про існування μ_{sj}^{*r} неправильне, а цей шлях необхідно виключити з подальшого аналізу, що й було потрібно довести.

Приклад 6.4. Потрібно максимізувати функціонал

$$f(\bar{x}) = 47x_1 + 45x_2 + 38x_3 + 32x_4 + 22x_5 + 22x_6 + 20x_7 + 18x_8$$

при обмеженні

$$10x_1 + 4x_2 + 10x_3 + 8x_4 + 9x_5 + 5x_6 + 1x_7 + 1x_8 \leq 30$$

у відповідності до стратегій L_3 і L_6 .

Розв'язання. Побудова шляхів наступного рангу буде здійснюватися на основі шляхів попереднього рангу. Усі шляхи, на основі яких і відбувається така побудова, виділені підкресленням (рис. 6.9). Для подальшого аналізу беруться тільки два шляхи $\mu_{s125}^{r=3}$ і $\mu_{s235}^{r=3}$, які підкреслені. Інші виключаються з множини $m_{s5}^{r=3}$. Аналогічно будуються множини шляхів усіх рангів.

197	s1(47,10)*						
		1					
152	s2(45,4)	s12(92,14)*					
		2	2				
114	s3(38,10)	s13(85,20)	s123(130,24)				
		s23(83,14)					
		3	3	3			
82	s4(32,8)	s14(79,18)	s124(124,22)*				
		s24(77,12)	s134(117,28)				
		s34(70,18)	s234(115,22)				
		4	4	4			
60	s5(22,9)	s15(69,19)	s125(114,23)				
		s25(67,12)	s135(107,29)				
		s35(60,19)	s235(105,23)				
		s45(54,17)	s145(101,27)				
		5	5	5	5		
		s245(99,21)					
38	s6(22,5)	s16(69,15)	s126(114,19)	s1236(152,29)			
		s26(67,9)	s136(107,25)	s1246(146,27)*			
		s36(60,15)	s236(105,19)	s1256(136,28)			
		s46(54,13)	s146(101,23)	s2356(127,28)			
		6	6	6	6	6	
		s246(99,17)	s2456(121,26)				
18			s127(112,15)	s1237(150,25)	s12367(172,30)		
				s1247(144,23)	s12467(166,28)*		
				s1257(134,24)	s12567(156,29)		
				s1267(134,20)			
		7	7	7	7	7	7
0							
	8	8	8	8	8	s124678(184,29)	8
	r=1	r=2	r=3	r=4	r=5	r=6	r=7

* r=6
 $\mu_{s124678} \rightarrow \bar{x}^{i\bar{0}} = (11010111)$

$f(\bar{x}^{i\bar{0}}) = 184.$

Рис. 6.9. Ілюстрація роботи стратегії L_3 і L_6

Так, наприклад, у множині $t_{s5}^{r=2}$ шляхи $\mu_{s35}^{r=2}$, $\mu_{s45}^{r=2}$ виключаються з подальшого розгляду, тому що згідно з теоремою 6.3 існує шлях $\mu_{s25}^{r=2}$ у вершину 5 більший за вагами функціонала й менший за вагами обмежень. На відміну від прикладу 6.3 застосування стратегій L_3 і L_6 дозволяє точно розв'язати дану (і будь-яку іншу) задачу.

Теорема 6.4 визначає поняття коридора.

Визначення. Під *одновимірним коридором* з множини t_{sj}^r в множину $t_{sp}^{r=r+1}$ будемо розуміти сукупність шляхів μ_{sj}^r , що знаходяться між верхньою межею множини t_{sj}^r і його нижньою межею, що задовольняє властивість v у множині $t_{sp}^{r=r+1}$. Верхня

межа визначається шляхом з максимальною $d_c(\mu_{sj}^r)$, а нижня – теоремою 6.3.

Виходячи з поняття коридора, можна запропонувати таку стратегію L_7 , яка являє собою ще одну стратегію вибору шляхів, що полягає в тому, щоб з множини m_{sj}^r рангу r у множини $m_{sp}^{r=r+1}$, $p = (\overline{r, n})$ наступного рангу вибрати шляхи, що задовольняють властивість v_i є максимальними за вагами функціонала c_j і мінімальними за вагами обмежень a_{lj} , що відповідає рекурентним співвідношенням (6.17), (6.20).

Приклад 6.5. Потрібно максимізувати функціонал

$$f(\bar{x}) = 50x_1 + 50x_2 + 45x_3 + 41x_4 + 37x_5 + 36x_6 + 31x_7 + 27x_8$$

при обмеженні

$$9x_1 + 7x_2 + 6x_3 + 9x_4 + 6x_5 + 6x_6 + 5x_7 + 1x_8 \leq 27$$

у відповідності до стратегій L_3 і L_7 .

Розв'язання. Даний приклад (рис. 6.10) розкриває поняття одновимірного коридора, тобто під коридором у множині $m_{s5}^{r=3}$ розуміється підмножина векторів, що складає зі шляхів $\mu_{s125}^{r=3}$, $\mu_{s135}^{r=3}$, $\mu_{s235}^{r=3}$, а у відповідності до стратегії L_7 для подальшого аналізу беруться тільки два шляхи $\mu_{s125}^{r=3}$ і $\mu_{s235}^{r=3}$, які підкреслені. Інші виключаються з множини $m_{s5}^{r=3}$. Аналогічно будуються множини шляхів усіх рангів.

Розглянемо тепер правила фільтрації безперспективних шляхів усередині виділеного коридора, починаючи зі стратегії L_8 , заснованої на твердженні 6.3.

η								
267	$\mu_1(50,9)$							
217	$\mu_2(50,7)^*$	$\mu_{12}(100,16)$						
172	$\mu_3(45,6)$	$\mu_{13}(95,15)$ $\mu_{23}(95,13)^*$	$\mu_{123}(145,22)$					
131	$\mu_4(41,9)$	$\mu_{14}(91,18)$ $\mu_{24}(91,16)$ $\mu_{34}(86,15)$	$\mu_{124}(141,25)$ $\mu_{134}(136,24)$ $\mu_{234}(136,22)$					
94	$\mu_5(37,6)$	$\mu_{15}(87,15)$ $\mu_{25}(87,13)$ $\mu_{35}(82,12)$ $\mu_{45}(78,15)$	$\mu_{125}(137,22)$ $\mu_{135}(132,21)$ $\mu_{235}(132,19)^*$ $\mu_{145}(128,24)$ $\mu_{345}(123,21)$					
58	$\mu_6(36,6)$	$\mu_{16}(86,15)$ $\mu_{26}(86,13)$ $\mu_{36}(81,12)$ $\mu_{46}(77,15)$ $\mu_{56}(73,12)$	$\mu_{126}(136,22)$ $\mu_{136}(131,21)$ $\mu_{236}(131,19)$ $\mu_{146}(127,24)$ $\mu_{156}(123,21)$ $\mu_{346}(122,21)$ $\mu_{356}(118,18)$	$\mu_{2356}(168,25)^*$				
27	$\mu_7(31,5)$	$\mu_{17}(81,14)$ $\mu_{27}(81,12)$ $\mu_{37}(76,11)$	$\mu_{127}(131,21)$ $\mu_{137}(126,20)$ $\mu_{237}(126,18)$ $\mu_{147}(122,23)$ $\mu_{157}(118,20)$	$\mu_{1237}(176,27)$ $\mu_{1257}(168,27)$ $\mu_{2347}(167,27)$ $\mu_{1267}(167,27)$ $\mu_{2357}(163,24)$ $\mu_{3567}(149,23)$				
0					$\mu_{23568}(195,26)^*$			
	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	

$* r=5$
 $\mu_{s_{23568}} \rightarrow \bar{x}^{i\bar{0}} = (01101101)$
 $f(\bar{x}^{i\bar{0}}) = 195.$

Рис.6.10 Ілюстрація роботи стратегії L_3 і L_7

Твердження 6.3. Якщо в коридорі існують два шляхи $\mu_{sj}^r \in t_{sj}^r$ і $\mu_{sj}^{**r} \in t_{sj}^r$, для яких $d_c(\mu_{sj}^{*r}) > d_c(\mu_{sj}^{**r})$; $d_a(\mu_{sj}^{*r})_1 \leq d_a(\mu_{sj}^{**r})_1$,

то вектор \bar{x} , що відповідає шляхи μ_{sj}^{*r} не може належати оптимальному рішенню задачі (6.10) - (6.11).

Покажемо справедливість твердження 6.3. Для цього припустимо, що на основі шляхи μ_{sj}^{**r} можна побудувати оптимальне розв'язання. Останнє означає, що на наступних рангах шлях μ_{sj}^{**r} набуде більшого значення за вагами функціонала, ніж шлях μ_{sj}^{*r} , тобто $d_c(\mu_{sj}^{*r}) < d_c(\mu_{sj}^{**r})$. Але шлях

μ_{sj}^{*r} може бути побудований у тій же множині m_{sj}^r , що й шлях μ_{sj}^{**r} , який задовольняє властивість ν і має довжину $d_a \left(\mu_{sj}^{*r} \right) \leq d_a \left(\mu_{sj}^{**r} \right)$.

При цьому за вагами функціонала до цього моменту шлях μ_{sj}^{*r} має довжину більшу, ніж шлях μ_{sj}^{**r} . Отже, кожний з них набере ту саму величину за вагами c_j , але на початку $d_c \left(\mu_{sj}^{*r} \right) > d_c \left(\mu_{sj}^{**r} \right)$.

Отже, припущення про те, що на основі шляху μ_{sj}^{**r} можна побудувати оптимальне розв'язання, неправильне, а твердження 6.3 справедливе.

Приклад 6.6. Потрібно максимізувати функціонал

$$f(\vec{x}) = 47x_1 + 45x_2 + 38x_3 + 32x_4 + 22x_5 + 22x_6 + 20x_7 + 18x_8$$

при обмеженні

$$10x_1 + 4x_2 + 10x_3 + 8x_4 + 9x_5 + 5x_6 + 1x_7 + 1x_8 \leq 30$$

відповідно до стратегій L_3 і L_8 (рис. 6.11).

Розв'язання. Розглянемо застосування стратегії L_8 до однієї з множин графа $D\Delta$ – множини $m_{s5}^{r=3}$. У результаті в ній залишаються лише два шляхи $\mu_{s125}^{r=3}$, $\mu_{s245}^{r=3}$, що відповідають двом векторам: (11001000) і (01011000), а без цього правила їх було б 5 (див. приклад 6.4). Аналогічний виграш і для інших множин.

Далі розглянемо стратегію L_9 фільтрації рішень у коридорі на основі каліброваних векторів. Для цього введемо поняття каліброваного вектори $\vec{y}_j, j = (1, n)$ для кожної вершини j графа $D\Delta$. Множина $\{\vec{y}_j\}$ утворює калібровану шкалу, що дає верхню оцінку \hat{r}_g для будь-якого вектори \vec{x} , що відповідає шляху μ_{sj}^r , і

що дозволяє визначити, на скільки рангів можна ще продовжити шлях μ_{sj}^r , що задовольняє властивість v . Побудову каліброваних векторів \vec{y}_j почнемо з першої вершини графа $D\Delta$, спочатку відсортувавши коефіцієнти $\{a_{lj}\}$ у порядку зростання. Утвориться вектор \vec{a}'_{lq} , $q = (\overline{1, n})$, у якому $a'_{l1} \leq a'_{l2} \leq \dots \leq a'_{ln}$. У цьому векторі \vec{a}'_{lq} викреслюємо $a'_{lq} = a_{l1}$, після чого формуємо компоненти каліброваного вектори \vec{y}_l за правилом

$$y_{lk} = a_{lk} + y_{l(k-1)}; \quad k = (\overline{1, n-1}); \quad y_{l0} = 0. \quad (6.22)$$

Далі у \vec{a}'_{lq} викреслюємо $a'_{lq} = a_{l2}$, після чого аналогічно правилу (6.22) продовжуємо формувати компоненти вектори \vec{y}_2 , \vec{y}_3 і т.д. На j -му кроці викреслюємо в \vec{a}'_{lq} елемент $a'_{lq} = a_{lj}$, тоді

$$y_{jk} = a_{jk} + y_{j(k-1)}; \quad k = (\overline{1, n-j}); \quad (6.23)$$

$$y_{j0} = 0; \quad y_{jn} = b_j; \quad j = (\overline{1, n-1}).$$

197	$s1(47,10)^*$								
	1								
152	$s2(45,4)$	$s12(92,14)^* 2$							
114	$s3(38,10)$	$s13(85,20)$	$s123(130,24)$						
	3	$s23(83,14)$	3						
82	$s4(32,8)$	$s14(79,18)$	$s124(124,22)^*$						
		$s24(77,12)$	$s134(117,28)$						
		$s34(70,18)$	$s234(115,22)$	4					
60	$s5(22,9)$	$s15(69,19)$	$s125(114,23)$						
		$s25(67,12)$	$s135(107,29)$						
		$s35(60,19)$	$s235(105,23)$						
		$s45(54,17)$	$s145(101,27)$						
		5	$s245(99,21)$	5					
38	$s6(22,5)$	$s16(69,15)$	$s126(114,19)$	$s1236(152,29)$					
		$s26(67,9)$	$s136(107,25)$	$s1246(146,27)^*$					
		$s36(60,15)$	$s236(105,19)$	$s1256(136,28)$					
		$s46(54,13)$	$s146(101,23)$	$s2456(121,26)$					
		6	$s246(99,17)$	6					
18			$s127(112,15)$	$s1237(150,25)$	$s12367(172,30)$				
				$s1247(144,23)$	$s12467(166,28)^*$				
				$s1257(134,24)$					
				$s1267(134,20)$					
0		7	7	7	7	7	7	7	7
		8	8	8	8	8	$s124678(184,29)$	8	8
	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$		

$$* r=6$$

$$\mu_{s124678} \rightarrow \vec{x}^{\#0} = (11010111).$$

$$f(\vec{x}^{\#0}) = 184.$$

Рис. 6.11. Ілюстрація застосування стратегій L_3 і L_8

Приклад 6.7. Потрібно максимізувати функціонал

$$f(\vec{x}) = 20x_1 + 15x_2 + 12x_3 + 8x_4 + 2x_5 \quad (6.24)$$

при обмеженні

$$10x_1 + 3x_2 + 7x_3 + 4x_4 + 2x_5 \leq 10. \quad (6.25)$$

Для задачі (6.24),(6.25) вектор $\vec{a}'_{1q} = \{2,3,4,7,10\}$. З нього видаляємо $a'_{15} = a_{11} = 10$, тоді $\vec{a}'_{1q} = \{2,3,4,7\}$, а $\vec{y}_1 = \{0,2,5,9,16\}$. Далі викреслюємо $a'_{1q} = a_{12} = 3$. Вектор $\vec{a}'_{1q} = \{2,4,7\}$, а калібрований вектор $\vec{y}_2 = \{0,2,6,13\}$. Аналогічно за допомогою (6.23) будуються калібровані вектори для вершин 3, 4, 5: $\vec{y}_3 = \{0,2,6\}$; $\vec{y}_4 = \{0,2\}$; $\vec{y}_5 = \{0\}$. Сформулюємо правила, за якими за допомогою каліброваного вектори \vec{y}_j для вершин $j = (\overline{1, n})$ можна визначити шляхи μ_{sj}^r у графі $D\Delta$ верхню оцінку \hat{r}_6 максимального значення рангу, що продовжує шлях μ_{sj}^r у графі $D\Delta$ (тобто яке число одиниць може додатися у вектор \vec{x}).

Правило K_1 :

- а) обчислюємо для шляху μ_{sj}^r величину $\Delta d = b_1 - d_a(\mu_{sj}^r)$;
- б) у каліброваному векторі \vec{y}_j для вершини j знаходимо номер k , $k = (\overline{0, n-j})$ елемента y_{jk} , починаючи з якого $y_{jk} > \Delta d$ (якщо $\forall y_{jk} < \Delta d$, то $k = n - j + 1$);
- в) визначаємо значення \hat{r}_6 для шляху μ_{sj}^r у графі $D\Delta$: $\hat{r}_6 = k - 1$.

У прикладі (6.24), (6.25) шлях $\mu_{s_2}^{r=1}$ має $d_a(\mu_{s_2}^{r=1})=3$. Шляху $\mu_{s_2}^{r=1}$ відповідає вектор $\vec{x}=(01000)$. Визначимо $\hat{r}_6(\mu_{s_2}^{r=1})$ за правилом K_1 :

- а) $\Delta d = b_1 - d_a(\mu_{s_2}^{r=1}) = 10 - 3 = 7$;
- б) в \vec{y}_2 елемент $y_{23} > \Delta d$ ($13 > 7$), отже, $k=3$;
- в) значення $\hat{r}_6(\mu_{s_2}^{r=1}) = k - 1 = 3 - 1 = 2$.

Отже, шлях у $\mu_{s_2}^{r=1}$ у графі $D\Delta$ можна продовжити, задовольняючи обмеження, не більш ніж на два ранги. Справедливість застосування правила K_1 для визначення \hat{r}_e підтверджується теоремою 6.5.

Теорема 6.5. Якщо шлях μ_{sj}^r у вершину j рангу r у графі $D\Delta$ має довжину за вагами обмежень $d_a(\mu_{sj}^r) = d$, то верхня оцінка \hat{r}_e , побудована за правилом K_1 , визначає максимальне число рангів r , на яке може бути продовжений шлях μ_{sj}^r у графі $D\Delta$.

Доведення. Припустимо, що для шляху μ_{sj}^r з вершини j найшлося таке його продовження $\mu_{sj}^{r'}$, ранг якого $r' > \hat{r}_e + 1$. Але, відповідно до пункту «а» правила K_1 $\Delta d'$ для $\mu_{sj}^{r'}$ буде більше Δd для шляху $\mu_{sj}^{r+\hat{r}_e}$, що неможливо, тому що b_l фіксоване й обидва шляхи є продовженням шляху μ_{sj}^r з довжиною $d_a(\mu_{sj}^r) = d$. Отже, $d' = \Delta d$, припущення про існування $\mu_{sj}^{r'}$ неправильне й теорема доведена.

Дотепер ми не враховували величину коефіцієнтів c_j при функціоналі, а прогноз здійснювали тільки на основі коефіцієнтів a_{ij} . Поставимо у відповідність кожному вектору \vec{y}_j вектор \vec{z}_j , елементи якого є верхніми оцінками \hat{z}_e довжини шляху μ_{sj}^r за вагами функціонала c_j . Компоненти вектори \vec{z}_j для вершини j формуються відповідно до правила K_2 .

Правило K_2

Нехай перша компонента вектори \vec{z}_j вершини j дорівнює $\hat{z}_{j1}^e = c_{j+1}$, друга – $\hat{z}_{j2}^e = c_{j+2} + \hat{z}_{j1}^e$; третя – $\hat{z}_{j3}^e = c_{j+3} + \hat{z}_{j2}^e$; й т.д. k -та компонента –

$$\hat{z}_{jk}^e = c_{j+k} + \hat{z}_{j(k-1)}^e; k = (\overline{1, n-j}); \hat{z}_{j0}^e = 0; \hat{z}_{n0}^e = 0; j = (\overline{1, n-1}). \quad (6.26)$$

Тоді за (6.26) побудуємо \bar{z}_j для задачі (6.24), (6.25). У вершині $j = 1$ елементи вектори \bar{z}_1 відповідно до правила K_2 будуть

$$\hat{z}_{10}^6 = 0; \quad \hat{z}_{11}^6 = c_{j+1} = c_2 = 15; \quad \hat{z}_{12}^6 = c_3 + \hat{z}_{11}^6 = 12 + 15 = 27;$$

$$\hat{z}_{13}^6 = c_4 + \hat{z}_{12}^6 = 8 + 27 = 35; \quad \hat{z}_{14}^6 = c_5 + \hat{z}_{13}^6 = 2 + 35 = 37.$$

Для вершини $j=2$ згідно з (6.26) маємо:

$$\hat{z}_{20}^6 = 0; \quad \hat{z}_{21}^6 = c_3 = 12; \quad \hat{z}_{22}^6 = c_4 + \hat{z}_{21}^6 = 8 + 12 = 20;$$

$$\hat{z}_{23}^6 = c_5 + \hat{z}_{22}^6 = 2 + 20 = 22.$$

Аналогічно будуються вектори для інших вершин графа $D\Delta$. Таким чином, для прикладу (6.19) - (6.20) калібровані оцінні вектори для вершин $j = (\overline{1, n})$ мають вигляд

$$j = 1; \quad \bar{y}_1 = \{0, 2, 5, 9, 16\}; \quad \bar{z}_1 = \{0, 15, 27, 35, 37\}; \quad \gamma_1 = 37;$$

$$j = 2; \quad \bar{y}_2 = \{0, 2, 6, 13\}; \quad \bar{z}_2 = \{0, 12, 20, 22\}; \quad \gamma_2 = 22;$$

$$j = 3; \quad \bar{y}_3 = \{0, 2, 6\}; \quad \bar{z}_3 = \{0, 8, 10\}; \quad \gamma_3 = 10;$$

$$j = 4; \quad \bar{y}_4 = \{0, 2\}; \quad \bar{z}_4 = \{0, 2\}; \quad \gamma_4 = 2;$$

$$j = 5; \quad \bar{y}_5 = \{0\}; \quad \bar{z}_5 = \{0\}; \quad \gamma_5 = 0.$$

Неважко бачити, що у векторів \bar{z}_j останні елементи збігаються з γ_j – вагою (6.18), що може набрати шлях μ_{sj}^r з вершини j на всіх наступних рангах при задоволенні їм властивість v . Як випливає з (6.19), на основі рангового підходу можна більш точно визначати верхню оцінку $\hat{z}_e = f(\hat{r}_e)$ за рахунок того, що шлях μ_{sj}^r не завжди може бути продовжений на всі ранги, що як не задовольняє властивість v . Інакше кажучи, на підставі рівності

$$\gamma_j = \hat{z}_{j\hat{r}_e}^6(\mu_{sj}^r) \quad (6.27)$$

можна збільшити ефективність фільтрації безперспективних шляхів, замінюючи в (6.19) γ_p на (6.27), що дає стратегія L_9 на основі такої нерівності:

$$d_c(\mu_{sp}^r) + \hat{z}_{p\hat{r}_e}^e(\mu_{sp}^r) < \max_{c_j} \{ d_c(\mu_{sp}^{*r}) \}. \quad (6.28)$$

Перевірка нерівності (6.28) дозволить вилучити шлях μ_{sp}^r з подальшого аналізу як неперспективний, тому що ми заздалегідь знаємо, що більш ніж на \hat{r}_e рангів шлях μ_{sp}^r не може бути продовжений.

Приклад 6.8. Потрібно максимізувати функціонал

$$f(\bar{x}) = 47x_1 + 45x_2 + 38x_3 + 32x_4 + 22x_5 + 22x_6 + 20x_7 + 18x_8$$

при обмеженні

$$10x_1 + 4x_2 + 10x_3 + 8x_4 + 9x_5 + 5x_6 + 1x_7 + 1x_8 \leq 30$$

відповідно до стратегій L_3 , L_8 і L_9 .

Розв'язання. На початку кожної вершини j графа ДД (рис. 6.12) зазначені калібровані вектори \bar{y}_j й $\bar{z}_j, j = (\overline{1, n})$. За дужками кожного шляху μ_{sj}^r показана величина $r_{\hat{a}}$. Так, для шляху $\mu_{s14}^{r=2}$ оцінка дорівнює $r_{\hat{a}}=3$. Додаткове використання правила K_2 сприяло ефективному відсіванню безперспективних варіантів розв'язання.

$\bar{y}_1 = \{0,1,2,6,11,19,28\}$ $\bar{z}_1 = \{0,45,83,115,137,159,179\}$	$\S 1(47,10)5^*$						
$\bar{y}_2 = \{0,1,2,7,15,24,34\}$ $\bar{z}_2 = \{0,38,70,92,114,134,152\}$	$\S 2(45,4)5$	$\S 12(92,14)4^*$					
$\bar{y}_3 = \{0,1,2,7,15,24\}$ $\bar{z}_3 = \{0,32,54,76,96,114\}$	$\S 3(38,10)4$	$\S 13(85,20)3$ $\S 23(83,14)4$	$\S 123(130,24)2$				
$\bar{y}_4 = \{0,1,2,7,16\}$ $\bar{z}_4 = \{0,22,44,64,82\}$	$\S 4(32,8)4$	$\S 14(79,18)3$ $\S 24(77,12)4$ $\S 34(70,18)3$	$\S 124(124,22)3^*$ $\S 134(117,28)2$ $\S 234(115,22)3$				
$\bar{y}_5 = \{0,1,2,7\}$ $\bar{z}_5 = \{0,22,42,60\}$	$\S 5(22,9)3$	$\S 15(69,19)3$ $\S 25(67,12)3$ $\S 35(60,19)3$ $\S 45(54,17)3$	$\S 125(114,23)3$ $\S 235(105,23)3$ $\S 145(101,27)2$ $\S 245(99,21)3$				
$\bar{y}_6 = \{0,1,2\}$ $\bar{z}_6 = \{0,20,38\}$	$\S 6(22,5)2$	$\S 16(69,15)2$ $\S 26(67,9)2$ $\S 36(60,15)2$ $\S 46(54,13)2$	$\S 126(114,19)2$ $\S 136(107,25)2$ $\S 236(105,19)2$ $\S 146(101,23)2$ $\S 246(99,17)2$	$\S 1236(152,29)1$ $\S 1246(146,27)2^*$ $\S 1256(136,28)2$ $\S 2456(121,26)2$			
$\bar{y}_7 = \{0,1\}$ $\bar{z}_7 = \{0,18\}$			$\S 127(112,15)1$	$\S 1237(150,25)1$ $\S 1247(144,23)1$	$\S 12367(172,30)0$ $\S 12467(166,28)1^* 7$		
$\bar{y}_8 = \{30\}; \bar{z}_8 = \{0\}$						$\S 124678(184,29)0^* 8$	
	r=1	r=2	r=3	r=4	r=5	r=6	

* r=6
 $\mu_{s124678} \rightarrow \bar{x}^{i\hat{a}} = (11010111)$
 $f(\bar{x}^{i\hat{a}}) = 184.$

Рис. 6.12. Ілюстрація роботи стратегій L_3 , L_8 і L_9

Дотепер розглядалася множина $\{L_w\}$ стратегій відсікань для одновимірної задачі ЦЛП з БЗ. Сформулюємо стратегії відсікань для m -вимірних задач. У випадку розв'язання m -вимірних задач найбільш простою стратегією вибору є стратегія L_{10} , заснована на рекурентному співвідношенні (6.17). Принцип формування шляхів у множинах ідентичний стратегії L_1 , за винятком необхідності виконувати m раз перевірку (6.8).

Приклад 6.9. Потрібно максимізувати функціонал

$$f(\bar{x}) = 16x_1 + 15x_2 + 12x_3 + 10x_4 + 8x_5 + 5x_6 + 2x_7$$

при обмеженні

$$4x_1 + 3x_2 + 1x_3 + 2x_4 + 3x_5 + 2x_6 + 1x_7 \leq 9,$$

$$3x_1 + 1x_2 + 2x_3 + 1x_4 + 1x_5 + 3x_6 + 1x_7 \leq 9,$$

відповідно до стратегій L_3 і L_{10} (рис. 6.13).

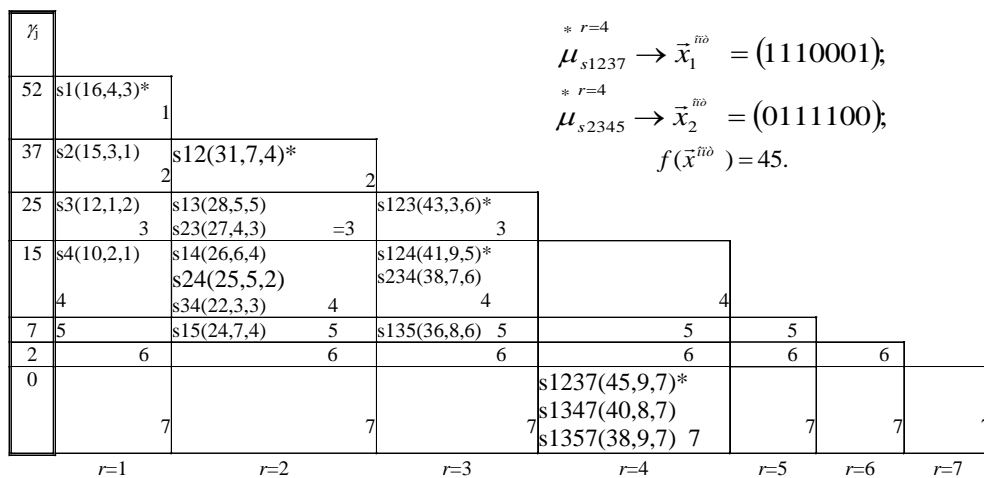


Рис.6.13. Ілюстрація роботи стратегій L_3 і L_{10}

Розв'язання. Побудова множин m_{sj}^r здійснюється так само, як і в одновимірній задачі. У цьому прикладі два оптимальних розв'язання: шляхи $\mu_{s_{1237}}^{r=4}$ і $\mu_{s_{2345}}^{r=4}$. Однак другий втрачається через наявність у множині $m_{s_3}^{r=2}$ шляху $\mu_{s_{13}}^{r=2}$, що домінує над $\mu_{s_{23}}^{r=2}$. Сполучення стратегії L_7 зі стратегією L_3 для випадку

m -вимірної задачі утворює правило вибору L_{11} , що полягає в тому, щоб з множини m_{sj}^r в множину $m_{sp}^{r=r+1}$, $p = (\overline{r, n})$, що впливає з рангу, вибирати шляхи, що задовольняють властивість ν і є максимальними за вагами функціонала c_j і мінімальними за кожною i -ю вагою обмежень $a_{ij}, i = (\overline{1, m})$. Перше відповідає рекурентному співвідношенню (6.12), а друге описується таким співвідношенням, заснованим на (6.20).

$$\mu_{sp}^{r=r+1} = \min \{ \mu_{sj}^r \cup (j, p) \}; \quad j = (\overline{r, n}); \quad p = (\overline{j+1, n}); \quad i = (\overline{1, m}). \quad (6.29)$$

$$a_{ij}$$

Приклад 6.10. Потрібно максимізувати функціонал

$$f(\vec{x}) = 29x_1 + 28x_2 + 26x_3 + 25x_4 + 20x_5 + 17x_6 + 15x_7 + 10x_8$$

при обмеженні

$$1x_1 + 4x_2 + 2x_3 + 3x_4 + 4x_5 + 1x_6 + 2x_7 + 2x_8 \leq 9,$$

$$2x_1 + 2x_2 + 1x_3 + 2x_4 + 4x_5 + 2x_6 + 1x_7 + 4x_8 \leq 9$$

відповідно до стратегій L_3 і L_{11} (рис. 6.14).

η_j								
137	s1(29,1,2)*							
	1							
113	s2(28,4,2)	s12(57,5,4)						
	2	2						
87	s3(26,2,1)	s13(55,3,3)*	s123(83,7,5)					
		s23(54,6,3)						
	3	3	3					
62	s4(25,3,2)	s14(54,4,4)	s124(82,8,6)					
		s24(53,7,4)	s134(80,6,5)*					
	4	s34(51,5,3)	4	4				
42	s5(20,4,4)	s15(49,5,6)	s125(77,9,8)					
		s25(48,8,6)	s135(75,8,8)					
	5	s35(46,6,5)	s145(74,8,8)					
		s45(45,7,6)	s345(71,9,7)	5				
25	s6(17,1,2)	s16(46,2,4)	s126(74,6,6)	s1236(100,8,7)*				
		s26(45,5,4)	s136(72,4,5)	s1246(99,9,8)				
		s36(43,3,3)	s146(71,5,6)	s1346(98,7,7)				
		s46(42,4,4)	s346(68,6,5)					
		s56(37,5,6)	s156(66,6,8)					
	6	6	s356(63,7,7)	6	6	6	6	6
10				s1237(98,9,6)	s13467(112,9,8)*			
				s1347(95,8,6)	7	7	7	7
0						8	8	8
	r=1	r=2	r=3	r=4	r=5	r=6	r=7	

$$* r=5$$

$$\mu_{s13467} \rightarrow \vec{x}^{no} = (10110110);$$

$$f(\vec{x}^{no}) = 112.$$

Рис. 6.14. Ілюстрація роботи стратегій L_3 і L_{11}

Розв'язання. Побудова множин шляхів здійснюється відповідно до принципу оптимізації за напрямком (6.16). У множинах t_{sj}^r виділені підкресленням шляхи μ_{sj}^r , що є мінімальними за кожним з обмежень (при рівності i -го обмеження вибирається шлях з найбільшою вагою за функціоналом) і максимальним за вагою функціонала. Збільшення числа обмежень приводить до невизначеності в оцінці верхньої межі за вагою функціонала $\hat{z}_{j\hat{r}_e}^e$, яку може набрати шлях μ_{sj}^r , і верхній оцінці кількості рангів \hat{r}_e , на яке може бути продовжений цей шлях. Щоб уникнути цієї невизначеності, для кожної вершини $j = (\overline{1, n})$ будемо будувати не один, а m каліброваних векторів \vec{y}_j^i , $i = (\overline{1, m})$ за правилом

$$\begin{aligned} y_{jk}^i &= a_{jk}^i + y_{j(k-1)}^i; \quad k = (\overline{1, n-j}); \quad y_{j0}^i = 0; \\ y_{n0}^i &= b_i; \quad j = (\overline{1, n-1}). \end{aligned} \quad (6.30)$$

У зв'язку з цим правило K_1 визначення \hat{r}_e для шляху μ_{sj}^r в багатовимірній задачі набуває нижчеподаного виду.

Правило K_3

- а) нехай $i = 1$;
- б) при поточному значенні i обчислюємо для шляху μ_{sj}^r величину $\Delta d = b_i - d_a(\mu_{sj}^r)_i$;
- в) у каліброваному векторі \vec{y}_j^i вершини j знаходимо номер k -го $k = (\overline{0, n-j})$ елемента y_{jk}^i вектора \vec{y}_j^i , починаючи з якого $y_{jk}^i > \Delta d$ (якщо $\forall y_{jk}^i < \Delta d$, то $k = n - j + 1$);
- г) для шляху μ_{sj}^r за i -м обмеженням у графі $D\Delta$: $\hat{r}_e^i = k - 1$;
- д) збільшуємо значення i на 1 і переходимо до п. «б» цього правила у випадку $i \leq m$ або до наступного пункту у випадку $i > m$;
- е) серед усіх \hat{r}_e^i , $j = (\overline{1, m})$ вибираємо найменший.

Таким чином, відповідно до правила K_3 верхня оцінка \hat{r}_e для шляху μ_{sj}^r в j -й вершині графа $D\Delta$ при m обмеженнях буде

$$\hat{r}_e = \min_i \{ \hat{r}_e^i \}, \quad i = (\overline{1, m}). \quad (6.31)$$

Неважко показати, що при реалізації правила K_3 справедливість теореми 6.4 не порушується. Так, якщо по одному з обмежень ми зможемо пройти на r_1 рангів, а по іншому на r_2 ранги й при цьому $r_1 > r_2$, то на ранг, більший ніж r_2 , неможливо побудувати шлях, що задовольняє властивість v , оскільки r_2 – верхня оцінка по даному обмеженню.

Правило K_2 для визначення верхньої оцінки за вагами функціонала в m -вимірних задачах не змінюється.

Уведемо за аналогією з поняттям одновимірного коридора поняття m -вимірного коридора в множині m_{sj}^r .

Позначимо підмножину векторів, що повинна залишитися в множині m_{sp}^r після фільтрації через m_{sp}^{rk} . Нагадаємо, що формування множин шляхів у m_{sp}^r здійснюється завжди таким чином, що довжина за вагами функціонала першого з них більше довжини за вагами функціонала другого й т. д., тобто $d_c(\mu_k) \geq d_c(\mu_k) \geq \dots \geq d_c(\mu_k)$.

Визначення. Під виділенням m -вимірного коридора будемо розуміти процес відсікання k -го шляху μ_k у множині m_{sp}^r , якому відповідає m -вимірний вектор $(d_a(\mu_k)_1, d_a(\mu_k)_2, \dots, d_a(\mu_k)_m)$, за правилом K_4 .

Правило K_4 .

Якщо $(\forall i \in (\overline{1, m})) d_a(\mu_v)_i \leq d_a(\mu_k)_i$, при цьому $v = (\overline{1, k-1})$, то вектор $(d_a(\mu_k)_1, d_a(\mu_k)_2, \dots, d_a(\mu_k)_m)$ і відповідний йому шлях μ_k можна видалити з подальшого аналізу, інакше $\mu_k \in m_{sp}^{r=r+1}$.

Справедливість відсікання за правилом K_4 впливає з такого досить очевидного твердження 6.4.

Твердження 6.4. Шляхи $\mu_{sp}^r \in m_{sp}^r$, що не потрапили в m_{sp}^{rk} , не можуть визначати оптимальний розв'язок задачі (6.1)-(6.3).

Доведення. Нехай деякий шлях $\mu_{sp}^r \notin m_{sp}^r$ є оптимальним розв'язанням задачі. Тоді відповідно до правила K_4 в m_{sp}^{rk} існує шлях μ_{sp}^{**r} , більший за вагами функціонала й менший за всіма вагами обмежень. Отже, шлях μ_{sp}^{**r} набуде на наступних рангах i більшого значення за вагами функціонала, при задоволенні властивість ν , тобто $d_c(\mu_k^{**r}) > d_c(\mu_k^{*r})$. Ми прийшли до протиріччя, а отже, наше припущення неправильне і твердження 6.4 справедливе.

Позначимо потужність множини $|m_{sj}^r|$ через ε . До відсікання $|m_{sj}^{rk}| = \emptyset$. Тоді процедуру формування m -вимірного коридора можна подати у вигляді послідовної перевірки векторів $(1, \varepsilon)$ з множини m_{sj}^r на можливість відсікання за правилом K_4 . Вектори, що залишилися, утворять підмножину m_{sj}^{rk} . Цю процедуру будемо називати стратегією L_{12} , що полягає у виділенні m -вимірного коридора усередині кожної множини m_{sj}^r рангу r , коли в ньому побудовані всі шляхи з вищих вершин графа $D\Delta$.

Приклад 6.11. Потрібно максимізувати функціонал

$$f(\vec{x}) = 17x_1 + 15x_2 + 13x_3 + 8x_4 + 7x_5 + 4x_6 + 4x_7$$

при обмеженні

$$\begin{aligned} 2x_1 + 4x_2 + 6x_3 + 9x_4 + 9x_5 + 2x_6 + 1x_7 &\leq 23; \\ 4x_1 + 6x_2 + 9x_3 + 1x_4 + 8x_5 + 10x_6 + 6x_7 &\leq 33 \end{aligned}$$

відповідно до стратегій L_9 і L_{12} (рис. 6.15).

3. У каліброваному векторі \bar{y}_1^1 вершини 1 знаходимо номер k -го елемента y_{1k}^1 вектори \bar{y}_j^i , починаючи з якого $y_{1k}^1 > \Delta d \rightarrow k = 5$.

4. Для шляху $\mu_{s1}^{r=1}$ по 1-му обмеженню: $\hat{r}_e^i = k - 1 = 5 - 1 = 4$.

5. Збільшуємо i на 1, тобто $i = 2$.

6. При поточному значенні i обчислюємо для шляху $\mu_{s1}^{r=1}$ величину

$$\Delta d = b_i - d_a(\mu_{s1}^{r=1})_2 = 33 - 4 = 29.$$

7. У каліброваному векторі \bar{y}_1^2 вершини 1 знаходимо номер k -го елемента y_{1k}^2 вектори \bar{y}_j^i , починаючи з якого $y_{1k}^2 > \Delta d \rightarrow k = 5$.

8. Для шляху $\mu_{s1}^{r=1}$ по 2-му обмеженню: $\hat{r}_e^i = k - 1 = 5 - 1 = 4$.

9. Серед усіх $\hat{r}_e^i, j = (\overline{1,2})$ вибираємо найменший, тобто $\hat{r}_e = 4$.

Маючи значення \hat{r}_e за співвідношенням (6.28) перевіряємо на можливість відсікання за стратегією L_9 .

Після формування множин усього рангу здійснюється виділення m -вимірного коридора. Наприклад, на ранзі $r = 2$ у множині $m_{s3}^{r=2}$ шлях $\mu_{s23}^{r=2}$ не потрапляє в коридор, тому що шлях $\mu_{s13}^{r=2} \in m_{s3}^{r=2}$ має більшу довжину за вагами функціонала й меншу довжину за вагами всіх обмежень. У вигляді підкреслення виділені шляхи, що потрапили в коридор.

Однак більш ефективною є стратегія відсікання L_{13} , при якій виділення m -вимірного коридора здійснюється на всьому ярусі r , після того, як на ньому сформується всі шляхи, що задовольняють у кожній вершині властивість v . Тоді здійснюється перевірка: чи є для довільного шляху $\mu_{sp}^r \in m_{sp}^r$ рангу r шлях μ_{sj}^r , що належить одній з вищих множин графа $D\Delta$ над m_{sp}^r , що має більшу довжину за вагами функціонала й не меншу довжину за вагами всіх m обмежень. Якщо такий шлях μ_{sj}^r

є, то відповідно до твердження 6.4, шлях μ_{sp}^r можна видалити з розгляду як безперспективний.

Приклад 6.12. Потрібно максимізувати функціонал

$$f(\bar{x}) = 17x_1 + 15x_2 + 13x_3 + 8x_4 + 7x_5 + 4x_6 + 4x_7$$

при обмеженні

$$\begin{aligned} 2x_1 + 4x_2 + 6x_3 + 9x_4 + 9x_5 + 2x_6 + 1x_7 &\leq 23; \\ 4x_1 + 6x_2 + 9x_3 + 1x_4 + 8x_5 + 10x_6 + 6x_7 &\leq 33 \end{aligned}$$

відповідно до стратегій L_9 і L_{13} (рис. 6.17).

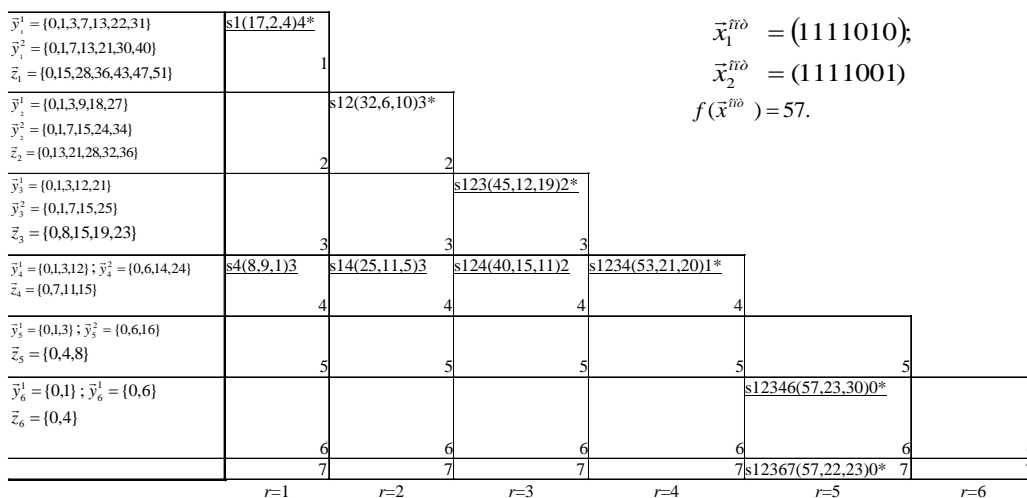


Рис. 6.17. Ілюстрація роботи стратегій L_9 і L_{13}

Розв'язування. Процес розв'язання ідентичний описаному процесу в прикладі 2.12, однак, виділення коридора відбувається на всьому ярусі. Таким чином, коли множини одного рангу сформовані, виконується виділення коридора як у множинах, так і між множинами одного рангу. Неважко помітити, що кількість оброблюваних векторів значно скоротилася (див. рис. 6.17).

Таким чином, процедуру A_0 зі стратегіями $\{L_w\}$ і правилами K_w можна покласти в основу побудови наближених і точних алгоритмів розв'язку задачі про рюкзак, тобто одержати різні модифікації процедури A_0 залежно від комбінацій

використовуваних правил відсікання $\{L_w\}$ безперспективних шляхів у множинах μ_{sj}^r на основі застосування принципу оптимізації за напрямком.

6.4. Алгоритми розв'язання задачі (0, 1)-рюкзак на основі рангового підходу

Алгоритми методу відсікань безперспективних варіантів

На основі висунутих у підрозд. 6.3 стратегій $\{L_w\}$ і правил $\{K_w\}$ необхідно розробити алгоритми розв'язання задачі цілочисельного лінійного програмування (0,1)-рюкзак. У результаті отримані одно етапні й багато етапні алгоритми, класифікація яких подана на рис.6.18.

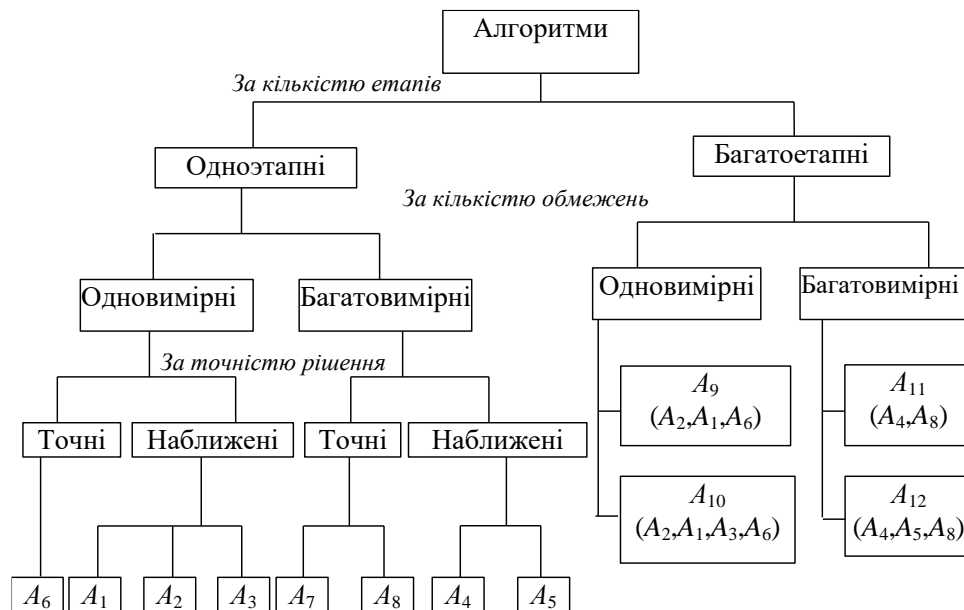


Рис 6.18. Класифікація алгоритмів - модифікацій узагальненої процедури A_0

Основними кваліфікаційними ознаками є кількість етапів і точність етапів розв'язання задачі. За першою ознакою алгоритми діляться на одно етапні й багато етапні. Даний поділ означає, що вихідна задача може вирішуватися один (одно етапні) або кілька разів (багато етапні) різними алгоритмами, перелік яких поданий

у дужках. Друга ознака визначає похибку одержуваного припустимого розв'язку задачі (6.7) - (6.9) за вагами функціонала від оптимального розв'язання. Якщо алгоритм принципово не може дати точного розв'язання за вагами функціонала, то такий алгоритм будемо називати наближеним.

6.4.1. Наближені алгоритми розв'язання задачі ЦЛП з БЗ

АЛГОРИТМ A_1

Розгляд наближених алгоритмів почнемо з одновимірної задачі (6.10), (6.11). Найбільш простим є алгоритм A_1 , побудований на основі процедури A_0 , що реалізує стратегії відсікань L_1 і L_3 . Його покроковий опис подано нижче.

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах $m_{sj}^{r=1}$ шляху максимальної довжини $\left\{ \begin{matrix} * & r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j . Для кожної вершини j визначається вага γ_j за правилом (6.18).

КРОК 2. Виключаються шляхи $\{\mu_{sp}^r\}$, $p = (\overline{r, n})$ в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність (6.19).

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі рекурентного співвідношення (6.17). В утворених множинах $m_{sp}^{r=r+1}$ виділяємо найдовші шляхи $\left\{ \begin{matrix} * & r=r+1 \\ \mu_{sp} \end{matrix} \right\}$. Якщо виявиться кілька шляхів максимальної довжини, то серед них вибирається шлях з меншим значенням довжини за вагами обмежень a_{1j} .

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5,

якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \begin{matrix} * r=1 & * r=2 & & * r=n \\ \mu_{sj} & , \mu_{sp} & , \dots & , \mu_{sw} \end{matrix} \right\}$ шлях

максимальної довжини й алгоритм A_1 закінчує роботу.

Одержання наближеного розв'язання обумовлено застосуванням стратегії вибору L_1 . Приклад 6.2 ілюструє роботу алгоритму A_1 .

АЛГОРИТМ A_2

Іншою стратегією вибору шляхів є стратегія L_4 , що за допомогою алгоритму A_2 дозволяє одержати шляхи максимально можливого рангу r у графі $D\Delta$. Її можна реалізувати, якщо в $D\Delta$ визначати найкоротші шляхи за вагами обмежень на основі процедури A_0 між вершиною s і всіма іншими вершинами графа $D\Delta$, використовуючи при цьому рекурентне співвідношення (6.20), і, природно, відсікання за правилом L_3 .

Покроковий опис алгоритму A_2 наведений далі:

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах $m_{sj}^{r=1}$ шляхи максимальної довжини

$\left\{ \begin{matrix} * r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j . Для кожної вершини j

визначається вага γ_j за правилом (6.18).

КРОК 2. Виключаються шляхи $\{\mu_{sp}^r\}$, $p = (\overline{r, n})$ в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність (6.19).

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу, які мають мінімальне значення в m_{sj}^r довжини за вагами обмежень, на основі

рекурентного співвідношення (6.20). В утворених множинах $m_{sp}^{r=r+1}$ виділяємо найдовші шляхи $\left\{ \begin{matrix} * r=r+1 \\ \mu_{sp} \end{matrix} \right\}$.

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \begin{matrix} * r=1 & * r=2 & * r=n \\ \mu_{sj}, \mu_{sp}, \dots, \mu_{sw} \end{matrix} \right\}$ шлях

максимальної довжини й номер останньої вершини w , що належить цьому шляху, після чого алгоритм A_2 закінчує роботу.

Приклад 6.3 відбиває хід розв'язання за алгоритмом A_2 .

АЛГОРИТМ A_3

Виходячи з поняття коридора й у відповідності до стратегії L_7 пропонується наступний наближений алгоритм A_3 розв'язку задачі (6.10), (6.11), який полягає в тому, що з множини m_{sj}^r рангу r у множини $m_{sj}^{r=r+1}$, $j = \overline{(r+1, n)}$ наступного рангу будуть вибиратися не більше двох шляхів, що задовольняють властивість ν .

Першим є шлях з максимальною вагою за функціоналом c_j (за рекурентним співвідношенням (6.17)), а другим – шлях з найменшою вагою за обмеженням $\{a_{lj}\}$ (за рекурентним співвідношенням (6.20)), що відповідним чином змінить крок 3 алгоритму A_1 . В іншому покроковий опис алгоритму A_3 повністю збігається з описом алгоритму A_1 . Втрата оптимального розв'язання обумовлена його влученням усередину коридора. Тому, як видно з прикладу 6.14, даний алгоритм A_3 є наближеним.

Приклад 6.14. Потрібно максимізувати функціонал

$$f(\bar{x}) = 20x_1 + 19x_2 + 17x_3 + 13x_4 + 12x_5 + 10x_6 + 7x_7 + 5x_8 + 2x_9$$

при обмеженні

$$9x_1 + 8x_2 + 8x_3 + 4x_4 + 3x_5 + 2x_6 + 1x_7 + 5x_8 + 1x_9 \leq 30,$$

алгоритмом A_3 (рис. 6.19).

\bar{y}							
85	s1(20,9)*						
66	s2(19,8)	s12(39,17)					
49	s3(17,8)	s13(37,17)*	s23(36,16)				
36	s4(13,4)	s14(33,13)	s124(52,21)				
		s24(32,12)	s134(50,21)				
		s34(30,12)	s234(49,20)				
24	s5(12,3)	s15(32,12)	s125(51,20)				
		s25(31,11)	s135(49,20)				
		s35(29,11)	s235(48,19)				
		s45(25,7)	s145(45,16)*				
			s245(44,15)				
14	s6(10,6)	s16(30,11)	s126(49,19)	s2456(54,17)			
		s26(29,10)	s136(47,19)				
		s36(27,10)	s236(46,18)				
			s146(43,15)				
			s246(42,14)				
			s156(42,14)				
			s456(35,9)				
7			s127(46,18)	s2347(56,21)	s24567(61,18)		
				s1257(58,21)			
				s2457(51,16)			
				s1267(56,20)			
2	8	8	8	8	8	8	8
0	9	9	9	9	9	s245679(63,19)	9
	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$

$$* r=6 \quad \mu_{s145679} \rightarrow \bar{x}^{opt} = (100111101);$$

$$f(\bar{x}^{opt}) = 64.$$

Рис.6.19. Ілюстрація роботи алгоритму A_3

Розв'язання. При побудові шляхів з кожної множини m_{sj}^r вибирається тільки два шляхи: максимальний за вагами функціонала й мінімальний за вагами обмеження (ці шляхи підкреслені), а інші – ігноруються. З цієї причини в множині $m_{s5}^{r=5}$ було виключено з подальшого розгляду шлях $\mu_{s145}^{r=3}$, що належить оптимальному, тому що домінували за алгоритмом A_3 шляхи

$\mu_{s125}^{r=3}$ і $\mu_{s245}^{r=3}$. Отже, результатом розв'язання вийшов шлях $\mu_{s245679}^{r=6}$, який відповідає вектору $\bar{x} = \{010111101\}$, що й дало відносну похибку $\Delta f = (64-63)/64 = 1,5 \%$.

АЛГОРИТМ A_4

Для багатовимірної задачі (6.7) - (6.9) за аналогією також можна побудувати ефективні наближені алгоритми. Таким є алгоритм A_4 , заснований на стратегії L_{10} , що з множини поточного рангу будує в множині наступного рангу $r=r+1$ тільки шлях з максимальною довжиною за вагами функціонала $\{c_j\}$. При цьому тимчасова складність алгоритму зростає в $2m$ раз, тому що необхідно виконати для кожного вектори m операцій додавання й m операцій порівняння, щоб переконатися: задовольняє вектор, що відповідає цьому шляху, обмеженню (6.8) чи ні. Отже, покроковий опис для алгоритму A_4 збігається з алгоритмом A_1 . Приклад 6.9 пояснює розв'язання задачі алгоритмом A_4 , використовуючи стратегії L_3 і L_{10} .

АЛГОРИТМ A_5

Застосування стратегії L_{11} дозволило побудувати алгоритм A_5 , суть якого полягає у виборі з кожної множини m_{sj}^r поточного рангу r одного шляху, максимального за вагами функціонала й, у найгіршому разі, m шляхів μ_{sj}^r , мінімальних за вагами обмежень, що задовольняють властивість v . Тоді, на відміну від алгоритму A_1 , покроковий опис алгоритму A_5 зміниться на наведений нижче.

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = (\overline{1, n})$ першого рангу r , що задовольняють властивість v , і визначаються в множинах $m_{sj}^{r=1}$ шляхи максимальної довжини

$\left\{ \begin{matrix} * r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j . Для кожної вершини j

визначається вага γ_j за правилом (6.18).

КРОК 2. Виключаються шляхи $\{\mu_{sp}^r\}$, $p = (\overline{r, n})$ в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність (6.19).

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі рекурентних співвідношень (6.17), (6.29). В утворених множинах $m_{sp}^{r=r+1}$ виділяємо найдовші шляхи $\left\{ \begin{matrix} * r=r+1 \\ \mu_{sp} \end{matrix} \right\}$. Якщо виявиться кілька шляхів максимальної довжини, то серед них вибирається шлях з меншим значенням довжини за вагами обмеження $\{a_{ij}\}$. Якщо виявиться кілька шляхів з мінімальним значенням за i -м обмеженням, то вибирається той, у якого більша довжина за вагами функціонала.

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \begin{matrix} * r=1 & * r=2 & * r=n \\ \mu_{sj} & , \mu_{sp} & , \dots, \mu_{sw} \end{matrix} \right\}$ шлях

максимальної довжини й алгоритм A_5 закінчує роботу.

Цей алгоритм має одну невизначеність у тому випадку, коли в множині виявиться кілька векторів з максимальною довжиною за вагами функціонала й з різними значеннями за вагами обмежень, тобто, наприклад, за вагами першого обмеження більше другий вектор, за вагами другого – перший, за вагами третього – знову другої й т. д. Тому для запобігання цієї невизначеності при рівності довжин за вагами функціонала вибирається шлях з меншим значенням за вагами всіх обмежень. Якщо таке неможливо, то вибирається кожний. Може виявитися, що відкинутий шлях буде належати оптимальному шляху. Ця обставина і є однією з причин одержання за алгоритмом A_5 наближеного розв'язання.

Приклад 6.10 ілюструє розв'язання задачі алгоритмом A_5 .

6.4.2. Точні алгоритми розв'язання задачі ЦЛП з БП

Побудова точних алгоритмів ґрунтується на понятті коридора й використанні стратегій $L_6, L_8, L_9, L_{12}, L_{13}$ відсікання. Для точних алгоритмів аналітична оцінка тимчасової складності й величини апаратурних витрат являє труднощі через NP-повноту [11] розв'язуваної задачі. Основною причиною неможливості оцінки є випадкове число векторів усередині виділюваного коридора. Природа випадковості кількості векторів усередині коридора визначається співвідношеннями між коефіцієнтами функціонала c_j і коефіцієнтами обмежень a_{ij} , а також приростами цих співвідношень між j -ми компонентами двійкового вектори $\vec{x} = x_1, \dots, x_n$ при заданому законі розподілу коефіцієнтів у функціоналі й обмеженнях.

АЛГОРИТМ A_6

На основі стратегій L_6, L_8, L_9 побудуємо точний алгоритм A_6 розв'язання одновимірної задачі (6.10), (6.11), покроковий опис якого поданий нижче.

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r-1}, j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах m_{sj}^{r-1} шляхи максимальної довжини $\left\{ \begin{matrix} * & r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j . Для кожної вершини j формуються калібровані вектори \vec{y}_j (6.23) і \vec{z}_j (6.26).

КРОК 2. Виключаються шляхи в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність (6.19).

КРОК 3. Для кожного шляху $\{\mu_{sj}^r\}, j = (\overline{r, n})$ поточного рангу r визначається за правилом K_1 значення \hat{r}_e . Виключаються шляхи $\mu_{sj}^r \in m_{sj}^r$, довжини яких задовольняють нерівність (6.28).

КРОК 4. Формуються множини шляхів $m_{sp}^{r=r+1}, p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі принципу

оптимізації за напрямком (6.16) з виділенням коридора за стратегією L_6 і виключенням векторів усередині коридора у відповідності до стратегії L_8 . Шлях у множині $m_{sp}^{r=r+1}$ може бути сформований, якщо він задовольняє властивість. Якщо властивість ν не виконується, то шлях виключається з подальшого аналізу. В утворених множинах $m_{sp}^{r=r+1}$ виділяємо

найдовші шляхи $\left\{ \mu_{sp}^{* r=r+1} \right\}$.

КРОК 5. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 6, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 6, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 6. Виділяємо в множині $\left\{ \mu_{sj}^{* r=1}, \mu_{sp}^{* r=2}, \dots, \mu_{sw}^{* r=n} \right\}$ шлях

максимальної довжини й алгоритм A_6 закінчує роботу.

Приклад 6.8 пояснює розв'язання задачі (6.10), (6.11) за алгоритмом A_6 . Як видно з цього прикладу, в кожній множині m_{sj}^r залишаються тільки шляхи, відсортовані й у порядку спадання за вагами функціонала й у порядку спадання за вагами обмеження (вони підкреслені). Інші виключаються з подальшого аналізу.

АЛГОРИТМ A_7

Для розв'язання багатовимірної задачі ЦЛП з БЗ (6.7) - (6.9) розглянемо два точних алгоритми A_7 і A_8 , заснованих на понятті m -вимірному коридора й стратегій вибору L_{12} , L_{13} відповідно. Принципове розходження алгоритмів A_7 і A_8 полягає в тому, що перший виділяє m -вимірний коридор усередині множини поточного рангу r графа $D\Delta$, а другий виділяє m -вимірний коридор безпосередньо на всьому ярусі графа $D\Delta$. Відсікання, засновані на стратегії L_9 і правилах K_2 , K_3 , K_4 , залишаються однаковими для обох алгоритмів. Покроковий опис алгоритму A_7 наведений далі

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах $m_{sj}^{r=1}$ шляхи максимальної довжини $\left\{ \begin{matrix} * & r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j . Для кожної вершини j формуються калібровані вектори \bar{y}_j^i (6.30) і \bar{z}_j (6.26).

КРОК 2. Для кожного шляху $\left\{ \mu_{sj}^r \right\}$, $j = (\overline{r, n})$ поточного рангу r визначається за правилом K_1 значення \hat{r}_g . Виключаються шляхи $\mu_{sj}^r \in m_{sj}^r$, довжини яких задовольняють нерівність (6.28).

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі принципу оптимізації за напрямком (6.16) з виділенням m -вимірного коридора у вигляді підмножини $m_{sp}^{rk} \in m_{sp}^{r=r+1}$ у відповідності до стратегії L_{12} і відсіканням безперспективних векторів за правилом K_4 . Шлях у множині $m_{sp}^{r=r+1}$ може бути сформований, якщо він задовольняє властивість. Якщо властивість ν не виконується, то шлях виключається з подальшого аналізу. В утворених множинах m_{sp}^{rk} виділяємо найдовші шляхи $\left\{ \begin{matrix} * & r=r+1 \\ \mu_{sp} \end{matrix} \right\}$.

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \begin{matrix} * & r=1 & * & r=2 & * & r=n \\ \mu_{sj}, \mu_{sp}, \dots, \mu_{sw} \end{matrix} \right\}$ шлях максимальної довжини й алгоритм A_7 закінчує роботу.

Приклад 6.11 відбиває розв'язання задачі за алгоритмом A_7 . Головною перевагою цього алгоритму є одержання точного розв'язку. Однак це досягається за рахунок обробки великої кількості векторів, що робить його малопридатним для практичного використання.

АЛГОРИТМ A_8

З метою усунення цього недоліку пропонується алгоритм A_8 , що здійснює виділення m -вимірного коридора на всьому ярусі. Суть цього алгоритму полягає в тому, що якщо на ярусі в множинах, що лежать у графі $D\Delta$ вище деякого шляху μ_{sp}^{*r} рангу r , існує шлях μ_{sp}^{**r} , що має не менше значення за вагами функціонала й менше значення за вагами всіх обмежень, то такий шлях $m_{sp}^{r=r+1}$ не потрапляє в m -вимірний коридор на ярусі. У цьому випадку покроковий опис A_8 набуде нижченаведеного виду.

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=l}$, $j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах $m_{sj}^{r=l}$ шляхи максимальної довжини $\left\{ \mu_{sj}^{*r} \right\}$ за вагами функціонала c_j . Для кожної вершини j формуються калібровані вектори \vec{y}_j (6.30) і \vec{z}_j (6.26).

КРОК 2. Виключаються шляхи в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність (6.19).

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі принципу оптимізації за напрямком (6.16). Шлях у множині $m_{sp}^{r=r+1}$ може бути сформований, якщо він задовольняє властивість ν . Якщо властивість ν не виконується, то шлях виключається з подальшого аналізу. В утворених множинах $m_{sp}^{r=r+1}$ виділяємо локальні екстремуми $\left\{ \mu_{sp}^{*r=r+1} \right\}$.

КРОК 4. Виконуємо процедуру виділення m -вимірного коридора на всьому сформованому ранзі $r=r+1$. Для цього послідовно перевіряємо всі множини $m_{sp}^{r=r+1}$, $p = (\overline{r+1, n})$ у

відповідності до стратегії L_{13} і здійснюємо відсікання безперспективних векторів за правилом K_4 .

КРОК 5. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 6, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності також переходимо до кроку 6, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 6. Виділяємо в множині $\left\{ \begin{matrix} * r=1 & * r=2 & & * r=n \\ \mu_{sj} & , \mu_{sp} & , \dots & , \mu_{sw} \end{matrix} \right\}$ шлях

максимальної довжини й алгоритм A_8 закінчує роботу.

У прикладах 6.11 і 6.12 та сама задача вирішувалася алгоритмами A_7 і A_8 відповідно. Однак число векторів, оброблених алгоритмом A_8 , у два рази менше, ніж алгоритмом A_7 .

Таким чином, на основі принципу оптимізації за напрямком у дискретному просторі станів, обумовленому графом D (рис. 6.4), удалося побудувати точні алгоритми для розв'язання як одновимірної, так і багатовимірної задачі ЦЛП з БЗ. Покажемо, що для зниження загальної тимчасової складності, за рахунок звуження ширини коридора за вагами функціонала, доцільно використовувати багато етапну фільтрацію.

6.4.3. Багатоетапні алгоритми розв'язку задачі ЦЛП з БЗ

Багатоетапні алгоритми призначені для зниження загальної тимчасової складності за рахунок звуження ширини коридора за вагами функціонала. Основою побудови багатоетапних алгоритмів служить багатоетапна фільтрація, у якій можна виділити два основних етапи. Перший етап – це визначення припустимого наближеного розв'язку f^* , що використовується на другому етапі для пошуку оптимального розв'язку точним алгоритмом, що приведе до підвищення ефективності фільтрації безперспективних шляхів у коридорі.

АЛГОРИТМ A_9

Спочатку побудуємо багато етапні алгоритми для одновимірної задачі (6.10), (6.11). На рис.6.20 показаний

трьохетапний алгоритм A_9 , у якому на першому етапі задача розв'язується алгоритмом A_2 з використанням додаткової стратегії L_3 . Застосування цього алгоритму дозволить швидко знайти припустимий розв'язок \vec{x}_1 і запам'ятати $f_{\max}^1 = f(\vec{x}_1)$, а також визначити у відповідності до стратегії L_5 величину k (6.21), що дозволить на наступних етапах не будувати шляхи першого рангу в множинах $m_{sj}^{r=1}, j = (\overline{k+1, n})$. На другому етапі задача розв'язується алгоритмом A_1 з урахуванням величини функціонала отриманого припустимого розв'язку f_{\max} і значення k . Це забезпечить, як видно з прикладу 6.1, більш ефективне відсікання за стратегією L_3 , і відітне на першому ранзі безперспективні множини за стратегією L_5 . Необхідність застосування зазначених стратегій відсікання приведе до зміни 1, 2 і 5-го кроків алгоритму A_1 , у результаті чого утвориться алгоритм A_1^m , покроковий опис якого набуває виду, що подано нижче.

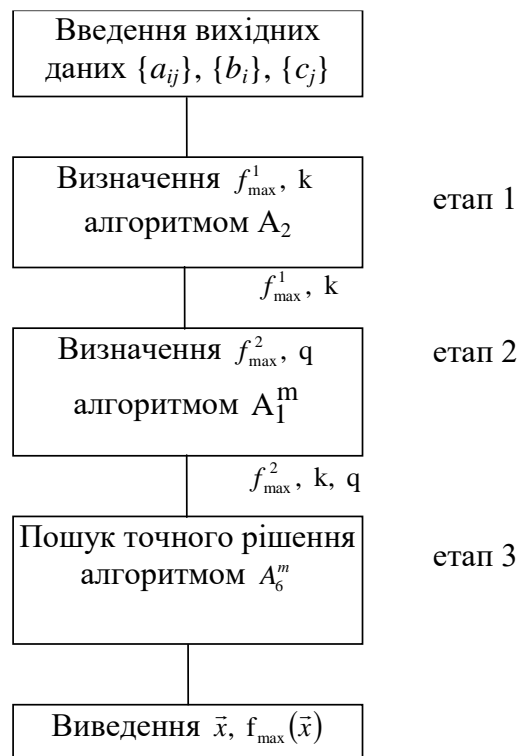


Рис. 6.20. Структура алгоритму A_9

Алгоритм A_I^m

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}, j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах $m_{sj}^{r=1}$ локальні екстремуми $\left\{ \mu_{sj}^{*r} \right\}$.

КРОК 2. Виключаються шляхи $\left\{ \mu_{sp}^r \right\}, p = (\overline{r, n})$ в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність

$$d_c(\mu_{sp}^r) + \gamma_p < \max_{\{c_j\}} \left\{ d_c \left(\mu_{sp}^{*r} \right), f_{max}^1 \right\}. \quad (6.32)$$

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}, p = (\overline{1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі рекурентного співвідношення (6.17). В утворених множинах $m_{sp}^{r=r+1}$ виділяємо найдовші шляхи $\left\{ \mu_{sp}^{*r=r+1} \right\}$. Якщо виявиться кілька шляхів максимальної довжини, то серед них вибирається шлях з меншим значенням довжини за вагами обмежень a_{lj} .

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r = (n - 1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \mu_{sj}^{*r=1}, \mu_{sp}^{*r=2}, \dots, \mu_{sw}^{*r=n}, f(\bar{x}_1) \right\}$

шлях максимальної довжини й запам'ятовуємо ранг цього шляху q . Після цього алгоритм A_I^m закінчує роботу.

Ефективність використання даного етапу залежить від точності припустимого розв'язку \bar{x}_2 . Він також дозволяє визначити величину q , що служить для відсікання безперспективних розв'язків на третьому етапі у відповідності до стратегії L_2 .

На третьому етапі задача розв'язується точним алгоритмом A_6 з урахуванням стратегій відсікань L_2 і L_5 . Застосування цього етапу відповідно до твердження 6.3 гарантує одержання точного розв'язку. Покроковий опис алгоритму A_6 зміниться, що приведе до одержання його модифікації – алгоритму A_6^m , що має нижчеподаний вигляд.

Алгоритм A_6^m

КРОК 1. З вершини s будуються множини шляхів m_{sj}^{r-1} , $j = (\overline{1, k})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах m_{sj}^{r-1} шляхи максимальної довжини $\left\{ \begin{matrix} * \\ \mu_{sj}^r \end{matrix} \right\}$ за вагами функціонала c_j .

КРОК 2. Для кожного шляху $\{\mu_{sj}^r\}$, $j = (\overline{r, n})$ поточного рангу r визначається за правилом K_1 значення \hat{r}_6 . Виключаються шляхи в множинах m_{sj}^r поточного рангу, довжини яких задовольняють нерівність

$$d_c(\mu_{sp}^r) + \gamma_p < \max_{\{c_j\}} \left\{ d_c \left(\begin{matrix} * \\ \mu_{sp}^r \end{matrix} \right), f_{max}^1, f_{max}^2 \right\}. \quad (6.33)$$

КРОК 3. У відповідності до стратегії L_2 , для кожного шляху μ_{sj}^r поточного рангу r здійснюється перевірка виконання умови

$$r + \hat{r}_6 < q. \quad (6.34)$$

Якщо умова виконується, то шлях $\mu_{sj}^r \in m_{sj}^r$ виключається з подальшого аналізу.

КРОК 4. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{r+1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі принципу оптимізації за напрямком (6.11) з виділенням коридора за стратегією L_6 і виключенням векторів усередині коридора у відповідності до стратегії L_8 . Шлях у множині $m_{sp}^{r=r+1}$ може бути сформований, якщо він задовольняє властивість. Якщо властивість ν не виконується, то шлях виключається з подальшого аналізу. В утворених множинах $m_{sp}^{r=r+1}$ виділяємо

найдовші шляхи $\left\{ \mu_{sp}^{* r=r+1} \right\}$.

КРОК 5. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 6, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 6, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 6. Виділяємо в множині $\left\{ \mu_{sj}^{* r=1}, \mu_{sp}^{* r=2}, \dots, \mu_{sw}^{* r=n}, f(\vec{x}_1), f(\vec{x}_2) \right\}$ шлях максимальної довжини й алгоритм A_6^m закінчує роботу.

АЛГОРИТМ A_{10}

Неважно бачити, що чим більш точний розв'язок одержуємо при розв'язанні задачі ЦЛП з БЗ наближеними алгоритмами, тим менше припустимих рішень доводиться аналізувати на останньому етапі й тим швидше буде знайдено точний розв'язок і доведена його оптимальність. Найбільш точним з наближених алгоритмів є алгоритм A_3 .



Рис.6.21. Структура алгоритму A_{10}

Тому пропонується чотирьох етапний алгоритм A_{10} , структурна схема якого зображена на рис. 6.21. Перші два етапи ідентичні першим двом етапам алгоритму A_9 . На третьому етапі використовується алгоритм A_3 , що реалізує стратегію L_7 . Для нього також справедливі стратегії відсікань L_2, L_5 , що приведе до створення алгоритму A_3^m . Для нього перші три кроки збігаються з першими трьома m кроками алгоритму A_6^m , а другі три відповідають крокам 3, 4 і 5 алгоритму A_3 з еквівалентною заміною адрес переходів у ньому.

Результатом розв'язання задачі алгоритмом A_3^m є вектор \bar{x}_3 і відповідне йому значення функціонала (6.10), рівне $f_{max} = f(\bar{x}_3)$. На заключному, четвертому етапі, задача розв'язується алгоритмом A_6^m . Однак на кроці 6 при виборі оптимального розв'язку буде враховуватися вже $n+3$ припустимих розв'язків.

Приклад чотирьохетапного алгоритму показаний на прикладі 6.2.

АЛГОРИТМ A_{11}

Розглянемо можливість побудови багато етапних алгоритмів для багатовимірної задачі ЦЛП з БЗ (6.7) – (6.9). У роботі пропонується два таких алгоритми – A_{11} і A_{12} . Перший алгоритм A_{11} , як показано на рис. 6.22, є двоетапним.



Рис. 6.22. Структура алгоритму A_{11}

На першому етапі швидкого пошуку припустимого розв'язку використовується алгоритм A_4 , заснований на стратегіях L_1, L_9, L_{10} . Покроковий опис алгоритму A_4^m має нижчеподаний вигляд.

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}, j = (\overline{1, n})$ першого рангу r , що задовольняють властивість v , і визначаються в множинах $m_{sj}^{r=1}$ шляхи максимальної довжини

$\left\{ \begin{matrix} * r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j . Для кожної вершини j формуються калібровані вектори \bar{y}_j^i (6.30) і \bar{z}_j (6.26).

КРОК 2. Для кожного шляху $\{\mu_{sj}^r\}, j = (\overline{r, n})$ поточного рангу r визначається за правилом K_3 і (6.31) значення \hat{r}_e . Виключаються шляхи в множинах m_{sj}^r поточного рангу r , довжини яких $d_c(\mu_{sp}^r)$ задовольняють нерівність (6.28).

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{r+1, n})$ наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі рекурентного співвідношення (6.17). В утворених множинах $m_{sp}^{r=r+1}$ виділяємо найдовші шляхи $\left\{ \begin{matrix} * & r=r+1 \\ \mu_{sp} \end{matrix} \right\}$. Якщо виявиться кілька шляхів максимальної довжини, то серед них вибирається шлях з меншим значенням довжини за вагами обмежень a_{lj} .

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \begin{matrix} * & r=1 & * & r=2 & & * & r=n \\ \mu_{sj}, \mu_{sp}, \dots, \mu_{sw} \end{matrix} \right\}$ шлях максимальної довжини й алгоритм A_4^m закінчує роботу.

Алгоритмом A_4^m буде знайдено припустимий розв'язок задачі \bar{x}_l й значення функціонала (6.7), рівне $f_{max}^l = f(\bar{x}_l)$, котре як параметр передається на другий етап алгоритму A_{11} .

Слід зазначити, що калібровані вектори, побудовані алгоритмом A_4^m , не змінюються і використовуються на другому етапі алгоритму A_8 . Його покроковий опис зміниться через необхідність урахувати f_{max}^l , що приведе до модифікації алгоритму A_8^m .

АЛГОРИТМ A_8^m

КРОК 1. З вершини s будуються множини шляхів $m_{sj}^{r=1}$, $j = (\overline{1, n})$ першого рангу r , що задовольняють властивість ν , і визначаються в множинах $m_{sj}^{r=1}$ шляхи максимальної довжини

$\left\{ \begin{matrix} * & r \\ \mu_{sj} \end{matrix} \right\}$ за вагами функціонала c_j .

КРОК 2. Для кожного шляху $\{\mu_{sj}^r\}$, $j = (\overline{r, n})$ поточного рангу r визначається за правилом K_3 і (6.31) значення \hat{r}_e . Виключаються шляхи $\mu_{sj}^r \in m_{sj}^r$, довжини яких задовольняють нерівність

$$d_c(\mu_{sp}^r) + \hat{z}_{p\hat{r}_e}(\mu_{sp}^r) < \max_{c_j} \left\{ d_c \left(\mu_{sp}^r \right), f_{max}^1 \right\}. \quad (6.35)$$

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (\overline{r+1, n})$, наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі принципу оптимізації за напрямком (6.16) з виділенням m -вимірного коридора у вигляді підмножини $m_{sp}^{rk} \in m_{sp}^{r=r+1}$ у відповідності до стратегії L_{12} і відсіканням безперспективних векторів за правилом K_4 . Шлях у множині $m_{sp}^{r=r+1}$ може бути сформований, якщо він задовольняє властивість. Якщо властивість ν не виконується, то шлях виключається з подальшого аналізу. В утворених множинах

m_{sp}^{rk} виділяємо найдовші шляхи $\left\{ \mu_{sp}^{r=r+1} \right\}$.

КРОК 4. Перевіряємо, чи всі множини шляхів наступного $(r+1)$ -го рангу порожні. Якщо це так, то переходимо до кроку 5, якщо ні, то перевіряємо $r = (n-1)$. У випадку виконання рівності переходимо до кроку 5, інакше збільшуємо r на 1 і виконуємо крок 2.

КРОК 5. Виділяємо в множині $\left\{ \mu_{sj}^{*r=1}, \mu_{sp}^{*r=2}, \dots, \mu_{sw}^{*r=n} \right\}$ шлях

максимальної довжини й алгоритм A_8^m закінчує роботу.

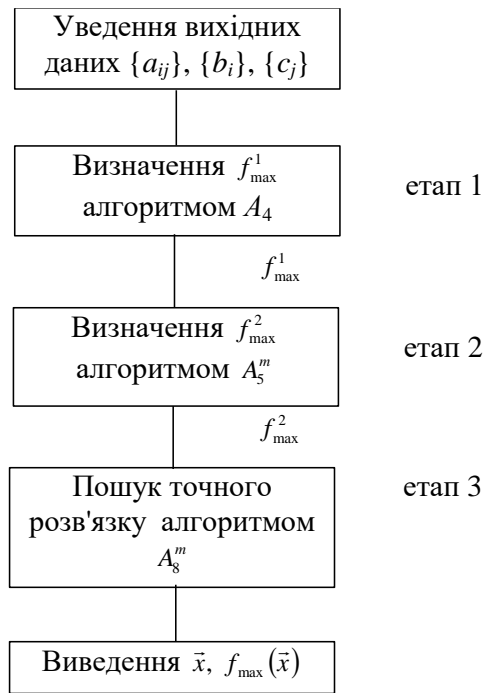


Рис. 6.23. Структура алгоритму A12

АЛГОРИТМ A12

Щоб зменшити кількість оброблюваних векторів усередині m -вимірного коридора, пропонується трьох етапний алгоритм A_{12} (рис. 6.23).

Перший етап відповідає виконанню алгоритму A_4^m , у результаті чого обчислюється припустимий розв'язок \bar{x}_1 і $f_{max}^1 = f(\bar{x}_1)$. На другому етапі використовується алгоритм A_5^m , заснований на стратегіях L_9, L_{11} . Покроковий опис алгоритму A_5^m збігається з описом алгоритму A_8^m , за винятком кроку 3, що має нижчеподаний вигляд.

КРОК 3. Формуються множини шляхів $m_{sp}^{r=r+1}$, $p = (r+1, n)$, наступного рангу, що задовольняють властивість ν , на базі множин шляхів m_{sj}^r попереднього рангу на основі принципу оптимізації за напрямком (6.16) і (6.20). В

утворених множинах m_{sp}^{rk} виділяємо найдовші шляхи $\left. \begin{matrix} * r=r+1 \\ \mu_{sp} \end{matrix} \right\}$.

Якщо виявиться кілька шляхів максимальної довжини, то серед них вибирається шлях з найменшим значенням довжини за вагами обмежень (6.8). Якщо виявиться кілька шляхів з мінімальним значенням за i -м обмеженням, то вибирається шлях більшої довжини за вагами функціонала.

На третьому етапі, як наведено на рис. 6.23, виконується точний алгоритм A_8^m , у якому перевірка нерівності (6.35) на кроці 2 замінена перевіркою такої нерівності:

$$d_c(\mu_{sp}^r) + \hat{z}_{p_{i_6}}^g(\mu_{sp}^r) < \max_{c_j} \{ d_c(\mu_{sp}^{*r}), f_{max}^1, f_{max}^2 \}. \quad (6.36)$$

Приклади 6.3 і 6.4 демонструють розв'язання алгоритмами A_{11} і A_{12} відповідно. Таким чином, багатетапне розв'язання задачі забезпечує одержання швидкого припустимого розв'язання й гарантує точний результат за рахунок використання принципу оптимізації за напрямком і виділення m -вимірного коридора.

6.4.4. Особливості розв'язання ЦЛП з БЗ у випадку рівності коефіцієнтів у функціоналі

Розглянемо випадок задачі (6.7 – 6.9) коли $C_j = \text{const}$ для всіх $j = (\overline{1, n})$, при цьому задача максимізації функціонала (6.7) при обмеженнях (6.8 – 6.9) природно зводиться до визначення шляху максимально можливого рангу в ациклічному графі D_Δ , оскільки максимальному рангу шляху в цьому випадку відповідає найбільше значення функціонала. Ясно, що для побудови шляху максимально можливого рангу, довжина якого не перевищує деяку величину b_i , необхідно, щоб вона з додаванням кожного певного ребра збільшувалася на мініимально можливу величину. Тоді варто очікувати, що в шлях вдається включити найбільше число ребер. Отже, що цікавить нас, шлях варто шукати серед множини найкоротших шляхів від вершини S до всіх інших вершин графа G з урахуванням обмежень (6.8 – 6.9). Позначимо

через $\{\mu_s^{*r}(i)\}$ множину найкоротших шляхів, при i різних варіантах розподілу ваг у графі D_Δ . Тут $\mu_s^{*r}(i = \nu)$ $i = (\overline{1, m})$ – найкоротший шлях максимального можливого рангу r для ν – варіанта розподілу ваг у графі, причому такої, що довжини цього шляху за всіма іншими варіантами розподілу ваг задовольняють обмеження (6.7), тоді задачу визначення шляху максимального можливого рангу задовольняючим обмеженням (6.8) можна визначити як знаходження $\max_r \{\mu_s^{*r}(i)\}$ $i = (\overline{1, m})$, з останнього випливає, що необхідно m разів шукати найкоротші шляхи, що задовольняють (6.8) між вершиною S та іншими вершинами графа D_Δ і виділяти серед них шлях максимального можливого рангу. Потім в отриманій підмножині знайти шлях з максимальним значенням r . Однак у ряді випадків число побудов таких шляхів може бути різко скорочено, якщо виконуються умови, обумовлені твердженням 6.5.

Твердження 6.5. *Максимальний ранг шляху з вершини s дорівнює рангу найкоротшого шляху по кожному з i -варіантів розподілу ваг у графі D_Δ $i = (\overline{1, m})$, якщо в i -варіанті його довжина досягла свого граничного значення b_i .*

Доведення. Нехай у графі D_Δ при визначенні найкоротших шляхів між вершиною S і всіма іншими вершинами графа D_Δ для ν -варіанта розподілу ваг виявилось, що шлях $\mu_s^{*r}(\nu)$ максимального можливого рангу r_ν^* і його довжина досягла свого граничного можливого значення b_ν , а довжини цього шляху, що залишилися $(m-1)$ по варіантам розподілу ваг, задовольняють обмеженням (6.8). Припустимо, що при k -варіанті розподілу ваг удалося побудувати найкоротший шлях $\mu_s^{*r}(k)$ рангу r_k^* , що задовольняє за обмеженнями (6.8) всі варіанти розподілів ваг, що залишилися, але тоді повинен існувати шлях $\mu_s^{*r}(\nu)$ рангу $r_\nu^* < r_k^*$, довжина якого $l \leq b_\nu$, що суперечить первісному припущенню про те, що найкоротший шлях $\mu_s^{*r}(\nu)$ рангу r_ν^* досяг свого граничного можливого значення b_ν . Отже, після припущення про можливість існування шляху $\mu_s^{*r}(k)$ рангу $r_k^* > r_\nu^*$ при задоволенні обмежень (6.8) неправильне, що й було потрібно довести. У такий спосіб виконання умов, визначених твердженням, дозволяє відразу відповісти на запитання розв'язуваної задачі.

Алгоритм розв'язання задачі

При визначенні шляху $\mu_s^{*r}(v)$, для v -варіанта розподілу ваг попередньо будемо здійснювати сортування вагових коефіцієнтів в v -обмеженні в порядку їхнього зростання й при цьому ребру, що входить у вершину 1 графа D_Δ буде завжди відповідати найменший ваговий коефіцієнт, у вершину r – шлях більший і т.д. у вершину n – найбільш вагові коефіцієнти. В основу розв'язку задачі буде покладена наступна спеціалізована процедура визначення найкоротших шляхів, складена з урахуванням того, що ваги в графі D_Δ закріплені за ребрами відповідно до зазначеного сортування.

Процедура A' визначення найкоротших шляхів у графі

КРОК 1. На базі шляху S_1 формуємо всі можливі шляхи рангу $r=2$.

КРОК 2. Для сформованих шляхів перевіряємо за всіма варіантами розподілу ваг $i = (\overline{1, m})$ виконання нерівності (6.8), якщо є серед них такі, для яких хоча б одна нерівність (6.8) не виконується, то ці шляхи виключаються з розгляду, а в тих, що залишилися, обчислюємо довжини за всіма варіантами розподілу ваг і уточнюємо серед тих, що залишилися, чи є довжини шляхів гранично можливого значення, що досягли $i = (\overline{1, m}) b_i$. Якщо так, то переходимо до кроку 5, а якщо ні, то виконуємо наступний крок.

КРОК 3. Перевіряємо сформовану підмножину шляхів: порожня чи ні. Якщо підмножина порожня, то переходимо до кроку 6, якщо ні, то у сформованій множині шляхів вибираємо найкоротший шлях за поточним варіантом розподілу ваг і переходимо до виконання наступного кроку.

КРОК 4. На базі отриманого найкоротшого шляху формуємо всі можливі шляхи рангу $r = r+1$ і переходимо до кроку 2.

КРОК 5. Закінчення роботи алгоритму й фіксація шляху, що досяг свого граничного значення b_i .

КРОК 6. Закінчення роботи алгоритму й фіксація найкоротшого шляху, отриманого останній раз на кроці 4.

З урахуванням розробленої процедури A' визначення найкоротших шляхів алгоритм розв'язку задачі (6.7)-(6.8) буде складатися з нижченаведених кроків.

- КРОК 1.* $i=1$, сортуємо ваги i -ї множини в порядку зростання.
- КРОК 2.* Виконуємо процедуру A' для поточного значення ваг.
- КРОК 3.* Перевіряємо процедурою A' отриманий кінцевий результат чи ні, якщо так, то алгоритм закінчує роботу, якщо ні, то виконуємо наступний крок.
- КРОК 4.* Формуємо множину $\{\mu_s^{*r}(i)\}$ і перевіряємо $i = m$, якщо $i \neq m$, то $i = i + 1$ і переходимо до кроку 2, у протилежному випадку виконуємо наступний крок.
- КРОК 5.* У множині $\{\mu_s^{*r}(i)\}$ виділяємо шлях максимально можливого рангу й на цьому алгоритм закінчує роботу.

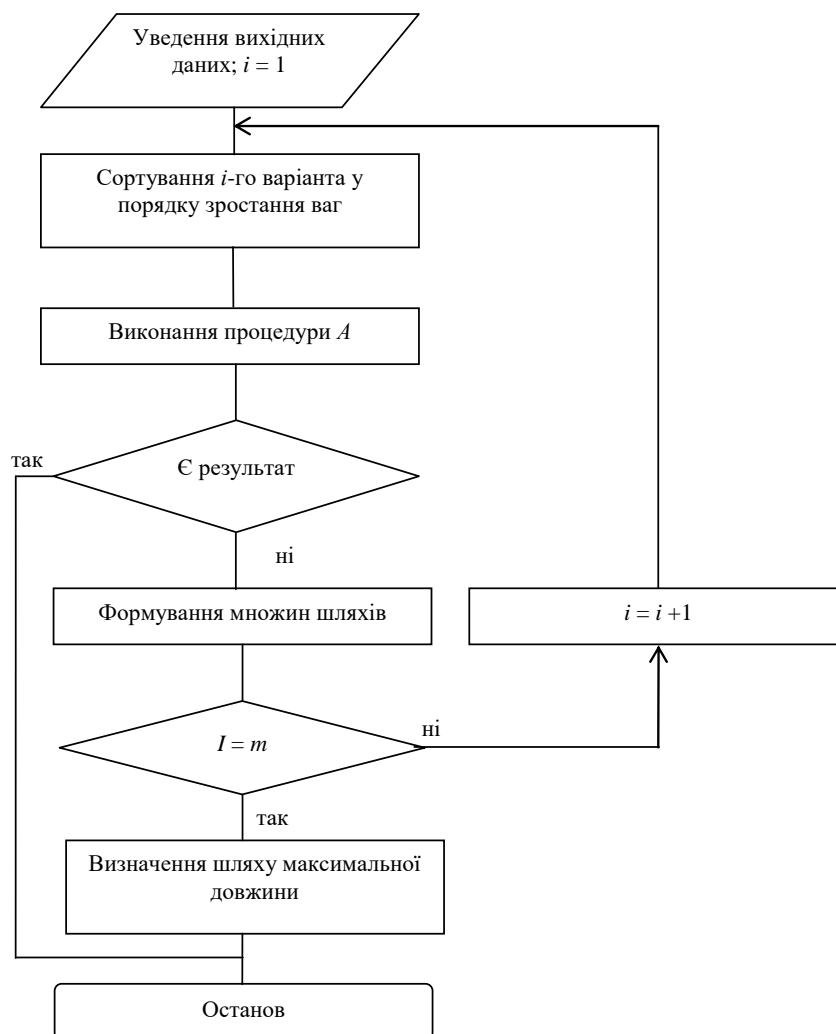


Рис. 6.24. Узагальнена структурна схема алгоритму

Узагальнена структурна схема алгоритму подана на рис. 6.24. Приклади розв'язання задачі за допомогою даного алгоритму й алгоритмів різних стратегій відсікань, що

використовують, розглянуті в [32 - 35]. Аналогічно була розв'язана й задача мінімізації функціонала (6.7) при обмеженнях (6.8). У цьому випадку коефіцієнти функціонала необхідно відсортувати в порядку зростання й у структурній схемі алгоритму всі операції знаходження найдовшого шляху за вагою C_j замінити на операції визначення найкоротшого шляху за ваги C_j .

6.5. Ранговий підхід до розв'язку задачі про найменше покриття й розбиття (ЗНП і ЗНР)

6.5.1. Формальна модель ЗНП

Нехай A^m – транспонована матриця суміжності графа $D\Delta$ з одиничними діагональними елементами. Задача визначення найменшої домінуючої множини графа G еквівалентна задачі знаходження такої найменшої множини стовпців у матриці A^m , що кожний рядок матриці містить одиницю хоча б в одному з обраних стовпців. Ця задача про пошук найменшої множини стовпців, «покриваючих» одиницями всі рядки, одержала назву задачі про найменше покриття (ЗНП). У загальному випадку матриця, що складається з 0 і 1, не обов'язково є квадратною. Крім того, кожному стовпцю j (вершині x_j) у матриці A^m протиставляється деяка вага, і потрібно знайти покриття з найменшою загальною вартістю. У випадку ж рівності коефіцієнтів задача трансформується в задачу мінімізації числа стовпців, що покривають усі рядки в матриці A^m .

Інакше кажучи, необхідно мінімізувати цільову функцію

$$L = \sum_{j=1}^n c_j x_j \rightarrow \min \quad (6.37)$$

при обмеженнях

$$\begin{aligned} \sum_{j=1}^n a_{ij} x_j &\geq 1, \quad i = \overline{1, m}; \\ x_j &\in \{0, 1\}; \quad c_j \geq 0, \end{aligned} \quad (6.38)$$

де

$$a_{ij} = \begin{cases} 1, & \text{якщо } i\text{-та змінна може бути покрита змінною } x_j; \\ 0 & \text{- у протилежному випадку.} \end{cases} \quad (6.39)$$

6.5.2. Ранговий підхід до розв'язання ЗНП

Нехай задана булева матриця $V = \|\beta_{ji}\|$ з n стовпцями й m рядками. Стовпці матриці V будемо задавати m -вимірними векторами $B^{(i)} = (\beta_1^i, \beta_2^i, \dots, \beta_m^i)$, у яких компоненти β_j^i набувають значення, рівне j , якщо j -й рядок покривається однією з одиниць i -го стовпця, й рівне 0 у протилежному випадку. Число компонентів β_j^i у векторі $B^{(i)}$, рівних нулю, назвемо ваговою характеристикою γ_i вектори $B^{(i)}$. Геометрично ЗНП, обумовлена співвідношеннями (6.37) – (6.38), інтерпретується як задача знаходження оптимальної вершини одиничного куба в R^n . Побудуємо на основі матриці V еквівалент n -вимірному одиничного куба R^n у вигляді стягнутого дерева всіх шляхів, заданого графом $G\Delta$ (рис. 6.4), вершини якого утворюють n горизонтальних лінійок. Як показано в [40], множини шляхів від фіктивної вершини s до всіх інших вершин цього графа відповідає 2^{n-1} варіантам можливих рішень. Припустимо, що споконвічно вектори $B^{(i)}$ відсортовані за вагами γ_i у порядку їхнього зростання й пронумеровані так, що заданій послідовності векторів B^1, B^2, \dots, B^n відповідає нерівність $\gamma_1 < \gamma_2 < \dots < \gamma_n$. Поставимо у відповідність кожній вершині графа $D\Delta$ вектор $B^{(i)}$, $i = \overline{1, n}$. Шлях $\mu_{st}^{p, l, \dots, t}$ у графі $D\Delta$, що проходить через k вершин, будемо розглядати як об'єднання k векторів, які відповідають вершинам p, l, \dots, t , що ввійшли в цей шлях, тобто якщо в шлях $\mu_{st}^{p, l, \dots, t}$ увійшли вершини p, l, \dots, t , то шляху $\mu_{st}^{p, l, \dots, t}$ відповідає вектор $B^{(p, l, \dots, t)} = B^{(p)} \cup B^{(l)} \cup \dots \cup B^{(t)}$, при цьому правило об'єднання однойменних компонентів β_j^i , що належать різним векторам $B^{(p)} \cup B^{(l)}$ ($p < l$), таке:

$$\beta_j^p \cup \beta_j^l = \begin{cases} 0, & \text{якщо } \beta_j^p = 0; \beta_j^l = 0, \\ \beta_j^p, & \text{якщо } \beta_j^p = \beta_j^l \text{ або } \beta_j^l = 0, \\ \beta_j^l, & \text{якщо } \beta_j^p = 0. \end{cases} \quad (6.40)$$

Тоді ЗНП можна розглядати як задачу визначення найкоротшого шляху від вершини s у графі $D\Delta$ до всіх інших

вершин цього графа за вагами γ_i і утримуючого найменше число вершин. Покажемо, що введення попередньої розмітки графа $D\Delta$ на основі каліброваних векторів дозволяє звести ЗНП до задачі визначення найкоротшого шляху (КШ) у графі $D\Delta$ від вершини s до всіх інших вершин графа. Слід зазначити, що одержання точного розв'язку задачі про найменше покриття на основі визначення КШ в $D\Delta$, не може бути здійснено традиційними алгоритмами [25] визначення КШ. Оскільки КШ в $D\Delta$, що відповідає мінімальному покриттю, не задовольняє властивість того, що кожний відрізок цього КШ є КШ між відповідними вершинами в $D\Delta$. Відсутність цієї властивість для КШ в $D\Delta$ і спричиняється NP-Повноту [61-63] задачі про найменше покриття. Тому, якщо намагатися розв'язувати задачу (6.37-6.38) методами визначення КШ, що наводяться у [25], то можна розраховувати тільки наближені розв'язки задачі (6.37-6.38).

Уведемо поняття каліброваних векторів $H^{(i)}$ висот h_j^i , $i = (\overline{1, n})$, $j = (\overline{1, m})$. Формування векторів почнемо з останньої вершини n графа $D\Delta$. Перший калібрований вектор $H^{(n)}$ одержуємо таким чином. Усі висоти h_j^n , для яких β_j^n в $B^{(n)}$ рівні 0, одержують значення $h_j^n = \infty$, а ті, для яких $\beta_j^n = j$, значення $h_j^n = 0$. Наступний калібрований вектор $H^{(n-1)}$ будується на основі попереднього вектори $H^{(n)}$ і поточного значення вектори $B^{(n-1)}$.

У загальному випадку на k -й ітерації процес формування вектори $H^{(k)}$ можна записати у вигляді такої рекурентної функціональної залежності:

$$H^{(k)} = f(B^{(k)}, H^{(k+1)}). \quad (6.41)$$

При цьому висоти h_j^k вектори $H^{(k)}$ визначаються значеннями компонентів β_j^k і висот h_j^{k+1} вектори $H^{(k+1)}$ таким чином:

$$h_j^k = \begin{cases} 0, & \text{якщо } \beta_j^k = j (\forall h_j^{(k+1)}), \\ h_j^{(k+1)} + 1, & \text{якщо } h_j^{(k+1)} \neq \infty, \beta_j^k = 0, \\ \infty, & \text{якщо } h_j^{(k+1)} = \infty, \beta_j^k = 0. \end{cases} \quad (6.42)$$

На основі співвідношень (6.40) – (6.42) калібровані вектори будуються для всіх вершин графа $D\Delta$, починаючи з вершини $i = n$ і закінчуючи вершиною $i = 1$, тобто одержимо послідовність каліброваних векторів

$$H^{(n)}, H^{(n-1)}, \dots, H^{(1)} \leftrightarrow n, n-1, \dots, 1$$

відповідних вершинам $n, (n-1), \dots, 1$ у графі $D\Delta$.

Фізичний зміст висоти h_j^i , якщо вона не дорівнює 0 або ∞ , у тому, що вона вказує номер вершини $x = h_j^i + i$ в графі $D\Delta$, приєднання якої до деякого шляху в графі $D\Delta$ приводить до покриття рядка j одиницею в матриці B . Якщо $h_j^i = \infty$, то це означає, що в графі $D\Delta$ не існує шляху від вершини i , що дозволяє утворювати покриття рядка j у матриці B . Випадок рівності $h_j^i = 0$ означає, що вектор $B^{(i)}$ покриває рядок j у матриці B .

Далі, після калібрування, шляху $\mu_{st}^{p,l,\dots,t}$ в графі $D\Delta$, що проходить через вершини p, l, \dots, t , будемо ставити у відповідність об'єднання каліброваних векторів цих вершин

$$H^{(p)} \cup H^{(l)} \cup \dots \cup H^{(t)}.$$

Визначимо правила їхнього об'єднання. Оскільки вершини в $D\Delta$ пронумеровані в порядку зростання при переході від однієї горизонтальної лінійки графа $D\Delta$ до іншої, а сам граф орієнтований і ациклічний, то при формуванні шляху, наприклад, що складається з двох вершин p і l , ми будемо розглядати об'єднання векторів $H^{(p)} \cup H^{(l)}$ ($p < l$), не може мати місце об'єднання $H^{(l)} \cup H^{(p)}$. Тобто операція об'єднання тут не комутативна, що відповідає спрямованості шляхів в $D\Delta$. При об'єднанні довільної пари векторів $H^{(p)} \cup H^{(l)}$ ($p < l$) поєднуються попарно його однойменні висоти $h_j^{i=p} \cup h_j^{i=l}$.

Позначимо $h_j^{i=p} = x$, $h_j^{i=l} = y$. Тоді правила об'єднання висот можна записати у такому вигляді:

$$\begin{aligned}
x \cup y &= y, \quad \forall x \neq 0, \\
x \cup 0 &= 0, \\
0 \cup y &= 0.
\end{aligned}
\tag{6.43}$$

Правило (6.43) для об'єднання каліброваних векторів $H^{(i)}$, тобто формування шляхів в $D\Delta$, дозволяє ввести вагову характеристику d_i^n вектори $H^{(i)}$, рівну числу різних висот h_j^i , не рівних 0, в об'єднаному векторі $H^{(i)}$. Наприклад, у векторі $H^{(i)}=(10012014)$ містяться три одиниці, двійка й четвірка, отже, вагова характеристика цього вектори $d_i^n=3$. Фізичний зміст цієї характеристики в тому, що вона прогнозує існування не більше d_i^n вершин в $D\Delta$, об'єднання з якими дозволить одержати шлях з об'єднаним вектором $H^{(i)}=(0,0,0\dots 0)$; $d_i^n=0$, тобто шлях, що відповідає покриттю всіх рядків стовпцями, що ввійшли в цей шлях, у матриці B . Якщо ж у векторі є хоча б одна висота $h_j^i = \infty$, то й $d_i^i = \infty$, і це означає, що не існує продовження цього шляху в $D\Delta$, яке б відповідало покриттю в матриці B .

Отже, вагова характеристика d_i^n об'єднаного вектори деякого шляху μ_{st} у графі $D\Delta$ визначає, яку найбільшу кількість вершин необхідно приєднати до цього шляху, щоб одержати шлях з вагою $d_i^n=0$. Таким чином, якщо для кожної вершини в графі $D\Delta$ побудований калібрований вектор $H^{(i)}$ і для кожного $H^{(i)}$ визначена вагова характеристика d_i^n , то найкоротший шлях за вагами d_i^n у графі $D\Delta$ від вершини s до всіх інших вершин графа $G\Delta$ визначає мінімальне покриття матриці B стовпцями, що ввійшли в цей шлях.

Твердження 6.6. Якщо в $D\Delta$ побудований шлях $\mu_{st}^{p,l,\dots,t}$, що проходить через вершини p,l,\dots,t , довжина якого за вагами $d_i^n=0$, тобто йому відповідає об'єднаний калібрований вектор

$$H^{(p,l,\dots,t)} = (h_1^{p,l,\dots,t} = 0, h_2^{p,l,\dots,t} = 0, \dots, h_n^{p,l,\dots,t} = 0); d_{p,l,\dots,t}^i = 0,$$

а вершини, що ввійшли в цей шлях, утворюють покриття матриці B .

Доведення: Рівність $h_j^{p,l,\dots,t} = 0$ означає, що $\beta_j^{p,l,\dots,t} = j$, що відповідає, а об'єднаний вектор $B^{(p,l,\dots,t)}$ по даному шляху має вигляд

$$(\beta_1^{p,l,\dots,t} = 1, \beta_2^{p,l,\dots,t} = 2, \dots, \beta_m^{p,l,\dots,t} = m),$$

тобто дає покриття всіх рядків одиницями стовпців, що ввійшли в шлях, у матриці B , що й було потрібно довести.

Отже, якщо побудувати деяку процедуру A , що дозволяє ідентифікувати найкоротший шлях у вершинно зваженому графі $D\Delta$ за вагами d_i^n , то цей шлях буде відповідати об'єднанню векторів $B^{(i)}$, що утворюють мінімальне покриття в матриці B . Для зручності опису такої процедури в графі $D\Delta$ уведена фіктивна вершина s , завдяки якій ранг шляху (число ребер, що утворюють шлях) збігся з числом різних вершин у графі $D\Delta$. Розглянемо варіанти побудови правил відсікання $\{L_w\}$ для розв'язання ЗНП.

6.5.3. Правила відсікання безперспективних варіантів при розв'язанні ЗНП на основі рангового підходу

Почнемо розгляд правил відсікання шляхів у множинах m_{si}^r з варіанта незваженої задачі про найменше покриття, коли ваги у функціоналі (6.37) $c_1 = c_2 = \dots = c_n = 1$. Найбільш простим і природним правилом відсікання шляхів L_1 у множинах $m_{si}^r, i = (\overline{r, n})$ є видалення шляхів $\mu_{si}^r \in m_{si}^r$, у яких $d_i^n = \infty$, оскільки ці шляхи не задовольняють властивість v . Надалі ми будемо використовувати поняття песимістичного гарантованого прогнозу d_i^n числа вершин, приєднання яких до шляху μ_{si}^r дозволить одержати шлях з вектором $H^{(i)}(0,0,\dots,0)$, тобто шлях, що відповідає покриттю в матриці B , і оптимістичного негарантованого прогнозу d_i^o , визначення якому буде дано нижче. Кожна вершина в $D\Delta$ характеризується двома числовими величинами: числом непокритих рядків у векторі $H^{(i)}$ $ai = m - p_i$, де p_i дорівнює кількості нулів в $H^{(i)}$; і числом $\alpha_i = m - \gamma_i$, що визначає максимально можливе число рядків, що може бути

$$z_{i+j} \geq a_i^H ; j = \overline{1, k}; (i + j \leq n). \quad (6.47)$$

Якщо величина песимістичної оцінки шляху $\mu_{si}^r d_i^n = \infty$, то й $d_i^o = \infty$.

Уведемо поняття коридора в множині шляхів M_{si}^r на r ярусі графа $D\Delta$ ($i = \overline{r, n}$). Під виділенням коридора будемо розуміти процес відсівання шляхів у множині M_{si}^r відповідно до правила p .

Правило p

Шлях μ_{sk}^r можна виключити з подальшого аналізу, якщо виконується нерівність

$$d_k^o(\mu_{sk}^r) > d_i^{*i}(\mu_{si}^r), \quad (6.48)$$

де d_i^{*i} – найкращий на даний момент песимістичний прогноз на ранзі r .

Теорема 6.6. Використання правила p у процесі розв'язання ЗНП на основі рангового підходу дозволяє виключати з подальшого аналізу безперспективні шляхи, не втрачаючи при цьому оптимального розв'язання задачі.

Доведення. Припустимо, що, використовуючи правило p , ми відсікли шлях, що надалі дав би оптимальне розв'язку задачі на ранзі $r \sim r+x$. Але в цьому випадку виходить, що $x < d_i^{*i} < d_k^o$, тобто ми одержимо покриття за $x < d_k^o$ кроків, що суперечить визначенню оптимістичного прогнозу, що й було потрібно довести.

У принципі, доведення цієї теореми очевидне, тому що ми відсікаємо лише ті шляхи, оптимістичний прогноз яких гірший, ніж кращий існуючий на даний момент песимістичний. А це означає, що виключений шлях μ_{sk}^r у найкращому разі може дати покриття за $r+d_k^o$ кроків, що явно більше значення $r + d_i^{*i}$, яке ми

одержимо в найгіршому разі, використовуючи шлях μ_{si}^r , песимістичний прогноз якого є на даний момент найкращим.

Таким чином, розглянуті нами правила відсікання безперспективних шляхів μ_{sk}^r у множинах $m_{sk}^r (k = \overline{r, n})$ L_1 і p можуть бути використані для одержання точного розв'язку ЗНП, якщо їх застосувати в процедурі A_0 .

Для зменшення тимчасової складності розв'язку задачі (6.37-6.38) можна використовувати процедуру відсікання L_2 , обумовлену рекурентним співвідношенням

$$\forall (\mu_{si}^r \in m_{si}^r): \left[\mu_{sp}^{r+1} = \min_{d_i^i} \{ \mu_{si}^r \cup (i, p) \} \right] \quad (6.49)$$

$$p = \overline{r+1, n}; i = \overline{r, n},$$

що характеризує процес формування шляхів наступного рангу з урахуванням принципу оптимізації за напрямком. Застосування правила (6.49) дозволяє відсікати більшу кількість шляхів, але при цьому є ймовірність втрати оптимального розв'язання.

Розглянемо тепер випадок, коли c_i у функціоналі (6.37) набувають довільні значення. У цьому випадку змінні в (6.37) необхідно перенумерувати так, щоб виконувалася нерівність

$$c_1 \leq c_2 \leq \dots \leq c_n. \quad (6.50)$$

У зв'язку з тим, що кожному стовпцю в матриці B тепер відповідає вага c_i , виникає необхідність у визначенні оцінок довжини шляху μ_{si}^r з урахуванням ваг вершин: оптимістичної негарантованої d_i^{os} і песимістичної d_i^{ne} .

Для одержання оптимістичної оцінки d_i^{os} ми не можемо скористатися співвідношеннями (6.44-6.47), оскільки характеристика α_i при зміні i змінюється через сортування (6.50). У зв'язку з цим для визначення характеристик d_i^{os} і d_i^{ne} уведемо систему каліброваних векторів $\{y_i\}$ і $\{z_i\}$. Для цього складемо вектор $a(a'_1, a'_2, \dots, a'_n)$, у якому a'_i розташовані в порядку спадання

$$a_1' \geq a_2' \geq \dots \geq a_n'. \quad (6.51)$$

Якщо позначити $a_q' = \alpha_{iq}$, де $q = \overline{1, n}$, а i – відповідає номеру вершини i у графі $D\Delta$, для якої визначалося a_i , то справедлива рівність

$$a = \alpha = (\alpha_{i_1 1}, \alpha_{i_2 2}, \dots, \alpha_{i_n n}) \quad (6.52)$$

Використовуючи вектор α , сформуємо множину векторів $\{y_i\}$. Для одержання вектори y_1 у векторі α викреслюємо елемент α_{i1} і для всіх α_{iq} , у яких $q > 1$, значення другого індексу q зменшуємо на 1. Після цього формуємо компоненти $\{y_{1k}\}$ вектори y_1 за правилом:

$$y_{1k} = \alpha_{1k} + y_{1(k-1)}; \quad k = \overline{0, n-1}; \quad y_{10} = 0. \quad (6.53)$$

Далі у векторі α викреслюємо елемент α_{2j} , змінюючи по q нумерацію так само, як і раніше. За правилом (6.53) продовжуємо формувати компоненти вектори y_2 і т.д. На p -му кроці викреслюємо в α елемент α_{pj} , змінюючи, як і раніше, нумерацію по q . Тоді

$$y_{pk} = \alpha_{pk} + y_{p(k-1)}; \quad k = \overline{0, n-p}. \quad (6.54)$$

Для $p = n$ вектор y_n буде мати тільки один елемент, рівний $\alpha_{nq} = \alpha_n$, де q відповідає номеру цього елемента у векторі α .

Формування величини z_p^o , що є оптимістичним прогнозом ваги вершин, які необхідно приєднати до шляху μ_{sp}^r , щоб можна було одержати покриття, здійснюється на основі такого співвідношення:

$$z_i^o = \sum_{j=p+1}^{y_{ik}^* + p + 1} c_j, \quad (6.55)$$

де y_{ik}^* – елемент у векторі y_i , починаючи з якого виконується нерівність $y_{ik}^* > \gamma_i^{(H)}$.

За допомогою вектори y_i і співвідношення (6.55) легко здійснити негарантований оптимістичний прогноз мінімального числа вершин, які необхідно додати до шляху μ_{si}^r в графі $D\Delta$, щоб одержати покриття в матриці B . Сформулюємо правило k_1 визначення d_i^{o6} і d_i^{n6} .

Правило k_1

1. Обчислюємо для шляху μ_{si}^r у вершині i величину $\gamma_i = m - pi$, де pi – число шляхів у векторі $H^{(i)}$, що відповідає цьому шляху.

2. У каліброваному векторі y_i для вершини i знаходимо номер k^* , починаючи з якого $y_{ik^*} > \gamma_i^{(H)}$. Значення $y_{ik^*} = d_i^o$ дає незважений негарантований оптимістичний прогноз.

3. Визначаємо зважений негарантований оптимістичний прогноз

$$d_i^{\hat{a}} = \sum_{j \in \mu_{si}^r} c_j + z_i^{\hat{t}}. \quad (6.56)$$

4. Величина зваженого песимістичного прогнозу визначається за формулою

$$d_i^{\hat{a}} = \sum_{j \in \mu_{si}^r} c_j + \sum_{j=n-d_i^{\hat{t}}+1}^n c_j. \quad (6.57)$$

Правило виділення коридора p^6 у випадку зважених стовпців у матриці B при розв'язанні задачі (6.37), (6.38) аналогічне (6.48).

Правило p^6

Якщо позначити через $d_l^{* \hat{a}}$ величину найменшого песимістичного прогнозу деякого шляху μ_{sl}^r , тоді всі шляхи, для яких виконується нерівність

$$d_k^{\hat{a}}(\mu_{sk}^r) > d_l^{\hat{a}}(\mu_{sl}^r) \quad (6.58)$$

можна видалити з подальшого аналізу як безперспективні. Правило L_2 , обумовлене співвідношенням (6.49), також залишається колишнім, але при цьому мінімум береться по d_i^{n6} .

$$\forall(\mu_{si}^r \in m_{si}^r): \left[\mu_{sp}^{r+1} = \min_{d_i^{\hat{a}}} \{ \mu_{si}^r \cup (i, p) \} \right], \quad (6.59)$$

$\overline{p=r+1, n}; \overline{i=r, n}.$

6.5.4. Наближені й точні алгоритми розв'язання ЗНП

Розгляд алгоритмів розв'язання ЗНП почнемо з наближеного алгоритму розв'язку задачі (6.37-6.39) для випадку рівності коефіцієнтів c_i з використанням правил об'єднання L_1 і L_2 . Тоді ЗНП можна розглядати як задачу визначення найкоротших шляхів на основі рангового підходу [51] у графі $D\Delta$, з урахуванням принципу оптимізації за напрямком (6.47). Покроковий опис алгоритму визначення найкоротших шляхів у графі $D\Delta$, що задовольняють властивість ν , має нижченаведений вигляд.

Наближений алгоритм розв'язання не зваженої ЗНП (процедура A_k)

КРОК 1. Формуємо калібровані вектори $H^{(i)}$, виділяємо множини шляхів $\{\mu_{si}^{r-1}\}$, що задовольняють властивість ν . Якщо всі ці множини порожні, то алгоритм закінчує роботу, тому що це означає, що покриття не існує. Якщо ж є хоча б одна порожня множина, то визначаємо в кожному з них шляхи μ_{si}^{*r-1} , мінімальні за вагою d_i^n і робимо перехід до кроку 3.

КРОК 2. На основі шляху множини шляхів $\{\mu_{si}^{*r}\}$ поточного рангу формуємо множини шляхів μ_{si}^{r+1} наступного рангу, що задовольняють властивість ν . Виділяємо в них шляхи μ_{si}^{*r+1} з найменшим значенням d_i^n (правило L_2) і переходимо до виконання наступного кроку.

КРОК 3. Якщо $\epsilon d_i^n = 0$, то алгоритм закінчує роботу, інакше – виконуємо крок 4.

КРОК 4. Якщо $\epsilon d_i^n = 1$, то формуємо шлях наступного рангу з $d_i^n = 0$ і алгоритм закінчує роботу, інакше – значення r збільшуємо на 1 і переходимо до кроку 2.

Наведений алгоритм відображає основне функціональне рівняння динамічного програмування, що для даної задачі можна подати в такому вигляді

$$d_{\min}^{r=k}(i) = \min_i [d_{\min}^{r=k-1}(s, j) + d^{r=k}(j, i)], \quad (6.60)$$

де $d_{\min}^{r=k-1}(s, j)$ – умовна мінімальна відстань від вершини s до вершини j в $D\Delta$, знайдена на $(k-1)$ кроці роботи процедури;

$d^{r=k}(j, i)$ – величини ваг вершин, що утворюють шляхи до вершин $i=(1, n)$ на k -му кроці.

Рівняння (6.60) дозволяє по ланцюжку знаходити умовні оптимальні розв'язки.

Таким чином, визначення каліброваних векторів $H^{(i)}$ еквівалентне маркуванню шляхів у графі $D\Delta$ від кінця до початку, а використання потім процедури A_k є ні що інше, як ідентифікація найкоротшого шляху по зробленому маркуванню на основі функціонального рівняння динамічного програмування (6.60), коли процес програмування розгортається від початку до кінця. У випадку зважених стовпців у матриці B покроковий опис наближеного алгоритму A_k^{ϵ} розв'язання задачі (6.37), (6.38) з використанням правила відсікання L_2 залишається тим же, але при цьому пошук найкоротших шляхів іде за ваговою характеристикою $d_i^{n\epsilon}$, обумовленою співвідношенням (6.57).

Наближений алгоритм розв'язання зваженої ЗНП (процедура A_k^{ϵ})

КРОК 1. Формуємо калібровані вектори $H^{(i)}$, виділяємо множини шляхів $\{\mu_{si}^{r=1}\}$, що задовольняють властивість v . Якщо

всі ці множини порожні, то алгоритм закінчує роботу, тому що це означає, що покриття не існує. Якщо ж є хоча б одна непорожня множина, то визначаємо в кожному з них шляхи $\mu_{si}^{*r=1}$, мінімальні за вагою d_i^{n6} , і переходимо до кроку 2.

КРОК 2. Якщо є шляхи з $d_i^{n6} = 0$, то виділяємо серед них шлях, мінімальний за вагою, що визначає локальний екстремум на ранзі, і переходимо до наступного кроку.

КРОК 3. На основі шляху множини шляхів $\{\mu_{si}^{*r}\}$ поточного рангу r формуємо множини шляхів μ_{si}^{r+1} наступного рангу відповідно до правила L_2 і переходимо до виконання наступного кроку.

КРОК 4. Якщо всі множини наступного рангу порожні, то виділяємо з множини локальних екстремумів глобальний і алгоритм закінчує роботу. Якщо ж є хоча б одна непорожня множина, то збільшуємо значення поточного рангу $r:=r+1$ і переходимо до кроку 2.

Точне розв'язання задачі (6.37), (6.38) у випадку рівності коефіцієнтів c_i можливе тільки з використанням правил L_1 і p . При цьому покроковий опис алгоритму, реалізованого на основі процедури A_0 , має нижче поданий вигляд.

Точний алгоритм розв'язання незваженої ЗНП (процедура А)

КРОК 1. Формуємо калібровані вектори $H^{(i)}$, виділяємо множини шляхів $\{\mu_{si}^{r=1}\}$, що задовольняють властивість v . Якщо всі ці множини порожні, то алгоритм закінчує роботу, тому що це означає, що покриття не існує. Якщо ж є хоча б одна непорожня множина, то для всіх вершин графа $G\Delta$ обчислюємо допоміжні вектори z_i , визначаємо для кожного шляху характеристики d_i^n і d_i^o відповідно до співвідношень (6.54) - (6.57) і переходимо до виконання наступного кроку.

КРОК 2. На основі шляху множини шляхів $\{\mu_{si}^{*r}\}$ поточного рангу r формуємо множини шляхів μ_{si}^{r+1} наступного рангу $r+1$ на основі рекурентного співвідношення (6.46), визначаємо характеристики цих шляхів d_i^n і d_i^o відповідно до (6.54) - (6.57) і виділяємо коридор у множині m_{si}^{r+1} за правилом p , обумовленим співвідношенням (6.58).

КРОК 3. Якщо $d_i^n = 0$, то алгоритм закінчує роботу, тому що саме цей шлях утворить мінімальне покриття, інакше – виконуємо крок 4.

КРОК 4. Якщо $d_i^n = 1$, то формуємо шлях наступного рангу з $d_i^n = 0$ і алгоритм закінчує роботу, інакше – значення r збільшуємо на 1 і переходимо до кроку 2.

У випадку, коли стовпці в матриці U зважені, покроковий опис алгоритму розв'язання задачі (6.37), (6.38) набуває нижченаведеного вигляду:

Точний алгоритм розв'язання зваженої ЗНП (процедура A^6)

КРОК 1. Формуємо калібровані вектори $H^{(i)}$, виділяємо множини шляхів $\{\mu_{si}^{r=1}\}$, що задовольняють властивість v . Якщо всі ці множини порожні, то алгоритм закінчує роботу, тому що це означає, що покриття не існує. Якщо ж є хоча б одна непорожня множина, то для всіх вершин графа $D\Delta$ визначаємо калібровані вектори y_i і z_i , обумовлені співвідношеннями (6.51) - (6.55), визначаємо для кожного шляху характеристики d_i^{n6} і d_i^{o6} відповідно до співвідношень (6.56), (6.57), виділяємо в них шляхи $\mu_{si}^{*r=1}$ відповідно до правил L_1 і p^6 і переходимо до виконання наступного кроку.

КРОК 2. Якщо є шляхи з $d_i^{n6} = 0$, то виділяємо серед них шлях, мінімальний за вагою, що визначає локальний екстремум на ранзі, і переходимо до наступного кроку.

КРОК 3. На основі шляху множини шляхів $\{\mu_{si}^{*r}\}$ поточного рангу формуємо множини шляхів μ_{si}^{r+1} наступного рангу на основі рекурентного співвідношення (6.59), визначаємо характеристики цих шляхів d_i^{n6} і d_i^{o6} відповідно до (6.56), (6.57) і виділяємо коридор у множині m_{si}^{r+1} за правилом p^6 , обумовленим співвідношенням (6.58) і переходимо до виконання наступного кроку.

КРОК 4. Якщо всі множини наступного рангу порожні, то виділяємо з множини локальних екстремумів глобальний і алгоритм закінчує роботу. Якщо ж є хоча б одна непорожня множина, то збільшуємо значення поточного рангу $r := r+1$ і

переходимо до кроку 2. Таким чином, показана можливість застосування методу розв'язання оптимізаційних задач з БЗ на основі рангового підходу до розв'язання ЗНП. Оскільки ЗНП поліноміально зводиться до задач: „Найменша домінуюча множина вершин у графі”; „Найбільша незалежна множина вершин у графі”; „Найбільша кліка в графі”; „Найбільше парозбігання в графі”, то розглянуті наближені й точні алгоритми розв'язання ЗНП можуть бути використані й для розв'язання перерахованих задач, що мають широке прикладне значення в теорії складних систем [26].

6.5.5. Особливості рангового підходу при розв'язанні ЗНР

Неважко показати, що, якщо в алгоритмах, розроблених для розв'язання ЗНП, правила об'єднаних векторів (6.43) замінити на (6.61)

$$\begin{aligned}
 x \cup y &= y, \text{ якщо } x \neq 0 \\
 x \cup 0 &= 0 \\
 0 \cup y &= 0 \\
 0 \cup 0 &= \infty,
 \end{aligned}
 \tag{6.61}$$

то їх можна застосувати для розв'язання ЗНР.

У роботах [81-84] показано, що при використанні рангового підходу в наближених алгоритмах на похибку розв'язання істотно можуть впливати сортування векторів n -вимірною одиничного куба в графі DA . У зв'язку з цим для зважених ЗНП і ЗНР уведемо коефіцієнти: χ_j визначальне число одиниць у векторі $B^{(i)}$ і $\lambda_j = c_j / \chi_j$, що характеризує частину вагової характеристики C_j доводиться на кожен одиницю вектори $B^{(i)}$. Тоді, якщо перед роботою алгоритму застосувати замість сортування (6.51) сортування векторів $B^{(i)}$ у порядку спадання

$$\lambda_1 > \lambda_2 > \dots > \lambda_{n-1} > \lambda_n,$$

то проведемо дослідження впливу різних сортувань на похибку розв'язання зважених задач ЗНП і ЗНР.

При розв'язанні ЗНР на деякій множині $A=\{a_i\}$ на основі уведеної узагальненої процедури A_0 виникає питання про можливість розв'язання задачі ЗНР, тобто про існування покриття з непересічних елементів заданої потужності m . Умову існування розв'язання ЗНР можна сформулювати у вигляді теореми 6.7.

Теорема 6.7. ЗНР має розв'язок тоді й тільки тоді, коли підмножина $A=\{a_i\}$ є розбиттям числа m .

Доведення Припустимо, що в A немає $\{a_i\}$ відповідній розбиттю m і побудована підмножина $L = \bigcup_{i=1}^z l_{si}^{r=1}$, що утворює покриття P потужності m , тоді, якщо для $\forall l_{si}^{r=1} \in L$ справедлива рівність $\bigcap_{i=1}^z l_{si}^{r=1} = \emptyset$, то з огляду на те, що $l_{si}^{r=1} = a_i$, повинна виконуватись рівність $\sum_i a_i = m$, але це суперечить первісному припущенню про відсутність розбиття числа m і, отже, L у цьому випадку може бути покриттям тоді й тільки тоді, коли виконуються умови теореми.

Наслідки. Якщо $a_1=a_2=\dots a_q=a$, то ЗНР має розв'язок тільки в тому випадку, якщо m ділиться на a без остачі.

Доведення: Якщо деякий шлях μ_{sv} у графі G_Δ рангу r задовольняє умову $\bigcap_{i=1}^r l_{si}^{r=1} = \emptyset$, де $l_{si}^{r=1} \in \mu_{sv}$, то ясно, що довжина цього шляху буде $d(\mu_{sv})=r*a$. Нам необхідно знайти шлях μ_{st}^* довжини $d(\mu_{st}^*)=\min_r \{ r*a \}$, де $r = (\overline{1, n})$, що еквівалентно знаходженню такого $r = r$, при якому $ra = m$, звідки $r = m/a$, але ранг шляху завжди є цілим числом, що й було потрібно довести.

У випадку, коли a_i різні, визначити, чи є розв'язок, ми зможемо тільки після закінчення роботи алгоритму на основі процедури A_0 . Розв'язок буде відсутній тільки в тому випадку, коли всі побудовані підмножини шляхів, за допомогою процедури A_0 і вироблених правил відсікання безперспективних варіантів, виявляться порожніми.

6.5.6. Ранговий підхід до розв'язання діофантових рівнянь з БЗ

Для розв'язання діофантових рівнянь з булевими змінними

$$\sum_{i=1}^n \alpha_i x_i = b \quad (6.62)$$

і систем діофантових рівнянь

$$\sum_{i=1}^n \alpha_{ij} x_i = b_j, \quad j = (\overline{1, m}) \quad (6.63)$$

відомі алгоритми, що здійснюють тільки повний перебір варіантів на бінарному дереві рішень, хоча розв'язання рівнянь (6.62), (6.63) має важливе значення в теорії складання розкладів і в теорії розпізнавання образів.

У літературі ці задачі відомі як задачі про камені [24]. Дано n каменів з цілими вагами a_1, a_2, \dots, a_n , потрібно вибрати деяку кількість з них так, щоб сума їхніх ваг дорівнювала даному числу b . Уведемо двійкові змінні x_1, x_2, \dots, x_n , тоді розглянута задача – це рівняння

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b. \quad (6.64)$$

У задачі про камені, описуваній співвідношенням (6.64), або немає розв'язку, або розмірність розв'язку менше розмірності опису розглянутого варіанта. Можливе й формулювання багатовимірної задачі про камені. Нехай камені мають кілька параметрів ваги $a_1^1, a_2^1, \dots, a_n^1$, об'єми $a_1^2, a_2^2, \dots, a_n^2$ й ціни в різні часи $a_1^j, a_2^j, \dots, a_n^j$, потрібно знайти значення двійкових змінних x_1, x_2, \dots, x_n , щоб виконувалися рівності

$$a_1^j x_1 + a_2^j x_2 + \dots + a_n^j x_n = b_j, \quad j = (1, 2, \dots, m). \quad (6.65)$$

Значення a_i^j і b_j – цілі числа, звичайно не від'ємні. Часто розглядають задачу з нерівностями, які, як відомо, за рахунок

збільшення розмірності задачі зводяться до задачі з рівностями. У [87] показано, що задача про камені належить до класу РС, тому розглянемо можливості побудови точного алгоритму розв'язку задачі на основі рангового підходу, уводячи відсікання безперспективних варіантів на основі поняття коридора. За аналогією з задачею лінійного програмування з булевими змінними зведемо задачу розв'язання систем діофантових рівнянь до задачі визначення шляхів заданої довжини в деякому ациклічному графі $D\Delta$. Ребрам, що входять у вершини графа $D\Delta$, що відповідають змінним x_1, x_2, \dots, x_n , присвоїмо ваги a_i^j , що стоять при x_i у рівнянні. Якщо ми будемо розглядати систему рівнянь, то ребрам, що входять у вершини $1, 2 \dots n$, будемо присвоювати кілька ваг a_i^j . Тоді задача розв'язання рівняння (6.64) зводиться до визначення в ациклічному графі $D\Delta$ шляхів довжиною b . Якщо розглядати систему рівнянь (6.63), то в графі $D\Delta$ ребру відповідає m ваг, і для її розв'язання потрібно знайти такий шлях, щоб його довжина за всіма вагами дорівнювала відповідно b_j ($j = 1, 2, \dots, m$). Алгоритм розв'язання даної задачі має нижчеподаний вигляд.

Алгоритм розв'язання діофантового рівняння з булевими змінними

КРОК 1. З вершини S формуються множини шляхів m_{sj}^r рангу $r = 1, j = (1 \div n)$, довжини яких $d\mu_{sj}^r \leq b$. Визначаємо коефіцієнти

$$\gamma_j = c_{j+1} + c_{j+2} + \dots + c_n \quad (6.66)$$

КРОК 2. Формуються множини шляхів $m_{sj}^{r=r+1}$ ($j = r + 1, n$) наступного рангу $r = r + 1$ на основі множини шляхів m_{sj}^r попереднього рангу r у відповідності до таких рекурентних співвідношень:

$$\begin{aligned} \mu_{sj}^{r=r+1} &= \{m_{sj} \cup (i, j)\}, \\ d\mu_{sj}^r &\leq b, j = (r \div n), i = (r \div n), j \neq i. \end{aligned} \quad (6.67)$$

При цьому враховується така умова:

$$d(\mu_{sj}^{r=r+1}) + \gamma_j < b. \quad (6.68)$$

І якщо воно не виконується, то шлях виключається з подальшого аналізу як неперспективний.

КРОК 3. Перевіряємо множина шляхів $m_{sj}^{r=r+1} = \emptyset$. І якщо вона порожня, то алгоритм закінчує роботу, у противному випадку здійснюється перехід до кроку 2.

Запропонований алгоритм являє собою процедуру перерахування корінь рівняння (6.62) з виділенням коридора на основі співвідношень (6.67), (6.68) і видалення шляхів, довжини яких $d\mu_{sj}^r \leq b$. У випадку розв'язання системи рівнянь (6.63) доцільно розв'язувати одне будь-яке рівняння й перевіряти потім підстановкою перерахованих корінь в усі рівняння, що залишилися, сумісність розглянутої системи.

6.6. Експериментальне дослідження алгоритмів розв'язання задачі ЦЛП з БЗ на основі рангового підходу

6.6.1. Показники ефективності алгоритмів

Порівняння розроблених алгоритмів розв'язання задачі (0, 1)-рюкзак проведений з адитивним алгоритмом Балаша, Н-методом і методом вектори спаду. Алгоритм Балаша на графіках позначений як алгоритм *D*. Порівняння алгоритмів розв'язання ЗНП проведено з кращими алгоритмами, що реалізують ідеї методу галузей і границь. Порівняння й дослідження властивостей алгоритмів проводилося за нижченаведеними параметрами.

1. Число елементарних операцій (ЕО) додавання й порівняння, виконуваних алгоритмами.

2. Ємність пам'яті *V* для реалізації конкретного алгоритму.

3. Для оцінки якості наближених алгоритмів обрані абсолютна й відносна похибки.

Для об'єктивного порівняння розроблених алгоритмів з відомими, а також з урахуванням їх специфіки, додатково будемо використовувати нижчеподані показники якості:

1. Число ЕО додавання й порівняння, виконуваних алгоритмами.

2. t_{cp} – середній час розв’язання задачі, с. (цей параметр залежить від обраної мови програмування, технічних даних ЕОМ, кваліфікації програміста).

3. K_n – кількість неточних розв’язків, що припадають на 100 тестових задач

4. Похибка наближених рішень Δf , %, дорівнює

$$\Delta f = \frac{|f^* - f^{\sim}|}{f^*},$$

де f – значення точного розв’язку тестової задачі;

f^{\sim} – значення наближеного розв’язку тестової задачі.

5. r_{cp} – середній ранг шляхів, що відповідають оптимальному розв’язку тестової задачі, обумовлений числом одиниць у векторі x .

6. Коефіцієнт, що характеризує якість фільтрації векторів при виділенні m -вимірного коридора в множинах на всьому ярусі в порівнянні з виділенням m -вимірного коридора в множинах, рівний

$$K_{\phi} = \frac{K(\bigcup_{j=r}^n m_{sj}^r)}{K(m_{sj}^r)},$$

де $K(\bigcup_{j=r}^n m_{sj}^r)$ – число векторів, що залишилися в коридорі, при виділенні його на ярусі;

$K(m_{sj}^r)$ – число векторів, що залишилися в коридорі, при виділенні його в множинах m_{sj}^r .

7. $K_y = \text{EO}(\Pi) / \text{EO}(n)$ – коефіцієнт прискорення, що може бути отриманий при реалізації розроблених алгоритмів на n процесорах у порівнянні з послідовними реалізаціями алгоритмів.

Такі показники, як займана ємність пам'яті V , Δf , \bar{r} , K_n і K_ϕ , утворюють групу $\{\Phi_k\}$ -показників, що не залежать від часу виконання алгоритму, а тимчасові показники: число N_{on} і K_v , утворюють групу $\{\Theta_m\}$ -показників.

Використання показників функціональної групи дозволяє оцінити ефективність застосування правил відсікання $\{L_w, K_i\}$ безперспективних варіантів в узагальненій процедурі A_0 , а група тимчасових показників дозволяє порівняти розроблені алгоритми з відомими. Такий підхід широко використовується при експериментальному дослідженні алгоритмів [78-81].

Для перевірки правильності й точності розв'язання розробленими алгоритмами $A_1 - A_{12}$ як еталон обраний модифікований алгоритм Балаша [23], що враховує вимоги позитивності коефіцієнтів у функціоналі. Надалі під алгоритмом D будемо розуміти саме цей алгоритм.

У ході розв'язання тестових задач за допомогою датчика випадкових чисел генерувалися коефіцієнти у функціоналі (6.7) у діапазоні $[1 \div 100]$ і в обмеженнях (6.8), (6.9) у діапазоні $[1 \div 10]$. Вибір інших діапазонів для функціонала змінив лише його абсолютне значення, але в середньому не вплинув на параметри алгоритмів. Зміна діапазону в обмеженнях впливає лише на ранг шляху (ранг шляху – кількість вершин графа $D\Delta$, які утворюють цей шлях).

Структура дослідження кожного алгоритму полягала в ось у чому. Розв'язувалося 1000 тестових задач з заданими вхідними параметрами m і n обраним алгоритмом. У ході розв'язання обчислювалися показники ефективності алгоритмів. Отримані значення показників наведені на графіках, апроксимованих за методом найменших квадратів.

Усі алгоритми розділимо на три групи. Першу групу становлять одновимірні алгоритми A_1, A_3, A_6 , що розв'язують задачу (6.10), (6.11). Другу групу становлять m -вимірні алгоритми A_4, A_5, A_7, A_8 , що розв'язують задачу (6.7)-(6.9). А третю групу утворюють багатоетапні одновимірні й m -вимірні алгоритми $A_9, A_{10}, A_{11}, A_{12}$.

Алгоритм A_2 не ввійшов у жодну з груп, тому що він, з одного боку, є одновимірним, а з другого – використовується як

один з етапів у багато етапних одновимірних алгоритмах. Найбільший інтерес являє не сам алгоритм A_2 , а властивість стратегії L_2 , на основі якої він побудований. Ця властивість використовується в багато етапних алгоритмах з метою більш ефективного відсікання векторів на першому ранзі. Тому, дослідження алгоритму A_2 не проводилося.

6.6.2. Результати експериментального дослідження алгоритмів

Метою експериментального дослідження є:

1. Практична перевірка правильності висунутих стратегій відсікань $\{L_w\}$ і $\{K_i\}$, заснованих на доведених у розд. 2 теоремах і твердженнях;
2. Оцінка тимчасової складності запропонованих алгоритмів;
3. Одержання числових значень обраних показників якості алгоритмів у середньому;
4. Порівняння показників ефективності розроблених алгоритмів з відомими алгоритмами;
5. Аналіз причин одержання неточних розв'язків наближеними алгоритмами.

Дослідження алгоритмів першої групи. Як виявилось, кількісні значення обраних показників істотно залежать від рангу \bar{r} одержуваного розв'язку, що визначає число одиниць в оптимальному розв'язку. Це підтверджують рис. 6.25 і 6.26, на яких подані залежності тимчасового показника N_{on} алгоритмів від \bar{r} при фіксованих значеннях вхідних параметрів n і $m=1$.

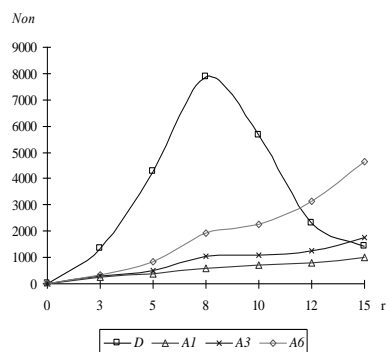


Рис.6.25. Залежність N_{on} від \bar{r} при $n=15$

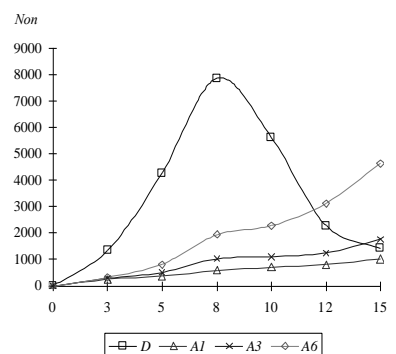


Рисунок 6.26. Залежність N_{on} від \bar{r} при $n=20$

З рис.6.25 і 6.26 видно, що діапазон зміни \bar{r} можна умовно розбити на три зони: 1-ша зона - $\bar{r} = [0 \div n/3]$; 2-га зона - $\bar{r} = [n/3 \div 2n/3]$; 3-я зона - $\bar{r} = [2n/3 \div n]$. У 1-й зоні алгоритм D знаходить розв'язок швидко, оскільки за рахунок зондування дуже ефективно відсікаються області дерева розв'язків, що відповідають одиничним розгалуженням. Аналогічно пояснюється й швидке одержання розв'язку в 3-й зоні, вектори в якій складаються з великої кількості одиниць. Прояв усієї експонентної складності алгоритму D відбувається саме в 2-й зоні. Для алгоритмів A_1, A_3, A_6 ріст числа ЕО зі збільшенням \bar{r} визначається числом локальних областей W , які необхідно обробити. Згідно з (6.10) кількість таких областей W у графі DA не перевищує $n^2/2$, тому при $\bar{r} \rightarrow n$ ріст числа ЕО також не перевищує $c \cdot n^2$, де $c = \text{const}$.

Таким чином, для об'єктивного порівняння алгоритмів необхідно вказувати до якої зони вони належать, тобто який відсоток одиниць (або нулів) містить оптимальний розв'язок. Відповідно до поділу на зони побудуємо залежність числа ЕО від n для кожної зони. Ці залежності для 1-3-ї зон показані на рис. 6.27- 6.29.

Умовимося, що для одержання величини якого-небудь показника для зони будуть використані його усереднені значення на рангах \bar{r} , що належать цій зоні.

З наведених рис. 6.27-6.29 видно, що для порівняння тимчасових показників алгоритмів необхідно чітко вказувати приналежність до умовних зон, тому що абсолютне значення цих показників у зонах істотно відрізняється. Потрапляння розв'язання задачі в будь-яку зону саме визначається співвідношенням діапазону зміни коефіцієнтів a_{1j} в обмеженні (6.9) і заданого діапазону для b_1 у правій частині нерівності (6.9).

Показник необхідної ємності оперативної пам'яті V залежить від конкретної програмної реалізації. Кількісні значення цього показника, природно, можуть бути змінені й прямо залежать від кваліфікації програміста. Однак, вид кривих не зазнає зміни через природу тих процесів, які вони описують. Покажемо, що і величина необхідної ємності оперативної пам'яті

для зберігання множини m_{sj}^r векторів графа $D\Delta$ також залежить від середнього рангу оптимального розв'язання тестової задачі. Залежність $V = f(\bar{r})$ для алгоритмів D , A_1 , A_3 , A_6 наводиться на графіках (рис. 6.30, 6.31).

Менше значення необхідної ємності оперативної пам'яті для алгоритму D пояснюється тим, що у випадку успішного зондування по (1) і (0) підмножин альтернатив нова задача, а отже, новий запис у пам'яті, не утворюється, тому що коректується список вільних вершин для обраної задачі. Тільки утворення нового розгалуження, що відповідає кроку 6 алгоритму, вимагає витрат оперативної пам'яті. Для алгоритмів A_1 , A_3 , A_6 максимальне число оброблюваних векторів знаходиться у 2-й зоні. Тому для зберігання потрібно більше пам'яті, ніж для 1-ї зони. Максимального значення витрати досягають у 3-й зоні, тому що в цьому випадку відмінність за вагами функціонала всередині множин незначна, що приводить до неефективного відсікання векторів за стратегією L_3 .

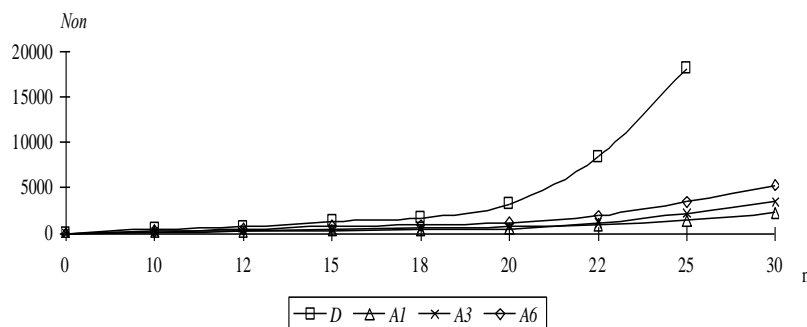


Рис. 6.27. Залежність N_{on} від n для 1-ї зони

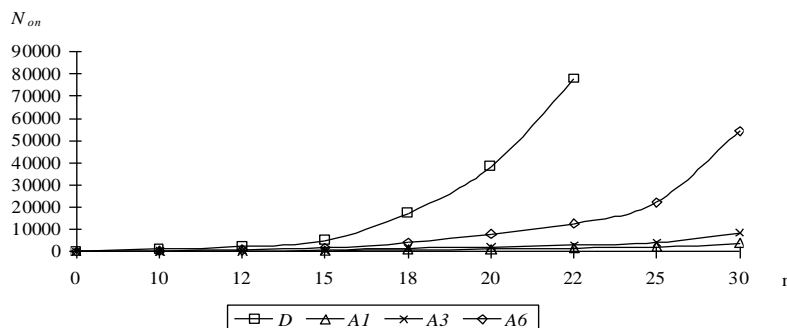


Рис. 6.28. Залежність N_{on} від n для 2-ї зони

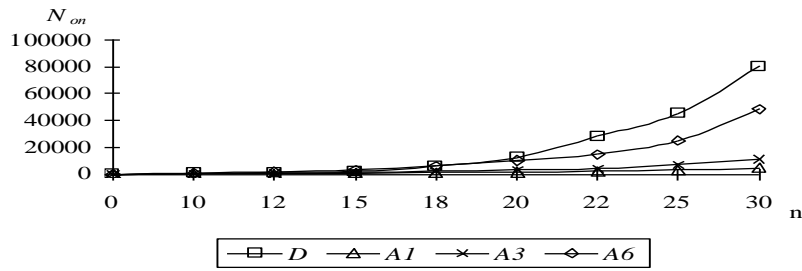


Рис. 6.29. Залежність N_{on} від n для 3-ї зони

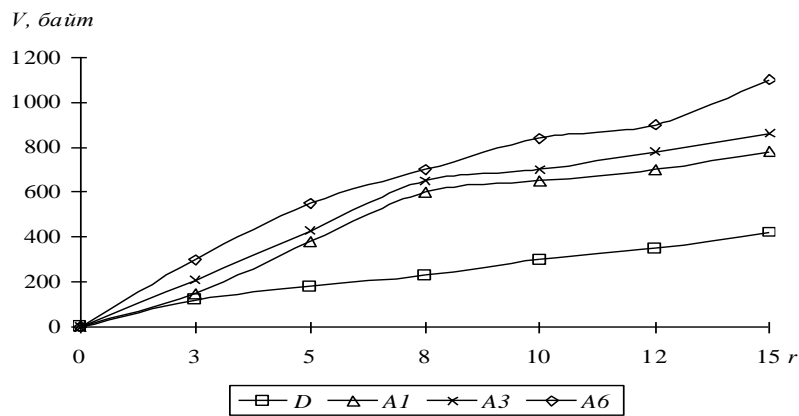


Рис. 6.30. Вплив r на V при $n=15$

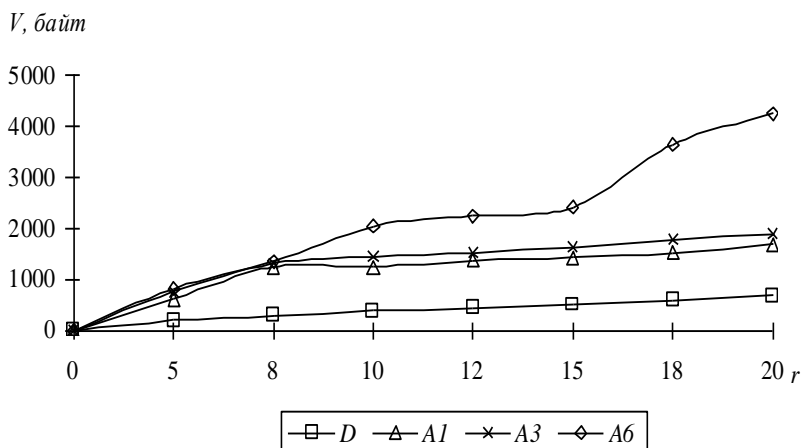


Рис.6.31. Вплив r на V при $n = 15$

Залежності величини V від розміру вхідного параметра n також не перевищують $c \cdot n^2$. Результати такого дослідження наведені на рис. 6.32 - 6.34, що відповідають 1-3-й зонам.

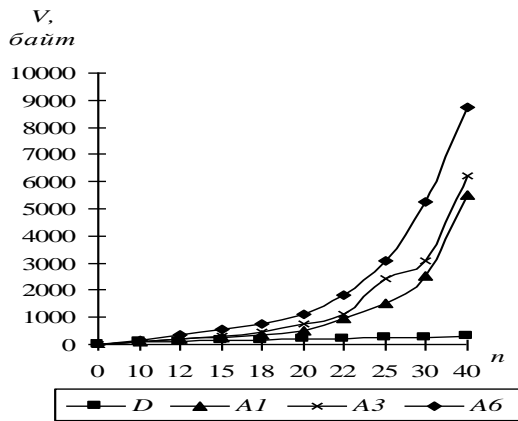


Рис. 6.32. Залежність $V=f(n)$ для 1-ї зони

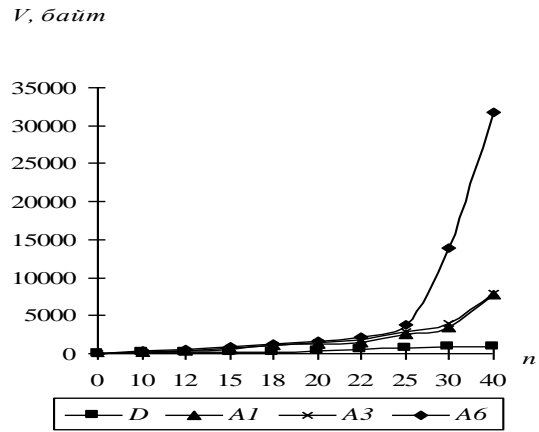


Рис. 6.33. Залежність $V=f(n)$ для 2-ї зони

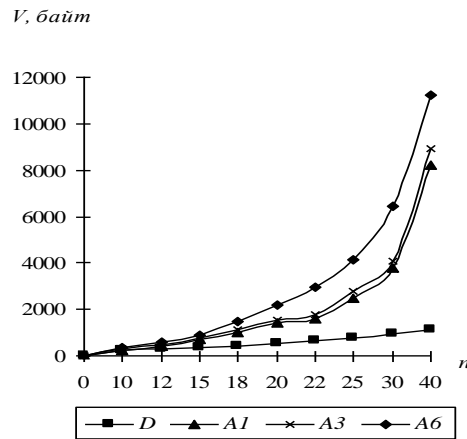


Рис. 6.34. Залежність $V=f(n)$ для 3-ї зони

Тимчасову складність алгоритмів A_1 і A_3 досить легко оцінити теоретично. Аналіз кожного вектори \vec{x} включає одну операцію додавання й одну операцію порівняння, тобто дві елементарні операції. Максимальне число векторів у найгіршому разі буде на другому ранзі й становить $n^2/2$. Число таких рангів у графі $D\Delta$ дорівнює n . Отже, тимчасова складність алгоритму A_1 складе $O(2n \cdot n^2/2) = O(n^3)$, для зберігання векторів в Ω -областях на всіх рангах буде потрібно пам'яті не більш ніж $O(n^3)$.

Оцінка тимчасової складності алгоритму A_2 збігається з оцінкою для алгоритму A_1 . Оцінимо тимчасову складність алгоритму A_3 . Тимчасова складність алгоритму A_1 становить $O(n^3)$. Однак в A_3 кожна множина буде містити після виділення

коридора й фільтрації в найгіршому разі два шляхи, що збільшить тимчасову складність алгоритму A_3 у два рази й складе $O(2n^3) \cong O(n^3)$.

Однією з важливих характеристик наближених алгоритмів служить відносна похибка Δf , що вона дає у випадку розв'язання задачі. З розглянутих алгоритмів наближеними є A_1 і A_3 . Відносна похибка залежить від зони, у яку потрапить оптимальний розв'язок. Однак одержання неточних рішень алгоритмами A_1 і A_3 для різних зон у середньому викликало певні утруднення. Так, для 1-ї зони експонентний ріст векторів у графі $D\Delta$ натрапляє на обмеження (6.11) і тому в множинах оптимальні шляхи будуть домінуючими за вагами функціонала. У 3-й зоні оптимальний шлях буде проходити по більш високих вершинах графа $D\Delta$, що не дозволить його відсікти. Тому на рис. 6.35 показана зміна похибки Δf і коефіцієнта K_n залежно від розмірності n тільки для 2-ї зони.

Як видно з рис. 6.35, для обох алгоритмів A_1 і A_3 середня похибка Δf стабілізується близько 5 % для першого алгоритму й близько 2 % для другого при $n > 25$.

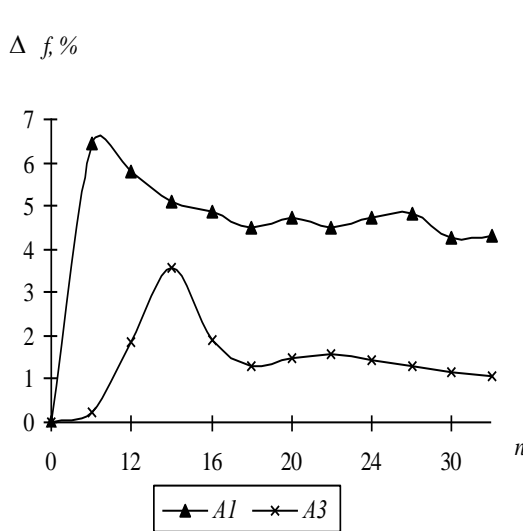


Рис. 6.35. Залежність Δf від n для 2-ї зони

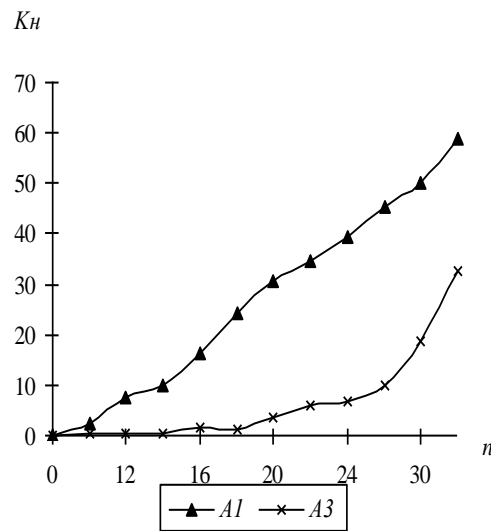


Рис. 6.36. Залежність K_n від n для 2-ї зони

Дослідження алгоритмів другої групи. Оцінимо властивість алгоритмів другої групи A_4 , A_5 , A_7 , A_8 за обраними показниками ефективності. Алгоритми A_4 і A_5 є наближеними, а

A_7 і A_8 – точними. Точно також як і для одновимірних задач, значення зазначених показників залежить від \bar{r} , що підтверджує рис. 6.37. Криві на цьому рисунку ідентичні за своєю формою відповідним показникам одновимірних алгоритмів. Тому зроблений поділ за зонами правочинний й для багатовимірних алгоритмів, а залежності обраних показників від n для кожної зони також збігаються з відповідними залежностями для одновимірних алгоритмів.

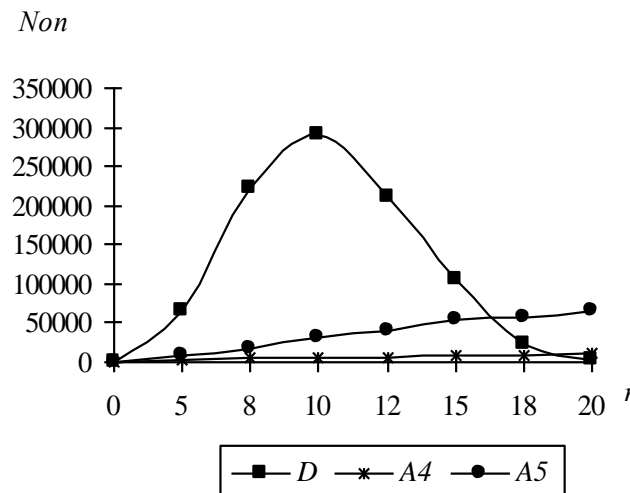


Рис. 6.37. Залежність N_{on} від \bar{r} для $n=20$ і $m=5$

У зв'язку з необхідністю виконання t раз перевірки умови (6.8) тимчасова складність алгоритму A_4 становитиме $O(2mn^3) \cong O(mn^3)$. Тимчасова складність алгоритму A_5 через обрану стратегію L_{11} в t раз більше, ніж тимчасова складність алгоритму A_4 і становить $O(mn^3t) = O(m^2n^3)$, тому що з кожної множини, згідно з (6.8), необхідно вибирати в найгіршому разі на t векторів більше. Отже, і величина апаратних витрат пам'яті, використовуваної для зберігання шляхів, також в t раз збільшиться й дорівнюватиме $O(mn^3)$. При $t = 5$ зроблене дослідження похибки алгоритмів A_4 і A_5 . Залежності f і K_n від n і t для 2-ї зони подані у вигляді графіків на рис. 6.38.

Якщо порівняти рис. 6.35 і рис. 6.38, то видно, що відносна похибка Δf не залежить від t , а залежить тільки від обраного алгоритму. Таким чином, можна стверджувати, що алгоритм A_4 має похибку – 5 %, а алгоритм A_5 – 2 %.

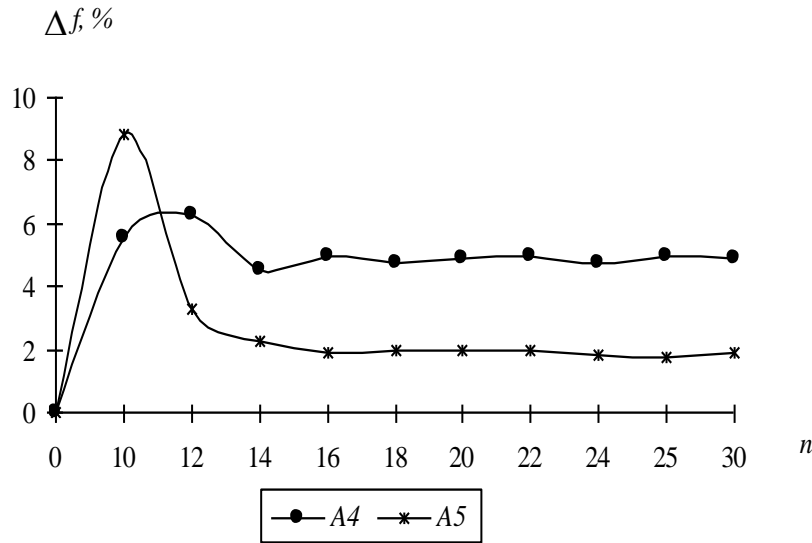


Рис. 6.38. Залежність Δf від n для 2-ї зони

Для точних алгоритмів A_7 і A_8 використовується підхід з виділенням m -вимірного коридора. При цьому отримані алгоритми мають експонентну складність, що полягає в експонентному числі векторів, які можуть потрапити в цей коридор при побудові шляхів, починаючи з рангу $r = 1$. Тому використання цих алгоритмів більш ефективно при вже знайденому припустимому розв'язку яким-небудь з наближених алгоритмів, а оцінка значень показників ефективності буде зроблена при побудові багато етапних точних алгоритмів.

Порівняємо алгоритми A_7 і A_8 між собою. Як основний критерій такого порівняння обраний коефіцієнт фільтрації K_ϕ . Для найгіршого розподілу (2-га зона) залежність K_ϕ від n показана рис. 6.39.

Як видно з рис. 6.39, зі збільшенням m при $n = \text{const}$ K_ϕ прагне до 1. Це пояснюється тим, що зменшується ймовірність існування шляху в більш високих вершинах графа $D\Delta$ на поточному ранзі r , що має менше значення за вагами всіх обмежень і більше значення за вагами функціонала.

Дослідження алгоритмів третьої групи. Проведемо експериментальне дослідження алгоритмів третьої групи: $A_9, A_{10}, A_{11}, A_{12}$. Перші два належать до одновимірних багато етапних алгоритмів. Другі – до багатомірних.

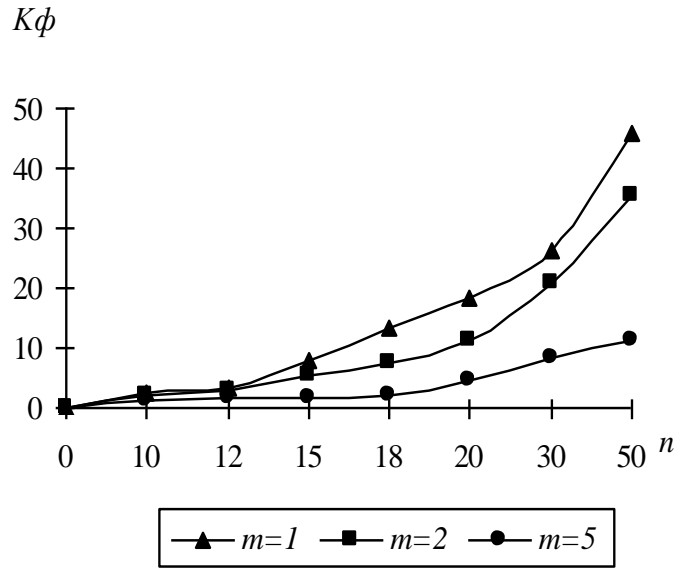


Рис. 6.39. Залежність K_{ϕ} від n для 2-ї зони

Експеримент полягав у розв'язанні 1000 тестових задач алгоритмом з заданою розмірністю вхідних параметрів n і m . Діапазон змін величин в обмеженнях підбирався так, щоб ранг оптимального розв'язку, одержуваного на останньому етапі алгоритму (він є точним), належав 2-й зоні, тобто $\bar{r} = [n/3 \div 2n/3]$.

Результати дослідження показали (рис. 6.40, 6.41), істотний виграв алгоритмів A_9 і A_{10} у порівнянні з алгоритмом A_7 . Тому використання значень отриманого припустимого розв'язку на попередніх етапах дозволяє зменшити число оброблюваних векторів алгоритмом. Аналогічні результати отримані й для алгоритмів A_{11} і A_{12} .

Швидкість розв'язання на останньому етапі для алгоритмів A_{11} і A_{12} залежить від максимального числа векторів, оброблюваних на одному ранзі \bar{r} .

Точно такий же експеримент був проведений для алгоритмів A_{11} і A_{12} . Тільки тепер змінювався вхідний параметр m , що задає кількість обмежень. Результати цього експерименту для найгіршого випадку (2-ї зони) наведені на рис. 6.42 та 6.43 відповідно.

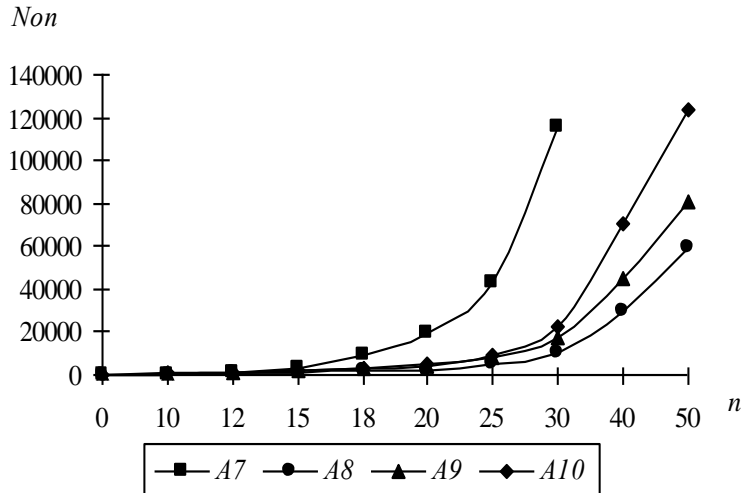


Рис. 6.40. Залежність N_{on} від n для 2-ї зони
 $V, Kб$

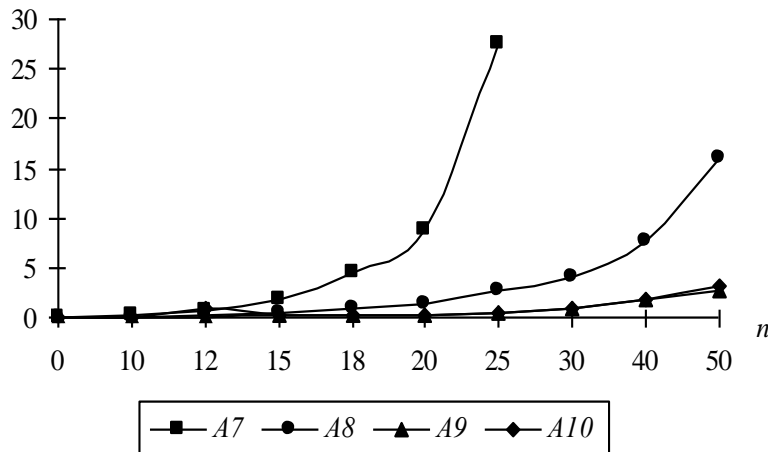


Рис. 6.41. Залежність V від n для 2-ї

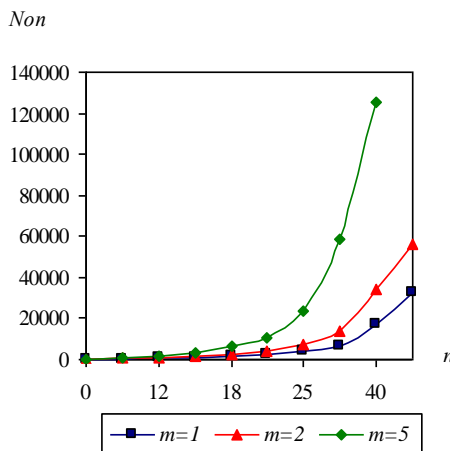


Рис. 6.42. Залежність N_{on} від n та t для алгоритму A_{11}

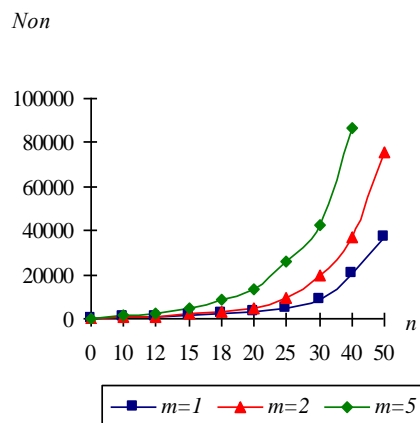


Рис. 6.43. Залежність N_{on} від n та t для алгоритму A_{12}

Таким чином, у підрозділі експериментально досліджені властивість розроблених алгоритмів і дана оцінка показників ефективності для них. З метою визначення практичної значущості алгоритмів $A_1 - A_{12}$ варто зробити порівняльний аналіз за обраними показниками з відомими алгоритмами.

6.6.3. Порівняльний аналіз рангових алгоритмів з відомими алгоритмами

Порівняльний аналіз зробимо у два етапи, що відповідають порівнянню одновимірних і m -вимірних алгоритмів з відомими.

Етап 1. Розробленими одновимірними алгоритмами в роботі є алгоритми A_1, A_2, A_3 , що дають наближений розв'язок, і A_6, A_9 , що дають точний розв'язок. З відомих точних одновимірних алгоритмів виберемо адитивний алгоритм Балаша [23], що має експонентну складність, і алгоритм динамічного програмування (ДП) [87], що має тимчасову складність $O(cn^2)$, де c – абсолютне значення цільової функції. Як показав експеримент (рис. 6.44), алгоритм ДП також залежить від \bar{r} .

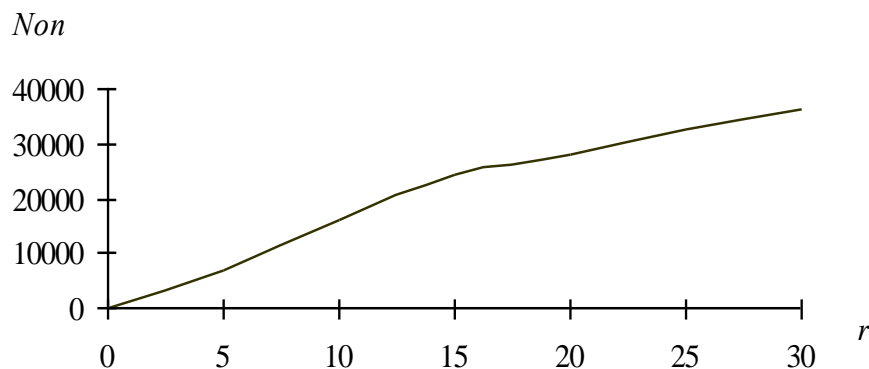


Рис. 6.44. Залежність N_{on} от \bar{r} для алгоритму ДП

Таким чином, можна зробити висновок: тимчасова складність точних алгоритмів для розв'язання задачі ЦЛП з БЗ залежить від приналежності оптимального розв'язання до якої-небудь зони. Звідси випливає то, що для точних алгоритмів розв'язання задач ЦЛП з БЗ існує хоча б одна зона, у якій спостерігається експонентний ріст виконуваних ЕО, що залежать

від кількості оброблюваних припустимих розв'язків. Приналежність до виділеної зони визначається кількістю одиниць в оптимальному розв'язку.

У випадку порівняння розроблених точних одновимірних алгоритмів A_6 , A_9 і A_{10} з алгоритмом Балаша й алгоритмом динамічного програмування можна зробити нижченаведені висновки.

При потраплянні оптимального розв'язку в 1-шу зону для задач невеликої розмірності ($n < 30$) з погляду часу розв'язання найбільш ефективними є алгоритми Балаша й ДП (для цього алгоритму додається умова малого абсолютного значення функціонала). У випадку, коли $n > 30$ і оптимальний розв'язок потрапляє в 1-шу зону, ефективніше застосовувати багато етапний алгоритм A_{10} . Якщо оптимальний розв'язок належить 2-й і 3-й зонам, то переважніше багато етапні алгоритми A_9 і A_{10} , причому за тимчасовими показниками спостерігається експонентний ріст виграшу цих алгоритмів над алгоритмом Балаша залежно від розмірності задачі n . Застосування алгоритму A_6 неефективне, тому що без початкового припустимого розв'язку йому доводиться перебирати експонентне число векторів, що лежать усередині коридора. Використання стратегій L_6 , L_8 , L_9 приводить до виділення коридора всередині множин, а таке виділення менш ефективне.

Отже, для точного розв'язання одновимірних задач малої розмірності й оптимального розв'язку, що лежить усередині 1-ї зони, виявляється переважніше алгоритм Балаша. В інших випадках доцільніше застосовувати алгоритми A_9 і A_{10} .

Якщо необхідно швидко знайти припустиме розв'язання, то за поліноміальний час алгоритмом A_1 буде отриманий такий розв'язок з похибкою близько 5 %, а алгоритмом A_3 – з похибкою близько 2 %, на що буде потрібно в 2 рази більше часу.

Етап 2. Розглянемо алгоритми другої групи. З числа розроблених алгоритмів до них належать (A_4 , A_5), і точні (A_7 , A_8 , A_{11} , A_{12}). Як видно з рис.6.39, застосування стратегії виділення m -вимірною коридора на всьому ярусі переважніше, ніж його виділення в множинах. Однак застосовувана стратегія відсікань L_{13} в алгоритмі A_8 використовується й у багато етапних алгоритмах A_{11} і A_{12} для одержання точного розв'язку. Тому для

порівняння з відомими алгоритмами розв'язку задачі ЦЛП з БЗ будуть обрані алгоритми A_{11} і A_{12} .

Почнемо порівняльний аналіз з тих методів і алгоритмів, які дають точний розв'язок задачі (6.7) - (6.9). До таких методів належать адитивний алгоритм Балаша [24], його модифікація для задач з позитивними коефіцієнтами [26], метод W послідовного аналізу й відсівання варіантів [34] і P -метод.

У таблиці 6.1 наводяться тимчасові значення розв'язання задачі (6.7) - (6.9) зазначеними методами і їхній виграш стосовно алгоритму Балаша.

За даними табл. 6.1 можна зробити такий висновок, що W -метод поступається перед алгоритмом Балаша для малих й середніх розмірності і більш ефективний для задач великої ($n > 100$) розмірності. У той же час для малих й середніх розмірностей більш ефективний P -метод, а при збільшенні розмірності він втрачає свою перевагу. Отже, якщо показати виграш розроблених алгоритмів A_{11} і A_{12} стосовно алгоритму Балаша й указати значення цього виграшу, то за аналогією можна зробити порівняння й з алгоритмами W -методу та P -методу.

Таблиця 6.1

Тимчасові оцінки відомих алгоритмів

n	m	Загальний час рахунку, с			Відношення до алгоритму Балаша	
		алгоритм Балаша	алгоритм методу W	алгоритм P -методу	алгоритм методу W	алгоритм P -методу
10	20	38	100	12	2.63	0.31
10	50	105	202	23	1.92	0.21
10	100	241	294	23	1.21	0.09
50	10	42	42	17	1	0.4
50	50	190	448	182	2.35	0.95
100	10	62	55	32	0.88	0.52

Найбільший виграш K_e в алгоритмі A_{12} над алгоритмом Балаша спостерігається в 2-й зоні, (рис. 6.28 та рис. 6.43), коли число оброблюваних векторів максимальне. У 1-й зоні час знаходження оптимального розв'язання порівняно з часом

розв'язання для алгоритму Балаша. Якщо оптимальне розв'язання належить 3-й зоні, то саме тут виявляються переваги алгоритму Балаша ($K_e < 1$).

Таким чином, можна зробити нижченаведені висновки.

1. Для задач малої розмірності й потраплянні точного розв'язку в 1-шу зону ефективніше використовувати P -метод.

2. Для задач великої розмірності й потраплянні в 1-шу зону – алгоритми A_{11} і A_{12} . При потраплянні оптимального розв'язку в 2-гу зону найбільший тимчасовий виграш, при незначних витратах пам'яті, досягають алгоритми A_{11} і A_{12} .

3. Для задач малої розмірності ($n < 50$) і потраплянні оптимального розв'язку в 3-ю зону переважний алгоритм Балаша, що з ростом n втрачає його. Тому для задач великої розмірності й розв'язків, що належать 3-й зоні, доцільніше застосування алгоритмів, заснованих на ранговому підході.

6.6.4. Аналіз впливу сортувань коефіцієнтів при функціоналі й обмеженнях на величину похибки розв'язання

Покажемо, що алгоритмам A_1, A_2, A_3 характерна властивість стабілізації похибки зі збільшенням розмірності задачі (рис.6.38). Становить інтерес також експериментальне дослідження впливу на похибку наближених алгоритмів A_1, A_2, A_3 різних видів сортувань коефіцієнтів при функціоналі й обмеженні (6.9):

$$a_1 \leq a_2 \leq a_3 \leq \dots \leq a_n ; \quad (6.69)$$

$$c_1 \geq c_2 \geq c_3 \geq \dots \geq c_n ; \quad (6.70)$$

$$\frac{c_{11}}{a_{11}} \geq \frac{c_{12}}{a_{12}} \geq \frac{c_{13}}{a_{13}} \geq \dots \geq \frac{c_{1n}}{a_{1n}} . \quad (6.71)$$

Для одержання точного розв'язку як еталон використовуємо алгоритм A_7 . Усі результати отримані з довірчою ймовірністю 0,95. Коефіцієнти у функціоналі й обмеженнях генеровані за рівномірним законом розподілу в діапазоні (1-100). Апроксимуючі криві результатів обчислень похибок і кількості неточних розв'язків для алгоритмів A_1, A_2, A_3 наведені на рис. 6.45 – 6.47 відповідно.

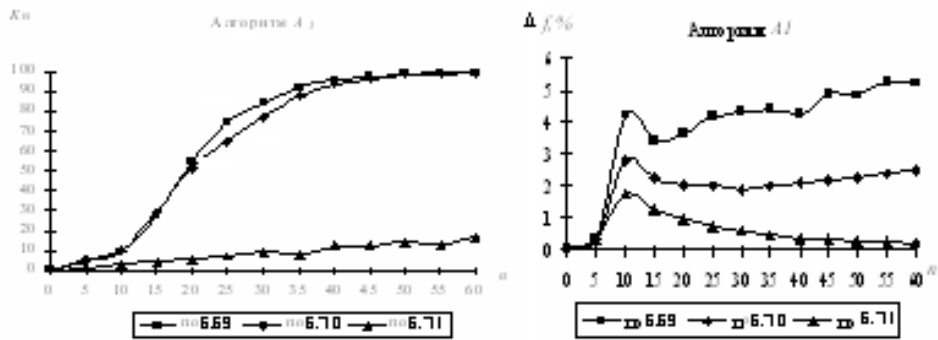


Рис. 6.45. Залежність K_n та Δf для алгоритму A_1

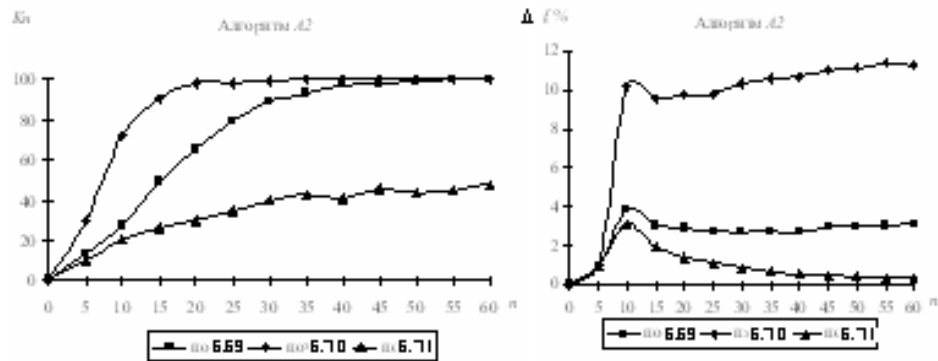


Рис. 6.46. Залежність K_n та Δf для алгоритму A_2

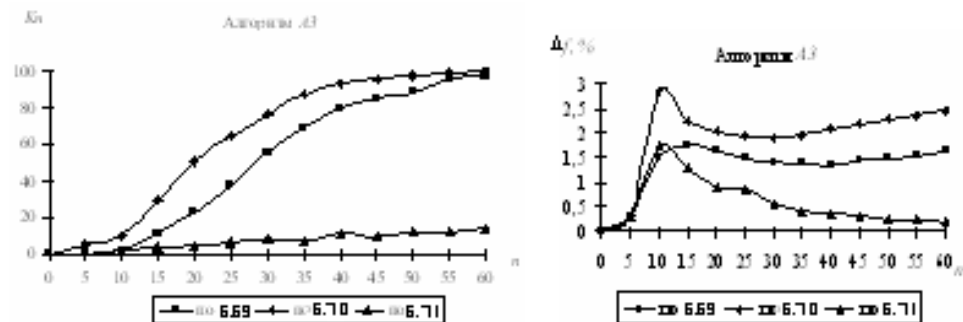


Рис. 6.47 Залежність K_n та Δf для алгоритму A_3

Апроксимація побудована за методом найменших квадратів.

З графіків видно, що найкращою точністю характеризується алгоритм A_3 , похибка якого при $n > 40$ не перевищує 0,5 %. Найбільш ефективним сортуванням є сортування в порядку спадання відношень коефіцієнтів при функціоналі до відповідних коефіцієнтів при обмеженнях.

Таким чином, наведені алгоритми мають тимчасову складність $O(n^3)$ і похибка не більше 10 %. Отже, алгоритми на основі рангового підходу мають похибку в 5-10 разів менше відомих алгоритмів. У випадку реалізації алгоритмів на

багатопроесорній обчислювальній структурі тимчасову складність алгоритмів можна зменшити в n раз, тобто до $O(n^2)$.

З аналізу графіків і таблиць можна зробити висновок, що найкращою точністю характеризується алгоритм MAX-MIN, похибка якого при $n > 50$ не перевищує 0,5 %. Найбільш ефективним сортуванням є сортування в порядку спадання відношень коефіцієнтів при функціоналі до відповідних коефіцієнтів в обмеженнях. За своїми складними показниками двох- і трьохетапні алгоритми майже еквівалентні. Останнє пояснюється тим, що більш точне попереднє розв'язання на першому етапі зменшує число аналізованих розв'язань на етапі роботи точного алгоритму. При дослідженні ЗНП і ЗНР як еталонний був узятий алгоритм методу галузей і границь (МГГ), що використовує стратегію пошуку в глибину. Вибір цього методу пов'язаний з тим, що він універсальний, простий у реалізації, не вимогливий до машинних ресурсів і вважається найбільш швидкодіючим з існуючих методів. Що ж стосується стратегії пошуку, то вона була обрана з таких міркувань. Уважається, що використання стратегії пошуку в глибину дозволяє розв'язувати задачу з меншою витратою ресурсів ЕОМ, але час пошуку при цьому збільшується в порівнянні з пошуком у шир. І на малих розмірностях це дійсно так. Але зі збільшенням розмірності задачі обсяг проміжних даних зростає настільки, що оперативної пам'яті (ОЗУ) для їхнього зберігання виявляється недостатньо. Тому ці дані доводиться зберігати на зовнішніх запам'ятовувальних пристроях (ВЗУ), що істотно збільшує час розв'язання задачі, тому що операція читання й запису на ВЗУ істотно більше, ніж в ОЗУ. Крім того, різні модифікації МГГ, що визначають до початку роботи основного алгоритму верхню оцінку розв'язання, дозволили практично вирівняти кількість аналізованих варіантів для обох стратегій. Оскільки ж нас у першу чергу цікавили задачі великої розмірності, то як еталонний був обраний МГГ до стратегії пошуку в глибину. Використання МГГ дозволяє зменшити кількість аналізованих варіантів, але проблема полягає в тому, що, навіть одержавши розв'язання, алгоритм продовжує перевіряти ті гілки дерева шляхів, вага (ранг) яких менше отриманого розв'язку. Таким чином, навіть якщо припустити, що верхня оцінка розв'язку збіглася з

оптимальним значенням R^* , кількість векторів (вузлів розгалуження),

$$\sum_{r=1}^{R^*} C_n^r,$$

побудованих МГГ, буде дорівнює сумі R^* елементів ряду трикутника Паскаля.

Алгоритми ж, засновані на ранговому підході, за рахунок попереднього калібрування й прогнозування кінцевого результату дозволяють на кожному ранзі, починаючи з першого, відсікати ті шляхи, які алгоритм визначає як безперспективні.

На рис. 6.48 показана залежність кількості оброблених векторів від рангу розв'язання задачі розмірності 200 для описаних алгоритмів. Для наочності на цьому ж графіку наведена крива «трикутника Паскаля».

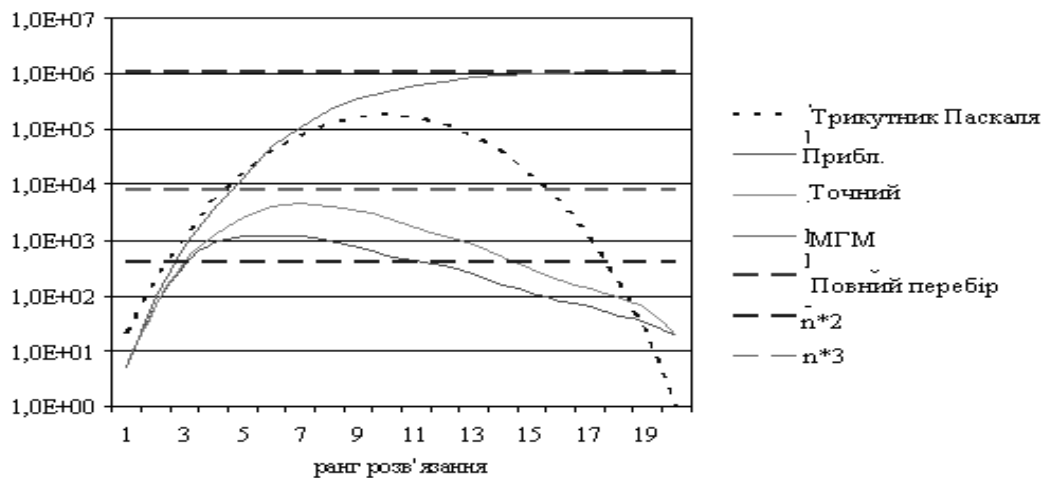


Рис.6.48. Залежність кількості оброблених векторів від рангу розв'язання задачі

Під безперспективними шляхами розуміються шляхи, які в принципі не можуть дати покриття, а також ті, які дадуть покриття на ранзі, що перевищує прогнозований оптимальний розв'язок. Інакше кажучи, на відміну від МГГ, відсікаються не тільки ті гілки, довжина яких більше оптимального значення, але й ті, які не містять розв'язання в принципі.

Криві на графіках підтверджують, що складність розглянутих наближених алгоритмів не може перевершувати $O(n^3m)$, оскільки всього ярусів n , на ярусі число формованих векторів не перевершує n^2 , і число елементів, аналізованих у кожному векторі, не перевершує m . Для точних алгоритмів отримана оцінка тимчасової складності в середньому також відповідає $O(n^3m)$. Що ж стосується МГГ, то зі збільшенням рангу розв'язання МГГ наближається до повно переборного алгоритму. Справедливості заради потрібно відзначити, що кількість елементарних операцій, необхідних для побудови одного вектори, у МГГ менше, ніж в алгоритмах, заснованих на ранговому підході, це компенсується зменшенням числа оброблюваних векторів. Графіки залежностей числа елементарних операцій і часу розв'язання від рангу шляху представлені на рис. 6.49 і 6.50.

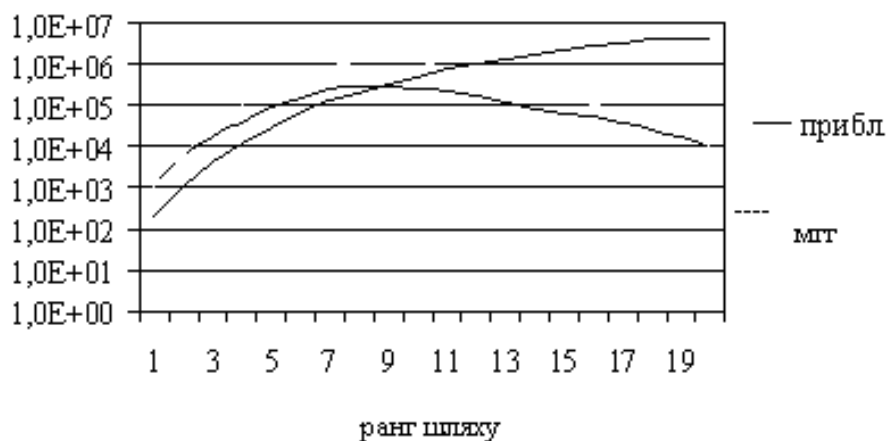


Рис. 6.49. Залежність кількості елементарних операцій від рангу шляху

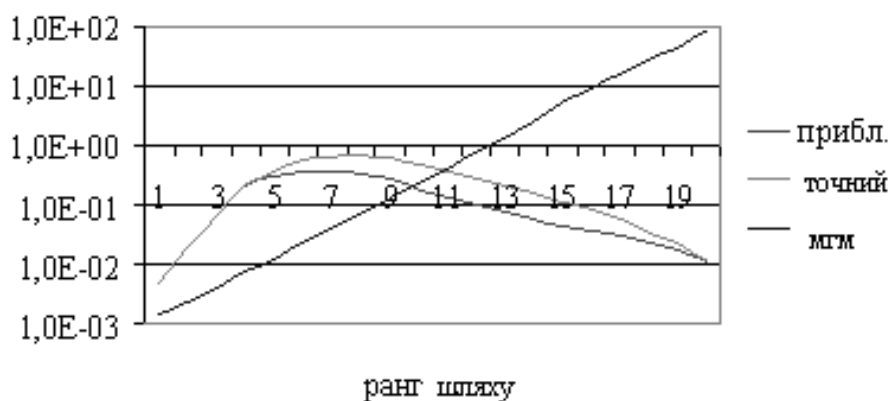


Рис. 6.50. Залежність часу розв'язання від рангу шляху

Що ж стосується залежності похибки наближеного алгоритму від рангу шляху, то вона виникає при потраплянні розв'язку в другу умовно виділену зону й коливається в межах від 0 до 15 % (рис. 6.51), причому кількість неточних розв'язків не перевершує 10 % від загального числа дослідів.

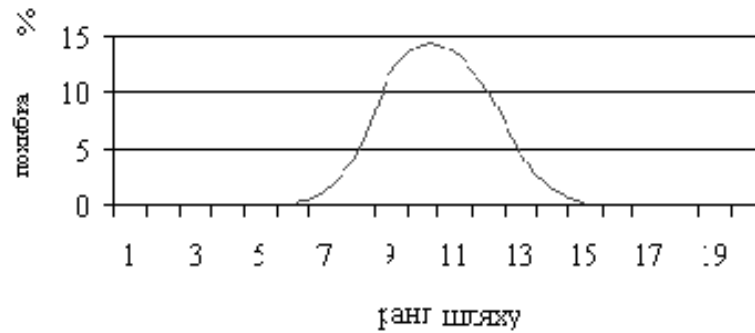


Рис. 6.51. Залежність похибки від рангу шляху

Для оцінки залежності числа елементарних операцій, кількості оброблених векторів і часу розв'язання від розмірності задачі вихідні дані за рахунок щільності заповнення одиницями генерувалися таким чином, щоб розв'язання потрапляло у 1-шу й у 2-гу умовно виділені зони, тому що для 3-ї зони переваги розроблених у роботі алгоритмів очевидні. З рис.6.52 - 6.54 впливає, що зі збільшенням розмірності задачі підхід до розв'язання ЗНП на основі методу галузей і границь з погляду тимчасової і алгоритмічної складності істотно поступається методам, заснованим на ранговому підході.

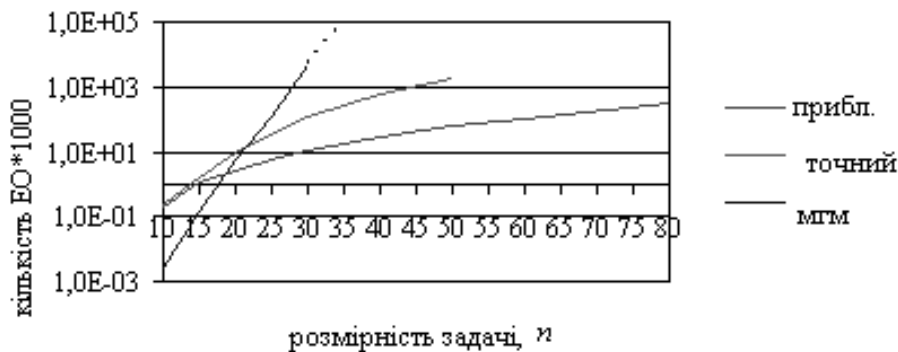


Рис. 6.52 Залежність кількості ЕО від розмірності задачі

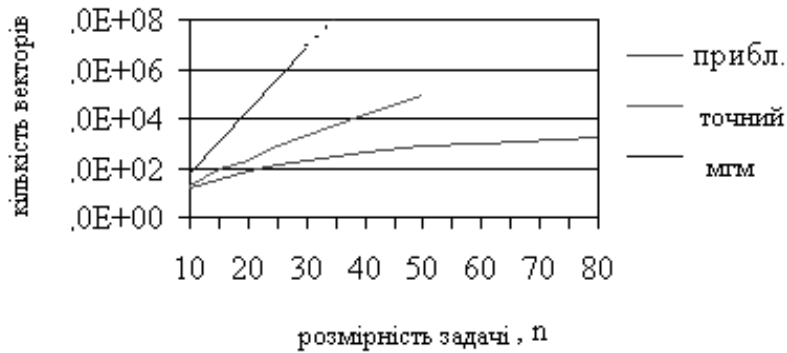


Рис. 6.53. Залежність числа оброблених векторів від розмірності задачі

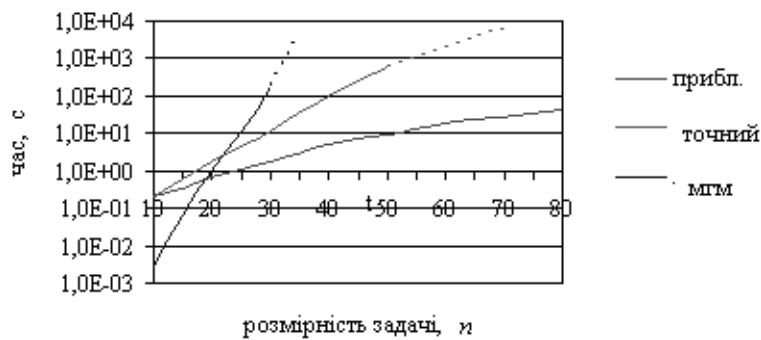


Рис.6.54. Залежність часу розв'язання від розмірності задачі

Похибка наближених алгоритмів (рис. 6.55 - 6.57) стабілізується зі збільшенням розмірності розв'язуваної задачі. Так, при $n > 30$ вона не перевищує 6 %, а починаючи з $n = 50$ похибка не перевищує 3 %, тобто зі збільшенням n спостерігається збіжність наближеного розв'язання до оптимального, однак кількість неточних розв'язків зросла до 30 %.

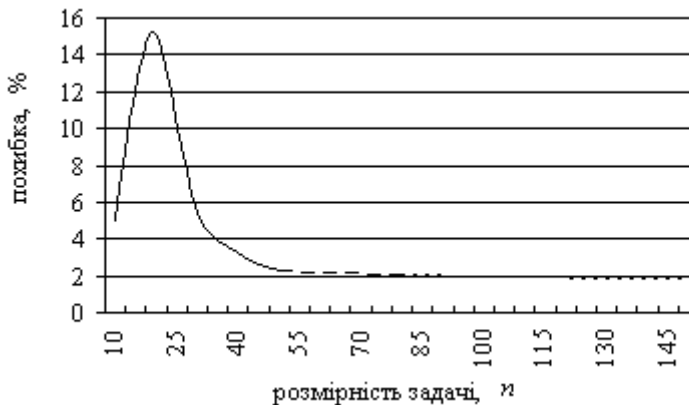


Рис. 6.55. Залежність похибки наближених алгоритмів від розмірності задачі

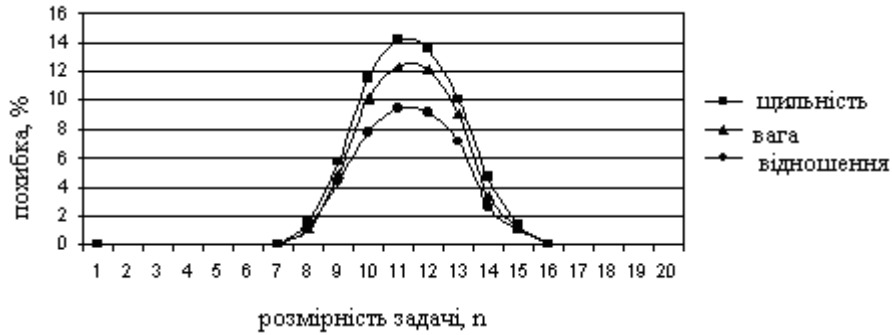


Рис. 6.56. Залежність похибки розв'язку ЗНП від рангу шляху для рзних сортувань

З рис. 6.56 та 6.57 видно, що використання сортувань векторів з множини B у порядку спадання відношень ваги стовпця до числа одиниць у стовпці дозволять в 1,4 рази зменшити похибку розв'язку, яка починаючи з $n=45$ не перевищує 2 %, причому кількість неточних розв'язків не перевершує 10 % від загального числа дослідів. На рис. 6.58 та 6.59 наведені аналогічні характеристики й для ЗНР.

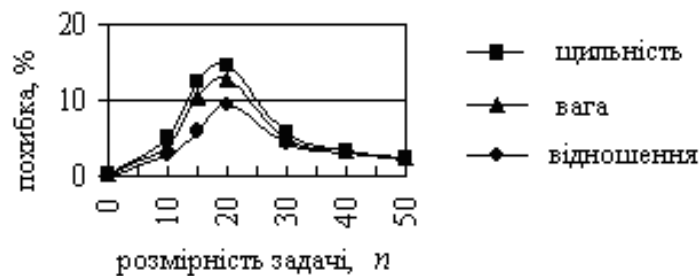


Рис. 6.57. Залежність похибки розв'язання ЗНП від розмірності задачі для рзних сортувань

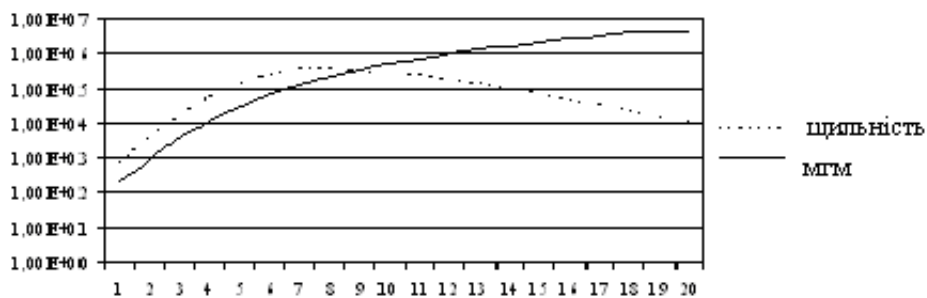


Рис. 6.58. Залежність кількості елементарних операцій від рангу шляху для ЗНР

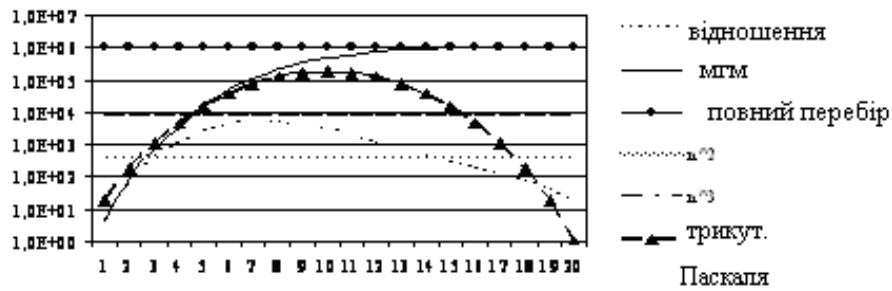


Рис. 6.59. Залежність кількості оброблених векторів від рангу шляху для ЗНР

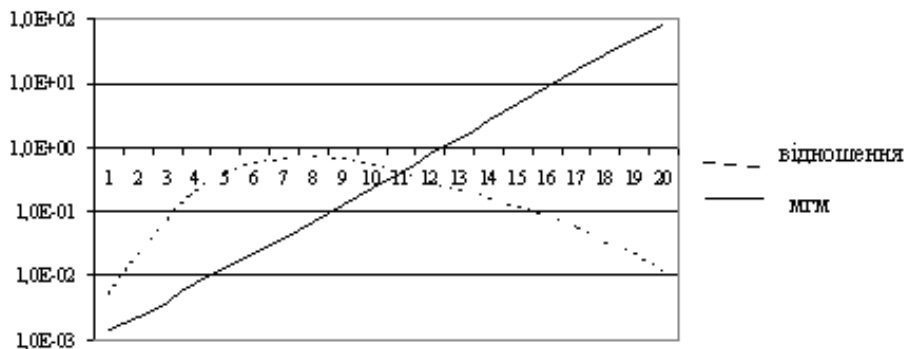


Рис. 6.60. Залежність часу розв'язання від рангу шляху для ЗНР

Дослідження залежності числа операцій додавання й порівняння від числа змінних у рівнянні при розв'язанні діофантових рівнянь алгоритмом виділення коридора, що реалізує ідею, показало, що при $n = 150$ тимчасова складність алгоритму не перевищувала $O(n^5)$, а при $n = 35 - O(5999n^4 + 49)$. Залежність МО й СКО числа елементарних операцій від n наведені на рис. 6.64.

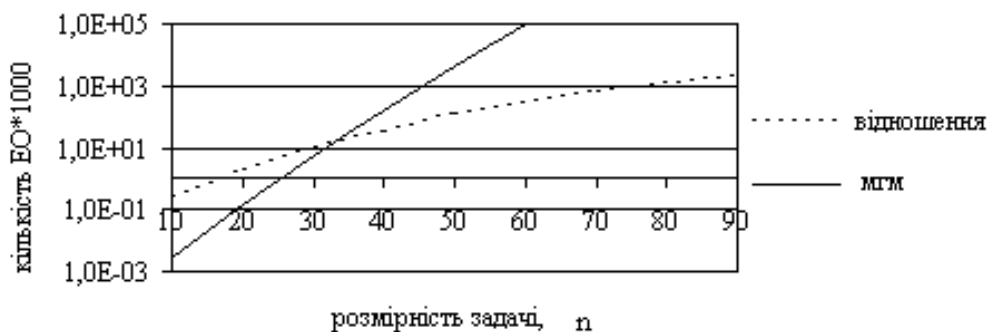


Рис. 6.61. Залежність кількості ЕО від розмірності задачі ЗНР



Рис. 6.62. Залежність числа оброблених векторів від розмірності задачі для ЗНР

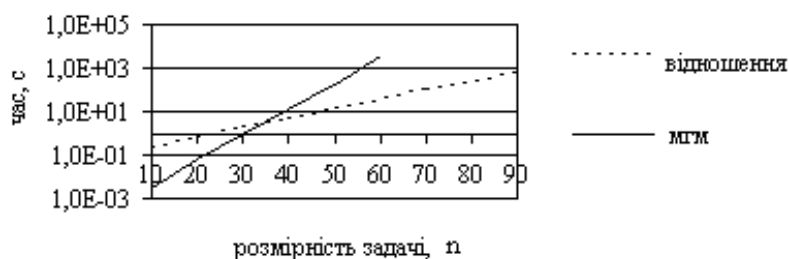


Рис. 6.63. Залежність часу розв'язання від розмірності ЗНР

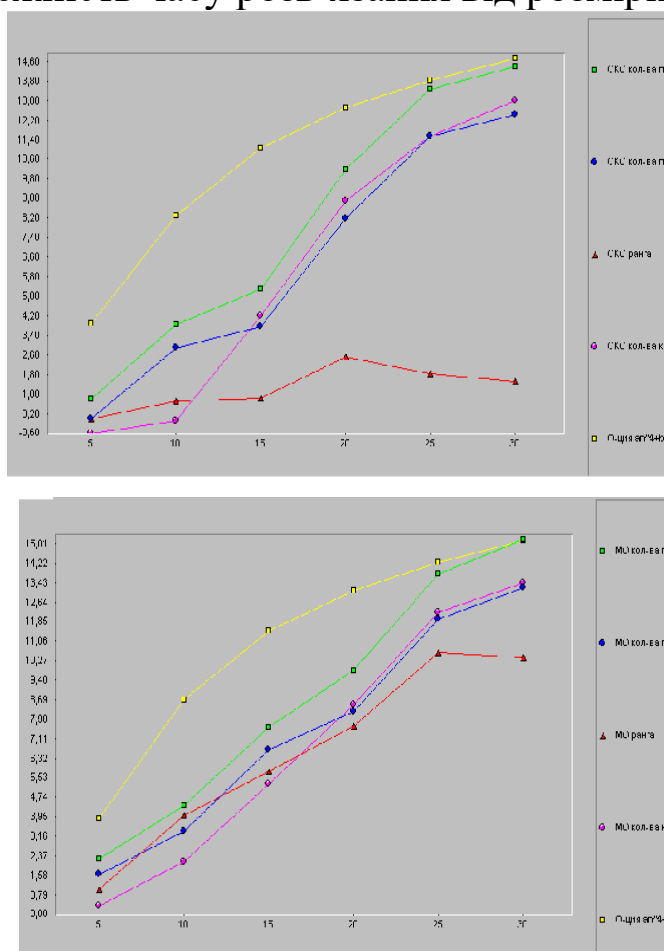


Рис. 6.64. Залежність МО й СКО числа елементарних операцій при розв'язанні діофантових рівнянь з БЗ від числа змінних у рівнянні

6.6.5. Гарантовані прогнози і їх вплив на тимчасову складність і похибка рангових алгоритмів розв'язку задачі 0, 1-рюкзак

В основі математичної моделі рангового підходу для побудови алгоритмів розв'язання задач ЦЛП з БП покладений принцип оптимізації за напрямком у дискретному просторі станів, заданому графом G_{Δ} [106]. Основна стратегія відсікання безперспективних варіантів розв'язань, використовувана в алгоритмах, розроблених в [106], заснована на перевірці нерівності

$$d_c(\mu_{sj}^r) + \gamma_p < \max_{\{c_j\}} \{d_c(\mu_{sp}^{*r})\},$$

де $d_c(\mu_{sp}^r)$ – довжина шляху μ_{sp}^r до вершини p рангу r за вагами $\{c_j\}$, що дозволяє виключити цей шлях з подальшого аналізу як безперспективний, якщо умова виконана.

$\gamma_p = c_{p+1} + c_{p+2} + \dots + c_n$, $\gamma_n = 0$; $p = (\overline{1, n-1})$ – верхня оцінка (негарантований прогноз) приросту величини значення локального екстремуму в області Ω_p на всіх наступних рангах.

У даній роботі показано, що заміна використовуваного в алгоритмах [40] негарантованого оптимістичного прогнозу збільшення довжини шляху в графі G_{Δ} на гарантований песимістичний дозволяє істотно знизити похибку наближених алгоритмів і тимчасову складність точних алгоритмів, а також досліджено вплив різних сортувань коефіцієнтів у функціоналі й обмеженнях на похибку наближених алгоритмів.

Формалізація й розв'язання задачі. Для формування гарантованого прогнозу введемо матрицю Π з n -стовпцями й p -рядками ($p = m+1$) і вектор стовпець b_j правих частин обмежень (6.8)

$$\Pi = \begin{array}{c} 1 \\ 2 \\ 3 \\ \dots \\ p \end{array} \left| \begin{array}{cccc} C_1 & C_2 & \dots & C_n \\ a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{array} \right| ; \left| \begin{array}{c} b_j \\ b_1 \\ b_2 \\ \dots \\ b_m \end{array} \right|, \quad (6.72)$$

де C_i – коефіцієнти при змінних x_i у функціоналі (6.7);
 a_{ij} – коефіцієнти при змінних x_i в j -му обмеженні.

На основі матриці (6.72) побудуємо процедуру A , на основі якої будемо визначати гарантований прогноз.

Процедура A

КРОК 1. Змінній Z присвоїмо значення, рівне 2, в рядку $h=Z$ і відсортуємо стовпці матриці Π у порядку зростання елементів a_{ij} у рядку h . Формуємо порожні множини R і S .

КРОК 2. Здійснюємо підсумовування елементів стовпців, починаючи з перших двох стовпців у відсортованій матриці Π , і перевіряємо для кожного рядка виконання нерівності $\sum_{i=1}^n a_{ij} \leq b_j, j = (\overline{1, m})$. Підсумовування стовпців припиняємо при досягненні n значення $r-1$, починаючи з якого для одного з рядків j нерівність $\sum_{i=1}^{n=r-1} a_{ij} \leq b_j$ перестане виконуватися, й значення r заносимо на вільну позицію в множину R (r – число додавань стовпців до моменту, коли в одному з рядків нерівність $\sum_{i=1}^n a_{ij} \leq b_j, j = (\overline{1, m})$ не виконується), а суму відповідних c_i , накопичену по першому рядку, розміщуємо на вільну позицію в множині S .

КРОК 3. Перевіряємо $Z=p$, якщо так, то переходимо до виконання наступного кроку, інакше присвоюємо $Z = Z + 1$ і переходимо до виконання кроку 2.

КРОК 4. У множині R знаходимо максимальний елемент r_{\max} , розташований на k -й позиції, й суму s_k , розташовану на k -й позиції в множині S .

Отримана пара значень (r_{max}, s_k) є гарантованим прогнозом того, на скільки рангів у графі G_Δ можна продовжити шлях не порушуючи обмежень (максимальне число вершин, яке можна ще приєднати до шляху не порушуючи обмежень), при цьому поточна сума s_k на k -й позиції множини S дає нижню гарантовану оцінку довжини шляху за вагами функціонала. У процедурі A на кроці 4 можливий також вибір максимальної суми s_k^* і відповідного їй значення r_k , при цьому пари (s_k^*, r_k) є тож гарантованим прогнозом того, що до шляху можна приєднати r_k вершин не порушуючи обмежень і при цьому максимальна нижня гарантована оцінка довжини шляху за вагами функціонала дорівнює s_k^* . Причому, якщо в графі G_Δ процедура A застосовується для конкретної вершини i , то стовпці з номерами, меншими ніж i , виключаються з матриці Π , оскільки, як видно з графа G_Δ , вершини з номерами, меншими i , уже не можуть зустрічатися в шляху. У випадку одновимірної задачі матриця Π буде містити тільки два рядки. В алгоритмах розв'язання задачі (6.7-6.9), наведених в [32], для підвищення ефективності відсікання безперспективних варіантів розв'язань уведено калібровані вектори, за якими здійснювалася оцінка того, скільки шлях може набрати за функціоналом, при цьому вони тож ґрунтуються на негарантованому оптимістичному прогнозі γ_p для довільної вершини p . Застосування процедури A дозволяє для кожної вершини графа G_Δ створити калібровані вектори шляхом сортування стовпців матриці Π у порядку зростання a_{ij} для одновимірних задач і в порядку спадання c_j для багатовимірних задач. У роботах [32, 40] показано, що при розв'язанні одновимірних задач найбільш ефективними сортуваннями коефіцієнтів функціонала є сортування в порядку спадання ваг функціонала й у порядку спадання відношень коефіцієнтів при x_i у функціоналі до відповідних коефіцієнтів при x_i в обмеженні. Причому в останньому випадку похибка була найменшою, а у випадках застосування стратегій *max* і *max-min* не перевищувала 3-5 %. Тому оцінимо, як вплине заміна негарантованого оптимістичного прогнозу на гарантований песимістичний прогноз і зміну сортувань у каліброваних векторах на основі запропонованої процедури A на тимчасову складність і похибку

алгоритмів, розроблених у [32-36]. Порівняння наближених алгоритмів проводилося з алгоритмом, розробленим у [40], що реалізує стратегію *max*, а як точний алгоритм, для порівняння, використовувався двохетапний алгоритм, наведений у [32]. Експериментально отримана оцінка в середньому для тимчасової складності аналізованих алгоритмів і їхньої похибки. Оцінка в середньому здійснювалася шляхом розв'язання 1000 тестових задач для кожної точки графіка. При цьому ваги коефіцієнтів у функціоналі й обмеженнях генерувалися за рівномірним законом розподілу. Середні значення тимчасової складності й відносної похибки алгоритмів отримані з довірчою ймовірністю 0,95. Результати експериментального дослідження наведені на рис. 6.65 - 6.72.

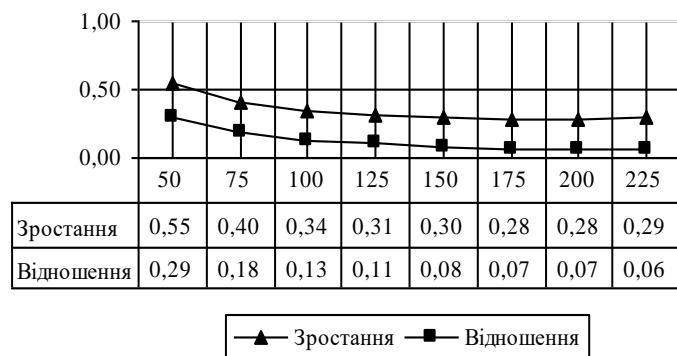


Рис. 6.65. Залежність похибки алгоритму *Max2* від початкового сортування функціонала

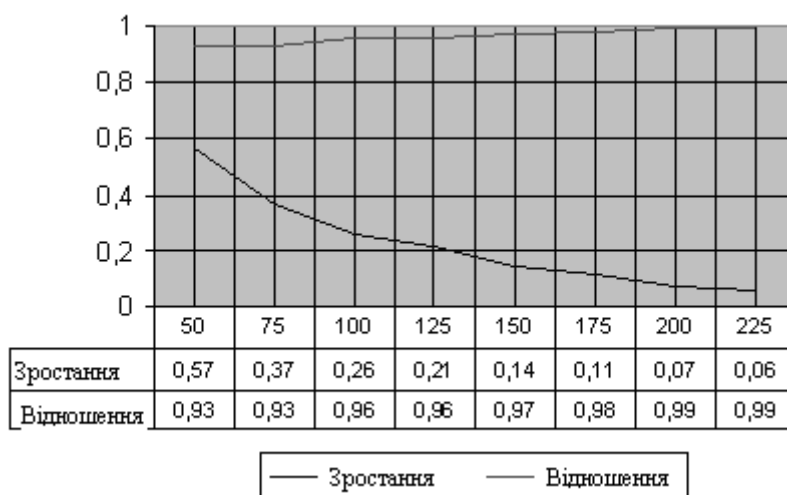


Рис. 6.66. Залежність частоти збігу розв'язання наближеного алгоритму *Max2* з точним розв'язанням від способу сортування функціонала

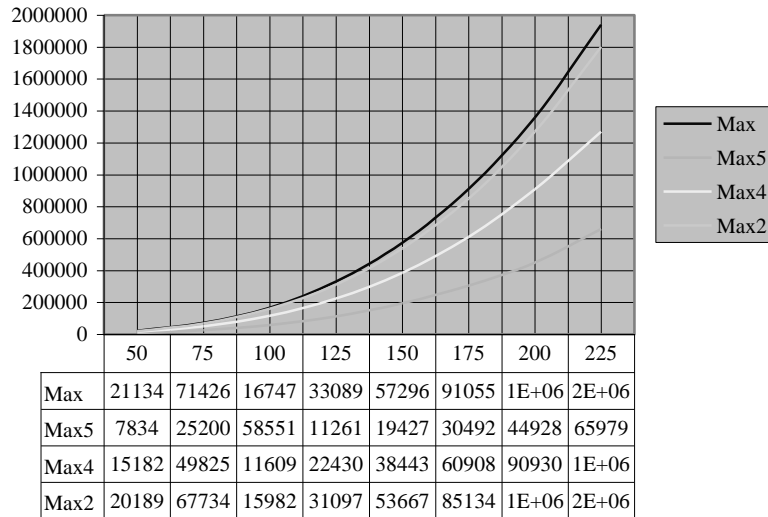


Рис. 6.67 Середнє число оброблюваних векторів для наближених алгоритмів

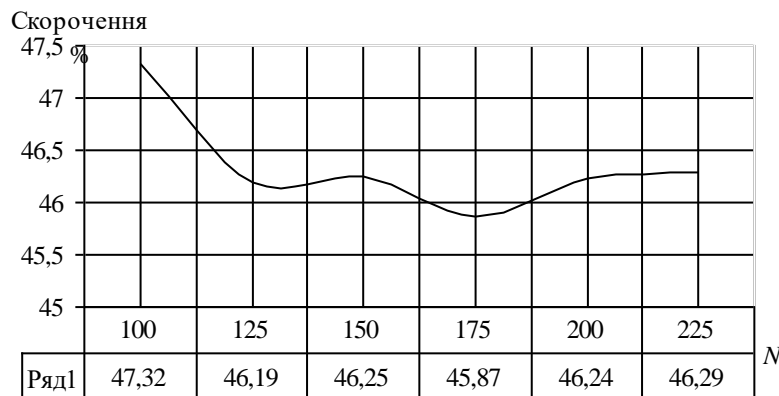


Рис. 6.68. Залежність скорочення числа аналізованих векторів при заміні негарантованого прогнозу на гарантований прогноз у точному методі

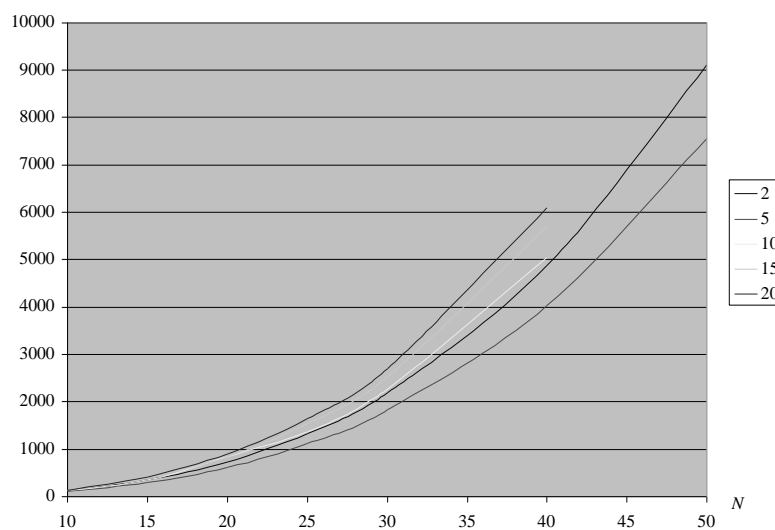


Рис. 6.69 Залежність середньої кількості аналізованих векторів від розмірності задачі при рзному числі обмежень для Max5

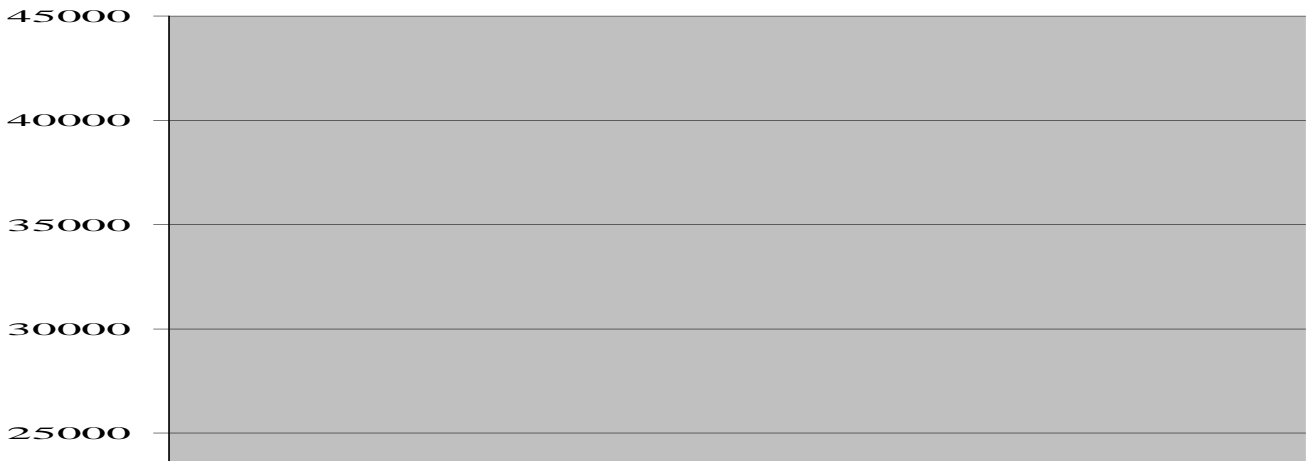


Рис. 6.70. Залежність середньої кількості аналізованих векторів від розмірності задачі при різному числі обмежень для Мах 2

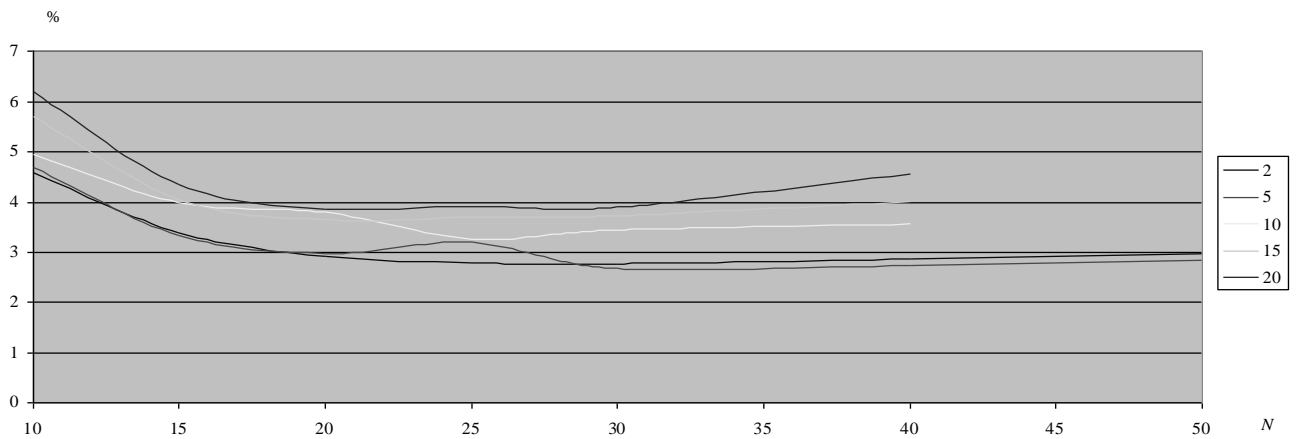


Рис. 6.71. Залежність похибки алгоритму Мах 5 від розмірності задачі при різному числі обмежень

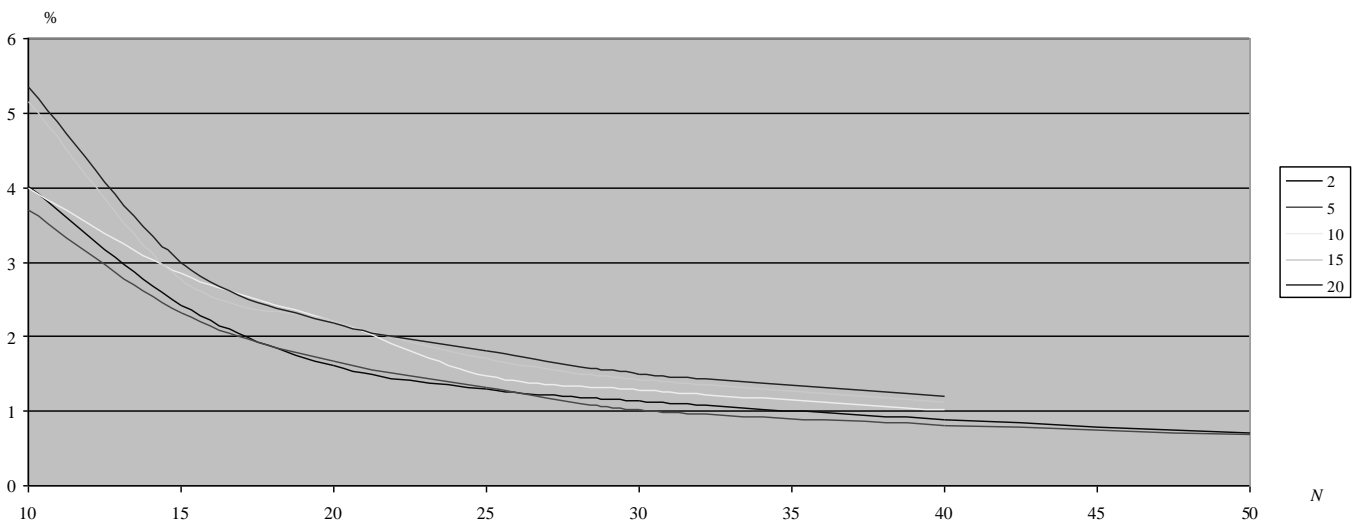


Рис. 6.72. Залежність похибки алгоритму Мах 2 від розмірності задачі при різному числі обмежень

У процесі експерименту досліджувалися такі стратегії формування шляхів у графі G_{Δ} :

- стратегія Мах 2 (при побудові векторів обмежень використовувалося сортування по спаданню відношень коефіцієнтів функціонала до відповідних коефіцієнтів відповідного обмеження, при цьому при формуванні шляхів наступного рангу на кожному кроці формувалися шляхи, що відповідали найкращому гарантованому прогнозу й шляху з кращим оптимістичним негарантованим прогнозом);

- стратегія Мах 4 (при побудові шляху з усіх вершин вибирається краща за прогнозом вершина на ярусі, і вона включається в шлях, що далі впливає з множини вершин, що залишилися, включається в шлях за тим же принципом);

- стратегія Мах 5 (відмінною рисою від *max* є то, що побудова починається не з першого рангу, а з рангу на одиницю меншого, ніж значення максимального рангу, на якому будь-який вектор побудованого шляху задовольняє обмеження).

З рисунків видно, що введення гарантованого прогнозу в алгоритми розв'язання одновимірної задачі (6.7 - 6.9) на основі процедури A дозволяє знизити похибку з 3 до 0,29 %, що зі збільшення розмірності асимптотичне прагне до 0 і вже при $n > 100$ не перевищує 0,08 %. Останнє обумовлено тим, що, як видно з рис. 6.66, зі збільшенням розмірності задачі число неточних розв'язків прагне до 0. Тимчасова складність як точних, так і наближених алгоритмів зменшується на 40 % і більше (див. рис. 6.67, 6.68). У випадку розв'язання багатомірної задачі (6.7 - 6.9) найкращою стратегією є стратегія Мах 2, у якій тимчасова складність практично не залежить від числа обмежень, а похибка асимптотичне прагне до нуля й при $n > 40$ не перевищує 1 %. Середнє значення тимчасової складності аналізованих алгоритмів з довірчою ймовірністю 0,95 не перевищувало $O(cn^3)$, де $c \leq 1$.

Вправи

6.1. Поясніть суть принципу оптимізації за напрямком в n -вимірному одиничному кубі й принцип виділення коридора при розв'язанні задачі $(0,1)$ - рюкзак.

6.2. Чому при розв'язанні задач булевого програмування на трикутному графі потрібне сортування коефіцієнтів функціонала в порядку їхнього спадання з номерів вершин трикутного графа, що є еквівалентом n -вимірного одиничного куба?

6.3. Чому при малих розмірностях оптимізаційної задачі алгоритми на основі ідей методу галузей і границь мають меншу тимчасову складність порівняно з ранговими алгоритмами, а зі збільшенням розмірності задач їхня тимчасова складність різко зростає?

6.4. Розв'яжіть задачу прикладу 6.1 з використанням стратегій \max і \min , порівняйте отриманий результат і поясніть, чому ці стратегії дають наближений розв'язок?

6.5. Як використовуючи ідеї рангового підходу можна розв'язати систему цілочисельних лінійних рівнянь з булевими змінними?

6.6. Як використовуючи рангові алгоритми для розв'язання ЗНП можна побудувати алгоритми розв'язання ЗНР?

6.7. Поясніть, чому зі збільшенням розмірності розв'язуваних задач у рангові в наближених рангових алгоритмах похибка зменшується зі збільшенням розмірності розв'язуваних задач.

7. ЗАГАЛЬНА СХЕМА РОЗВ'ЯЗАННЯ ЗАДАЧ КОМБІНАТОРНОЇ ОПТИМІЗАЦІЇ Й ЗАДАЧ НЕЛІНІЙНОГО БУЛЕВОГО ПРОГРАМУВАННЯ

7.1. Загальний підхід до розв'язання задач комбінаторної оптимізації й теорії графів

Розглянемо кінцеву множину довільних об'єктів теорії графів або конфігурацій комбінаторної оптимізації.

У загальному випадку об'єкт може бути довільним, але повинна бути визначена кінцева множина елементів $\Omega = \{\omega_l\}$ або підмножини $L_i \in \Omega$ і правила R , що дозволяють формувати об'єкти з вихідних елементів або підмножин L_i , приналежних множині Ω . Нехай задано деяке розбиття множини Ω на сімейства підмножин $\{L_i\}$ таке, що $\bigcup_i L_i = \Omega$, і L_i описують об'єкти, які цікавлять нас, що складаються з базових елементів $\{l_i\}$ таких, що $\bigcup_i l_i = \Omega$, і правило R , що дозволяє з базових елементів визначати вагові характеристики довільних об'єднань $L_k \cup L_p \in \Omega$, що характеризують властивість $\{v\}$, і потрібно визначити об'єкт з властивістю $v^* \in \{v\}$, що цікавить нас. Подамо множину всіх можливих об'єднань підмножин L_i у вигляді графа D_\square (рис. 7.1) з паралельно ярусною структурою, що складається з n горизонтальних лінійок з вершинами $1, 2, \dots, n$ і n ярусами, кожний з яких містить усі вершини графа D_\square , при цьому кожній вершині графа D_\square поставимо у відповідність базовий елемент l_i .

У графі D_\square довільна вершина i може бути досягнута шляхами рангів $r = 1, r = 2, \dots, r = n - 1$, а довільному шляху μ_{st} , що задовольняє правила побудови R і проходить через вершини (j, p, \dots, k, t) , відповідає об'єднання базових елементів $(l_j \cup l_p \cup \dots \cup l_k \cup l_t)$, визначальний деякий об'єкт $L_i \in \Omega$. Довжина цього шляху $d(\mu_{st})$ визначається за правилами, що належать множині R . Отже, множина всіх шляхів $m_{si}(r)$ у графі D_\square , що задовольняють правила R , визначає область припустимих розв'язків вихідної задачі щодо виділення об'єкта з властивістю $v^* \in \{v\}$. Як вихідну вершину в графі D_\square будемо використовувати

фіктивну вершину S , яку у деяких випадках зручно ототожнювати з нульовим або вихідним станом системи, це приводить до того, що максимальний ранг шляху в графі D_{\square} стає рівним n , а додавання вершини S до базових елементів системи не змінює їхніх властивостей, обумовлених правилами R .

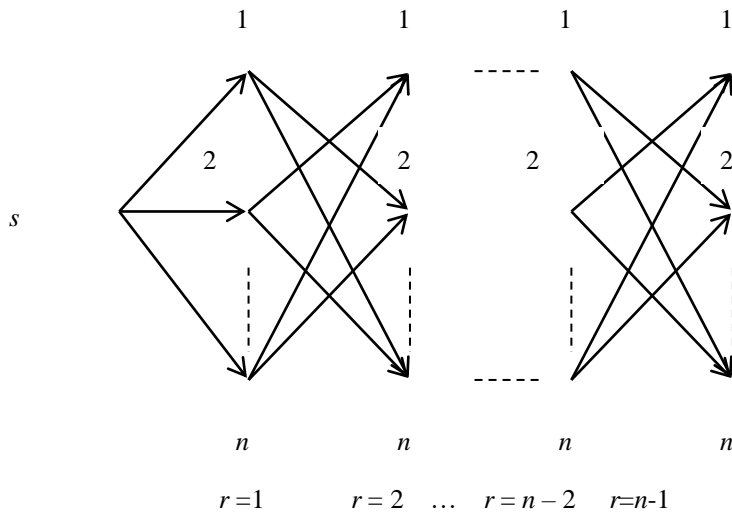


Рис. 7.1. Граф D_{\square}

Множина складних систем, що мають різні властивості $\{F_i\}$, може бути відображена за допомогою деякої підмножини графів $\{G_i\}$. Розглянемо довільний n верховий граф $G(V, E) \in \{G_i\}$, що описує стан системи $F \in \{F_i\}$ з кінцевим числом станів n . Вершини $\{i\} \in V$ графа $G(V, E)$ відповідають можливим станам системи, шляхи в графі $G(V, E)$, обумовлені послідовністю проходження вершин $\{v_i\}$ і ребер $\{(i, j)\}$, характеризують можливий порядок досягнення стану $i = p$ з деякого вихідного стану s . Важливою характеристикою шляху є ранг шляху r – число ребер, що утворює шлях. У графі $G(V, E)$ максимальне значення рангу $r = n - 1$, і в загальному випадку ранг довільного шляху μ_{sp} характеризує суму початкового стану, кінцевого стану й числа станів попередніх, через які може бути досягнутий стан p , з деякого вихідного стану s . Тоді множини шляхів m_{sj}^r ; $j = (\overline{1, n})$ визначають способи досягнення стану j . Як базові елементи $\{l_i\}$ виберемо підмножину $\{v_i\} \in V$ вершин графа $G(V, E)$, тоді

об'єктам $\{L_j\}$ буде відповідати вся множина об'єктів Ω , яких можна породити на множині V , використовуючи правила R , наприклад, кликі в довільних графах, незалежні множини, цикли графів, вектори й матриці, що задають деякий об'єкт в $G(V, E)$. Кожний об'єкт будемо характеризувати $m + 1$ ваговою характеристикою, де m – це деякі другорядні характеристики об'єкта, на які в загальному випадку можуть бути накладені обмеження на то, що вони не повинні перевищувати деяких величин $\{b_i\}$ $i = (1, 2, \dots, m)$, i є один визначальний показник якості об'єкта, побудований з вихідних елементів або їхніх підмножин об'єкт, що належать множині Ω і характеризується певною властивістю v . Правила P визначення вагових характеристик об'єктів природно повинні визначатися відповідно до правил формування самих об'єктів і $P \in R$. Таким чином, шляху μ_{sj}^r у графі D_{\square} відповідає об'єкт L_j , що може бути побудований з r базових елементів $\{v_j\}$, включаючи елемент j , на основі правил R , а множини шляхів m_{sj}^r ; $j = (\overline{1, n})$ визначають множину об'єктів L_j , які можна побудувати з r базових елементів $\{v_j\}$, включаючи елемент j . Уведемо узагальнені процедури A_0 – для формування шляхів у графі D_{\square} , що дозволяє перераховувати всі об'єкти множини $\{L_j\}$ і A'_0 – для визначення об'єктів $\{L_j\}$ з властивістю, що цікавить, при цьому розгляд почнемо з випадку, коли визначальною характеристикою об'єктів $\{L_j\}$ є одна вагова характеристика об'єкта.

Процедура A_0

КРОК 1. Формуємо в графі D_{\square} і з вершини S у множини шляхів $m_{sj}^{r=1}$ всі можливі шляхи рангу $r = 1$, що задовольняють правило R .

КРОК 2. На основі шляхів поточного рангу r формуємо всі можливі шляхи рангу $r = r + 1$ у підмножинах $m_{sj}^{r:=r+1}$, що задовольняють правило R .

КРОК 3. Перевіряємо підмножини $m_{sj}^{r:=r}$ поточного рангу r , порожні чи ні, якщо ні, то переходимо до виконання кроку 2,

інакше процедура закінчує роботу, оскільки всі об'єкти $L_j \in \Omega$ перераховані.

Процедура A_0 при переході від довільного рангу r до рангу $r + 1$ дозволяє будувати з об'єктів L_j , що містять r базових елементів v_j , об'єкти L_j , що містять $r + 1$ базовий елемент v_j .

Узагальнена процедура A'_0

КРОК 1. З вершини S будуються шляхи рангу $r = 1$, що задовольняють правила R , і відповідно до правил R визначаються їхні вагові характеристики.

КРОК 2. На основі шляхів поточного рангу r будуються всі можливі шляхи наступного рангу $r = r + 1$, що задовольняють правила R , на основі такого рекурентного співвідношення:

$$\mu_{SP}^{r+1} = \min_{d_j(m_{sj}^r \cup (j,p))} (\max)\{m_{sj}^r \cup (j,p)\}; j = (\overline{1,n}); p = (\overline{1,n}); j \neq p,$$

де вагові характеристики d_l об'єктів визначаються відповідно до правил $P \in R$.

КРОК 3. Перевіряємо підмножини m_{sj}^r поточного рангу r , порожні чи ні, якщо ні, то переходимо до виконання кроку 2, інакше процедура закінчує роботу, і з отриманих локальних екстремумів на всіх рангах вибирається глобальний екстремум.

Залежно від властивостей, що цікавлять, об'єктів і правил R може виникнути необхідність у роботі A'_0 процедури до досягнення деякого конкретного значення рангу $r = k$, наприклад, якщо об'єктом пошуку є кліка максимальної ваги, потужність якої не перевищує k . Слід також зазначити, що в загальному випадку в підмножинах $m_{sj}^r; j = (\overline{1,n})$ можлива поява однакових шляхів. Дублюючі шляхи в цих підмножинах можна видаляти й, крім того, видаляти їх на ярусах.

З'ясуємо, у яких випадках процедура дозволяє одержувати точний або наближений розв'язок задачі визначення об'єкта L_j з необхідною властивістю v , пов'язаною з мінімізацією або максимізацією вагової характеристики об'єкта L_j . Для цього

більш докладно проаналізуємо процес переходу від вихідного графа $G(V, E)$ з множиною вершин, що відповідають базовим елементам, з яких за правилами R у графі $G(V, E)$ можуть бути побудовані об'єкти, що нас цікавлять. Особливістю відображення графа $G(V, E)$ в D_{\square} є то, що якщо аналізується конкретний об'єкт L_j в $G(V, E)$, то йому в графі D_{\square} відповідає деякий шлях μ_{sj}^r , побудований на тій же множині вершин, що й об'єкт в $G(V, E)$, і вагова характеристика об'єкта L_j в $G(V, E)$ використовується для оцінки довжини шляху μ_{sj}^r в D_{\square} , але при цьому множини шляхів у графах D_{\square} і $G(V, E)$ збігаються. Припустимо, що об'єктами в $G(V, E)$ є самі шляхи графа $G(V, E)$, тоді відображення графа $G(V, E)$ є відображенням на самого себе й граф D_{\square} , у цьому випадку, як показано в роботах [9, 27], являє собою стягнуто дерево всіх шляхів даного графа, для прочитання яких на кожній горизонтальній лінійці дозволяється бувати тільки один раз. І отже, якщо ми в цьому випадку будемо розв'язувати задачу визначення, наприклад, найкоротших гамільтонових шляхів або найкоротших гамільтонових циклів на основі процедури A'_0 , то перехід до графа D_{\square} тільки впорядковує процес пошуку найкоротших гамільтонових шляхів у графі за рахунок формування на основі рекурентного співвідношення, наведеного у кроці 2 узагальненої процедури A'_0 . Неважко показати, що в цьому випадку оптимальний розв'язок буде точним, якщо процес роботи процедури безперервний, тобто в процесі її роботи або не виникає ситуацій, коли множини $m_{sj}^r = \emptyset$, або якщо така ситуація виникла, то всі шляхи, що ведуть у j у графі $G(V, E)$ рангів $r + 1, \dots, r = n$ довше, ніж шляхи, отримані в j на ранзі $r - 1$ (доведення справедливості даного твердження елементарне й тому не наводиться). Однак легко бачити, що якщо множина m_{sj}^r виявилася порожньою на деякому ранзі $r = q$, то може виявитися, що у вершині j існує довший, ніж побудований процедурою A'_0 , шлях $\mu_{sj}^{r=q}$, але продовження якого надалі може дозволити одержати коротший гамільтонів шлях. Дана ситуація фактично є «каменем спотикання» при розв'язанні всіх NP-повних задач і

більш ніж трьохсотрічний досвід показує, що в цьому випадку гарантовано одержати точний розв'язок можна тільки повним перебором або неявним повним перебором на основі методу галузей і границь, що при досить великій розмірності задачі одне й теж. Отже, якщо ми розглядаємо відображення графа $G(V, E)$ в D_{\square} , і об'єктами L_j в $G(V, E)$ є самі шляхи графа, то перехід до аналізу на графі D_{\square} не змінює дану ситуацію й ми стикаємося з тією ж проблемою, що й розв'язання даної задачі на графі $G(V, E)$ без використання перетворень. Однак ситуація в деяких випадках змінюється, з погляду одержання оптимального розв'язку, якщо аналізований об'єкт у графі $G(V, E)$ відмінний від шляху в даному графі, наприклад, цикл, кліка й т.д., а множини шляхів в обох графах, як впливає з побудови графа D_{\square} , при цьому збігаються. Уведемо процедуру B перетворення об'єктів $\{L_j\}$ у графі D_{\square} .

Процедура B

КРОК 1. Використовуючи правило R , формуємо множину всіх об'єктів $L_j^{r=2}$, що складаються з $r = 2$ базових елементів з вагами $d_j^{r=2}$, і виділяємо об'єкт L_j^{*r} з мінімальною вагою $d_j^{*r=2}$, якщо таких мало, то виділяємо їх усі.

КРОК 2. На основі виділеного об'єкта L_j^{*r} й правила R формуємо всі можливі об'єкти $L_j^{r:=r+1}$, що складаються з $r = r + 1$ базових елементів з вагами $d_j^{r:=r+1}$, і виділяємо об'єкт $L_j^{*r:=r+1}$ з мінімальною вагою $d_j^{*r:=r+1}$, якщо таких мало, то виділяємо їх усі.

КРОК 3. Перевіряємо на основі L_j^{*r} й правила R , чи можна побудувати об'єкти з $r = r + 1$ базових елементів, якщо ні, то процедура закінчує роботу, інакше переходимо до виконання кроку 2.

У результаті роботи процедури B одержимо об'єкти $L_j^{*r=2}$ $L_j^{*r=3}$... $L_j^{*r=k}$, які містять відповідно по 2, по 3 і т.д. по k базових елементів. Неважко побачити, що справедливе твердження 7.1.

Твердження 7.1. Якщо об'єкти $\{L_j\}$ задовольняють властивість B , яка полягає в тому, що застосування процедури B к множині $\{L_j\}$ дозволяє побудувати об'єкти $L_j^{*r=2}, L_j^{*r=3}, \dots, L_j^{*r=\kappa}$, які містять відповідно по 2, по 3 і т.д. по κ базових елементів з мінімальними вагами d_j^{*r} , то процедура A'_0 дає точний розв'язок задачі. Твердження справедливе, оскільки, якщо, припустити, що властивість B виконується, й на основі процедури A'_0 отриманий глобальний екстремум, який не є точним розв'язком, то це можливо, якщо існують об'єкти $L_j^{**r=2}, L_j^{**r=3}, \dots, L_j^{**r=\kappa}$ довжини яких менші, ніж в об'єктів $L_j^{*r=2}, L_j^{*r=3}, \dots, L_j^{*r=\kappa}$, що суперечить первісному припущенню про те, що властивість B виконується. Таким чином, якщо властивість B виконується, запропонована процедура A'_0 буде давати точне розв'язання в умовах безперервності роботи процедури, а якщо не виконується то наближене. Слід також зазначити, що у випадку виконання властивості B наявність порожніх множин m_{sj}^r при роботі процедури A'_0 означає, що на основі правил R побудувати об'єкт $L_j^{r=r}$, що містить базовий елемент j , або не можна в принципі (останнє потрібно обґрунтовувати для кожної задачі окремо), або що будувати даний елемент не має сенсу, оскільки його вагова характеристика буде істотно гірше тих, що побудовані на ярусі.

Твердження 7.2. Якщо аналізовані об'єкти $\{L_j\}$ у графі $G(V, E)$ відмінні від шляху в даному графі, наприклад, цикл, кліка й т.д., то вони завжди задовольняють властивість B у графі D_\square .

Доведення. Нехай на основі процедури B у графі D_\square побудовані всі об'єкти, які містять 2 базових елементи, що задовольняють правило R , виділимо серед них об'єкт $L_j^{*r=2}$, що має мінімальну довжину $d_j^{*r=2}$, побудуємо на його основі всі можливі об'єкти $L_j^{r=3}$, що містять 3 базових елементи, і виділимо серед них об'єкт $L_j^{*r=3}$, що має мінімальну довжину $d_j^{*r=3}$ щодо даного об'єкта. Можна стверджувати, що він є об'єктом з

мінімальною довжиною, який можна побудувати, якщо процес побудови безперервний. Дійсно, припустимо, що в $G(V, E)$ існує об'єкт $L_j^{**r=3}$, що має меншу довжину, ніж $L_j^{*r=3}$, але це можливо, якщо в $G(V, E)$, і відповідно в D_{\square} існує об'єкт $L_j^{**r=2}$ з довжиною, меншою, ніж довжина $L_j^{*r=2}$, але це суперечить установленому факту, що $L_j^{*r=2}$ об'єкт мінімальної довжини в $G(V, E)$, і отже, наше припущення неправильно, аналогічні міркування можна на основі принципу повної математичної індукції провести і для об'єктів $L_j^{*r=3} \dots L_j^{*r=k}$, і довільного об'єкта L_j^{*r} й, отже, властивість B для них виконується. Таким чином, із тверджень 7.1 і 7.2 випливає, що умова того, що відображення $G(V, E)$ у граф D_{\square} не є відображення $G(V, E)$ на самого себе, тобто аналізовані об'єкти відмінні від шляху в даному графі $G(V, E)$, є необхідною, але недостатньою умовою для виконання властивість B (достатність, на жаль, необхідно обґрунтовувати для кожної задачі окремо). Отже, для випадку, коли об'єкти $\{L_j\}$ характеризуються однією ваговою характеристикою, можна спростити процедуру A'_0 за рахунок виділення на ярусах глобальних екстремумів і наступного формування всіх можливих шляхів наступного рангу в графі D_{\square} . Позначимо цю процедуру A''_0 , що буде мати нижчеподаний вигляд.

Процедура A''_0

КРОК 1. З вершини S будуються шляхи рангу $r = 1$, що задовольняють правила R , і відповідно до правил R визначаються їхні вагові характеристики.

КРОК 2. На основі шляхів поточного рангу r будуються всі можливі шляхи наступного рангу $r = r + 1$, що задовольняють правила R , на основі такого рекурентного співвідношення:

$$\mu_{SP}^{r:=r+1} = \min_{d_j(m_{sj}^r \cup (j,p))} (\max)\{m_{sj}^r \cup (j,p)\}; j = (\overline{1,n}); p = (\overline{1,n}); j \neq p,$$

де вагові характеристики d_l об'єктів визначаються відповідно до правил $P \in R$, серед них на ярусі виділяється шлях μ_{sj}^{*r} , максимальний або мінімальний на ярусі по довжині, і на основі формуються будь-які шляхи рангу $r = r + 1$ у графі D_{\square} .

КРОК 3. Перевіряємо підмножини m_{sj}^r поточного рангу r , порожні чи ні, якщо ні, то переходимо до виконання кроку 2, інакше процедура закінчує роботу, і з отриманих на останньому ранзі шляхів вибирається кращий (максимальний або мінімальний) по довжині шлях.

Розглянемо застосування процедури A'_0 й A''_0 для розв'язання задач визначення незалежної максимальної множини й мінімального гамільтонового циклу в довільних графах. Остання задача в теорії графів широко відома як задача про комівояжера. Нехай заданий граф G (рис. 7.2), зважений по вершинах (ваги вершин наведені в табл. 7.1), потрібно визначити незалежну максимальну множину вершин.

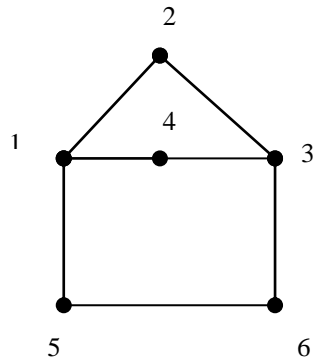


Рис. 7.2. Граф G

Таблиця 7.1

Ваги вершин графа

Номер вершини	1	2	3	4	5	6
Вага вершини	4	1	5	2	2	1

Тут правило R , буде полягати в тому, що в множини поєднаних вершин у графі D_{\square} допускається в тому випадку,

якщо вони в графі G не зв'язані, а ваговою характеристикою буде сума ваг вершин, об'єднаних у шлях у графі D_{\square} . Послідовність розв'язання задачі визначення незалежної максимальної множини вершин мінімальної ваги на основі процедури A'_0 показана в табл. 7.2.

Таблиця 7.2

Процес розв'язання задачі визначення незалежної
максимальної множини

S1(4) 1	— 1	1 —	— 1
S2(1) 2	S62(2) 2	2 —	— 2
S3(5) 3	— 3	3 —	— 3
S4(2) 4	— 4	S624(4) 4	— 4
S5(2) 5	— 5	5 —	— 5
S6(1) 6	— 6	6 —	— 6

Як впливає з табл. 7.2, оптимальному розв'язку задачі відповідають шляхи S624(4) і S625(4), тобто підмножини вершин $\{2, 4, 6\}$ і $\{2, 5, 6\}$ утворюють у графі G (рис. 7.2) незалежні максимальні множини з мінімальною вагою, рівною 4.

У випадку застосування процедури A''_0 процес розв'язання задачі може бути поданий у вигляді табл. 7.3.

Таблиця 7.3

Процес розв'язку задачі на основі процедури A_0''

S1(4) 1	S31(1) S61(5) 1	_____ 1	_____ 1
S2(1) 2	S42(3) S52(3) S62(2) 2	S542(5) S462(4) 2	_____ 2
S3(5) 3	S13(3) S53(7) 3	_____ 3	_____ 3
S4(2) 4	S24(3) S54(4) 4	S624(4) S254(5) 4	_____ 4
S5(2) 5	S25(3) S35(7) S45(4) 5	S625(4) S245(5) 5	_____ 5
S6(1) 6	S16(5) S26(2) S46(3) 6	S316(10) S426(4) 6	_____ 6

Розглянемо розв'язання задачі комівояжера на основі процедури A_0' для графа G повного графа G з числом вершин $n = 5$, заданого матрицею C .

$$C = 6 \begin{vmatrix} & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 2 & 3 & 0 & 2 & 1 & 1 \\ 3 & 1 & 2 & 0 & 2 & 3 \\ 4 & 4 & 1 & 2 & 0 & 4 \\ 5 & 2 & 1 & 3 & 4 & 0 \end{vmatrix}$$

Процес розв'язання даної задачі процедурою A_0' наведений у табл. 7.4.

Таблиця 7.4

Процес розв'язання задачі комівояжера процедурою A'_0

S1(0)	21(3) 31(1) 41(4) 51(2)	1	_____	2134(6)	_____
1		1	1	1	1
S2(0)	12(3) 32(2) 42(1) 52(1)	2	(234)' (5)	_____	_____
2		2	2	2	2
S3(0)	13(1) 23(2) 43(2) 53(3)	3	_____	_____	_____
3		3	3	3	3
S4(0)	14(4) 24(1) 34(2) 54(4)	4	_____	_____	_____
4		4	4	4	4
S5(0)	15(2) 25(1) 35(3) 45(4)	5	_____	_____	25134(7)
5		5	5	5	5

У загальному випадку, додаючи будь-яку вершину p до циклу, що містить r вершин, необхідно перевіряти всі r варіантів розміщення вершини p у циклі й серед них вибирати найкоротший. У табл. 7.4 цикл з трьох вершин 234, довжиною 5, є найкоротшим, на його основі можна побудувати цикли, додаючи вершини 1 і 5. Додаючи вершину 1 ми одержимо цикли (див. рис.7.3 і 7.4).

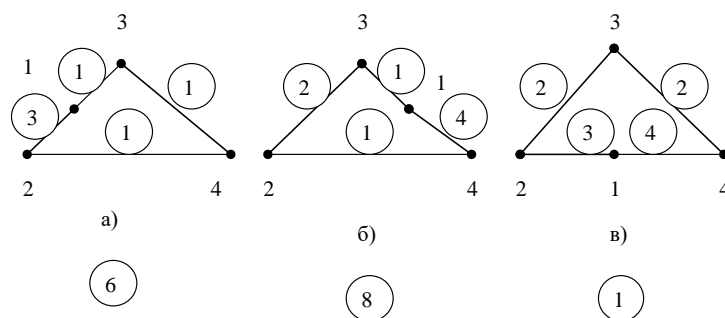


Рис. 7.3. Варіанти розміщення вершини 1

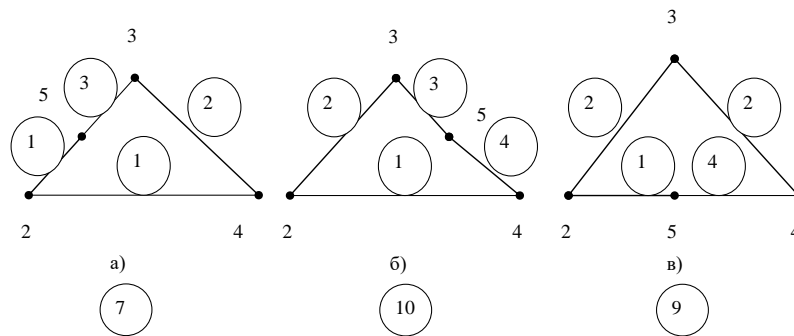


Рис. 7.4. Варіанти розміщення вершини 5

Як видно з рис. 7.3 і 7.4 на наступному ранзі ми можемо сформуванати цикли 2134 довжиною 6 і 2534 довжиною 7, оскільки цикл 2134 найкоротший на ярусі, то його тільки й залишаємо, і на його основі на наступному ярусі формуємо найкоротший цикл з п'яти вершин 25134 довжини 7. Оскільки кожний цикл доводиться перевіряти на можливість розширення r раз, а максимальне число шляхів в одній підмножині не може перевищити $(n - 1)$, число ярусів дорівнює n , максимальний ранг шляху теж не перевищує n , то загальна складність алгоритму розв'язання задачі комівояжера не перевищить $O(n^4)$. У загальному випадку при довільних об'єктах складність процедури A'_0 не перевищить $O(kn^3)$, де k – число операцій, необхідних для визначення оптимальної ваги об'єкта при додаванні в нього ще одного базового елемента. Тепер розглянемо можливість розв'язання задач дискретної оптимізації для випадку, коли об'єкт характеризується $m + 1$ вагою й на m ваг накладається обмеження, у цьому випадку ми приходимо до задач лінійного й нелінійного булевого програмування. Задачі булевого лінійного й нелінійного програмування є моделями широкого класу прикладних задач у теорії побудови складних систем і при цьому задачі булевого лінійного програмування належать до класу NP-повних, важко розв'язуваних задач, а ефективні методи розв'язання задач нелінійного булевого програмування з довільними нелінійностями практично відсутні. У цей час для кожного типу задачі розробляється свій метод розв'язання. Покажемо, що на основі запропонованої узагальненої процедури може бути розв'язана будь-яка задача булевого програмування.

7.2. Загальний метод розв'язання довільних задач булевого програмування

Задачі булевого лінійного й нелінійного програмування є моделями широкого класу прикладних задач у теорії побудови складних систем і при цьому задачі булевого лінійного програмування належать до класу NP-повних важко розв'язуваних задач [78, 79], а ефективні методи розв'язання задач нелінійного булевого програмування з довільними нелінійностями практично відсутні. У цей час для кожного типу задачі розробляється свій метод розв'язання. Актуальним є вироблення єдиного підходу до розв'язання даного класу задач, що забезпечує їхнє розв'язання з необхідною оперативністю й точністю. Розглянемо підхід, що дозволяє розв'язувати задачі булевого лінійного програмування й нелінійного програмування з довільними нелінійностями тим самим алгоритмом, заснованим на ідеях рангового підходу.

Формалізація й постановка задачі. Для опису всієї множини адитивних цілочисельних функцій з довільними нелінійностями, обумовленими на множині змінних $\{X_1, X_2, \dots, X_n\}$ уведемо поняття позначка функції $F(x)$ рівної

$$F(x) = \sum_{j=1}^{p_1} C_{1j} S_1(C_n^1) + \sum_{j=1}^{p_2} C_{2j} S_2(C_n^2) + \dots + \sum_{j=1}^{p_k} C_{kj} S_k(C_n^k) + \dots + \sum_{j=1}^{p_n} C_{nj} S_n(C_n^n), \quad (7.1)$$

де $S_r(C_n^r) = S_1 + S_2 + \dots + S_{p_r}$ – сума всіх можливих сполучень добутоків змінних, що містяться у кожному добутку $S_r = X_p X_k \dots X_m$ (нелінійності) r рзних змінних;

$$p_r = \frac{n!}{r!(n-r)!};$$

C_{rj} – цілочисельні коефіцієнти в добутках S_r , що містять r змінних.

Позначимо через H множини всіх функцій, яку можна породити на основі $F(x)$, вважаючи рівними нулю рзні

сполучення C_{ij} в (7.1). Множина H є повною у тому розумінні, що містить у собі всі можливі нелінійності, що складаються з усіх можливих сполучень змінних, утворюючих ці нелінійності, які можна взагалі побудувати на основі даної підмножини змінних $\{X_1, X_2, \dots, X_n\}$.

Неважко показати, що потужність даної множини дуже велика, але кінцева й дорівнює 2^{p_Σ} , де

$$p_\Sigma = 1 + \frac{1}{2} \left[\frac{n!}{1!(n-1)!} \left(\frac{n!}{1!(n-1)!} + 1 \right) + \frac{n!}{2!(n-2)!} \left(\frac{n!}{2!(n-2)!} + 1 \right) + \dots + \frac{n!}{k!(n-k)!} \left(\frac{n!}{k!(n-k)!} + 1 \right) + \dots + \frac{n!}{(n-1)!1!} \left(\frac{n!}{(n-1)!1!} + 1 \right) \right].$$

Слід зазначити, що за допомогою співвідношення (7.1) може бути визначений клас задач дискретної оптимізації, у яких розв'язок визначається тільки сполученням змінних і не залежить від перестановки змінних в $S_r(C_n^r)$. Тобто значення C_{ij} у цих задачах залежить тільки від сполучення змінних в $S_r(C_n^r)$. У загальному випадку задачу булевого програмування можна подати у вигляді

$$\begin{aligned} f(X_1, X_2, \dots, X_n) &\Rightarrow \max; \\ g_j(X_1, X_2, \dots, X_n) &\leq b_j; \quad j = (\overline{1, m}), \end{aligned} \quad (7.2)$$

де

$$\begin{aligned} f(X_1, X_2, \dots, X_n) &\in H \\ g_j(X_1, X_2, \dots, X_n) &\in H \\ b_j &\in Z; \quad Z - \text{множина цілих чисел} \\ X_i &\in \{0, 1\}. \end{aligned}$$

Розглянемо граф $G(X, E)$ на рис.7.5, у якому вершини X_i і X_j з'єднані ребром (i, j) , якщо вони можуть бути об'єднані в кліку. У графі $G(X, E)$ кожній вершині X_i відповідає змінна X_i .

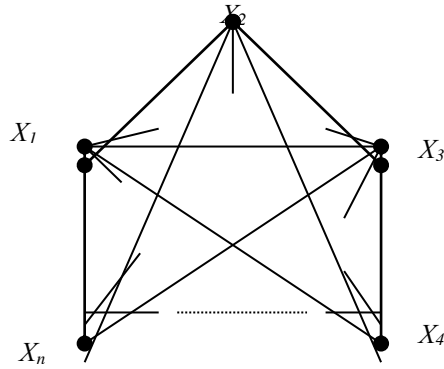


Рис.7.5. Граф G

Виділимо в графі G довільну кліку $Q = X_p X_r \dots X_m$, що складається з r вершин, де $r < n$, і розглянемо її перетинання з

$$S_r(C_n^r) \in f(X_1, X_2, \dots, X_n),$$

а також з

$$S_r(C_n^r) \in g_j(X_1, X_2, \dots, X_n).$$

Кожне перетинання можна охарактеризувати сумами коефіцієнтів C_{lj} , що стоять при $S_r(C_n^r)$ у функціоналі $f(X_1, X_2, \dots, X_n)$ й обмеженнях $g_j(X_1, X_2, \dots, X_n)$, при цьому в загальному випадку довільна кліка Q завжди буде характеризуватися відповідною вагою за функціоналом $f(X_1, X_2, \dots, X_n)$ й не більш ніж m вагами за обмеженнями $g_j(X_1, X_2, \dots, X_n)$. Таким чином, довільна задача булевого програмування може розглядатися як задача знаходження кліки Q^* максимальної ваги за вагами функціонала, у графі G , у якій усі m ваг за вагами обмежень не перевищують відповідно b_j . Так, наприклад, якщо розв'язується задача лінійного програмування

$$\begin{aligned} f(x) &= C_1 X_1 + C_2 X_2 + C_3 X_3 + C_4 X_4 \rightarrow \max; & (7.3) \\ B_1 X_1 + B_2 X_2 + B_3 X_3 + B_4 X_4 &\leq b_1; \\ K_1 X_1 + K_2 X_2 + K_3 X_3 + K_4 X_4 &\leq b_2, \end{aligned}$$

то граф G буде мати вигляд як на рис. 7.6.

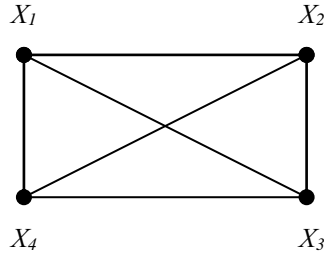


Рис. 7.6. Граф G для задачі лінійного програмування

Як видно з табл. 7.5, кожна вершина графа G характеризується трьома вагами (C_i, B_i, K_i) і для розв'язання задачі в графі G (рис. 7.6) варто знайти кліку Q^* зі зважених вершин таку, щоб її сумарна вага за вагами $\{C_i\}$ була максимальною і при цьому сумарні ваги за вагами $\{B_i\}$ і $\{K_i\}$ не перевищували відповідно величин b_1 і b_2 . Слід зазначити, що в задачах лінійного програмування з нелінійністю, вище другої (тобто при наявності у функції мети або обмеженнях добутоків, що складаються з числа змінних, більше двох), вагові характеристики ребер графа G припускаються рівними нулю. У випадку розв'язання задач квадратичного програмування зручно вводити й вагові характеристики ребер. Так, розглянемо задачу квадратичного програмування такого виду:

$$\begin{aligned}
 f(x) = & C_1X_1 + C_2X_2 + C_3X_3 + C_4X_4 + \\
 & + C_{12}X_1X_2 + C_{13}X_1X_3 + C_{34}X_1X_4 + \\
 & + C_{23}X_2X_3 + C_{24}X_2X_4 + C_{34}X_3X_4 \rightarrow \max, \\
 & B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 \leq b_1, \\
 & K_1X_1 + K_2X_2 + K_3X_3 + K_4X_4 \leq b_2.
 \end{aligned}
 \tag{7.4}$$

Таблиця 7.5

Ваги вершин графа			
X_1	C_1	B_1	K_1
X_2	C_2	B_2	K_2
X_3	C_3	B_3	K_3
X_4	C_4	B_4	K_4

Для неї можна поставити у відповідність граф на рис. 7.7.

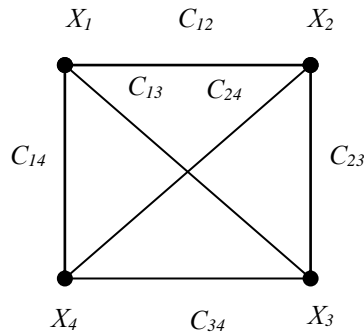


Рис. 7.7. Граф G для завдання квадратичного програмування

Як видно з рис. 7.7 і табл. 7.6, вершинам графа, як і в задачі лінійного програмування, відповідають ваги (C_i, B_i, K_i) , а ребрам ваги $\{C_{ij}\}$. Для розв'язання даної задачі в графі G потрібно знайти кліку Q^* максимальної сумарної ваги по C_i і C_{ij} і при цьому щоб сумарні ваги за вагами $\{B_i\}$ і $\{K_i\}$ не перевищували відповідно величин b_1 і b_2 . У цьому випадку граф G є зваженим і по вершинах і по ребрах, аналогічно зваженими є й усі кліки даного графа. Варто мати на увазі, що у квадратичній задачі $C_{ij} = C_{ji}$. Таким чином, для розв'язання довільної задачі булевого програмування необхідно побудувати алгоритм, що дозволяє знаходити в заданому графі G , зваженому по вершинах або по ребрах, кліку Q^* максимальної сумарної ваги за вагами функціонала $f(X_1, X_2, \dots, X_n)$, у якій усі m ваг за вагами обмежень $g_j(X_1, X_2, \dots, X_n)$ не перевищують відповідних b_j .

Таблиця 7.6

Ваги вершин і ребер графа

X_1	C_1	B_1	K_1
X_2	C_2	B_2	K_2
X_3	C_3	B_3	K_3
X_4	C_4	B_4	K_4

Розв'язання задачі. Для розв'язання задачі скористаємося поданням вихідного графа G у вигляді симетричного дерева

шляхів, запропонованого в роботах [6 – 8]. Зміст такого подання полягає ось у чому. Нехай усі можливі стани деякої системи визначаються графом $G(V, E)$ з n вершинами, де вершини відповідають можливим станам системи. Перейдемо до простору з $(n - 1)^2$ станами. Для цього кожному з n станів поставимо у відповідність ще $(n - 1)$ стан, що характеризує спосіб досягнення стану з множини $\{1, 2, \dots, n\}$. При цьому як додані стани визначимо ранг шляху в графі $G(V, E)$. Тобто з вершини s графа $G(V, E)$ у довільну вершину j можна потрапити шляхом рангу $r=1$, використовуючи одне ребро, шляхом рангу $r=2$, використовуючи 2 ребра й т.д. шляхом рангу $r = n - 1$, використовуючи $n - 1$ ребро. Такий простір станів можна подати у вигляді стягнутого дерева шляхів D , графічно воно може бути зображене як на рис. 7.9.

Дерево всіх шляхів D містить $(n - 1)$ горизонтальну лінійку й $(n - 1)$ ярус. Для прочитання шляхів на кожній горизонтальній лінійці можна бувати тільки один раз. Виходячи зі стягнутого дерева шляхів, для довільної вершини j множини шляхів, що ведуть у цю вершину з деякої вершини s , можна подати в такому вигляді

$$m_s(j) = m_{sj}^{r=1} \cup m_{sj}^{r=2} \cup \dots \cup m_{sj}^{r=n-1}; j = (\overline{1, n-1}), \quad (7.5)$$

де $m_{sj}^r = \{ \mu_{sj}^r \}$ – підмножини шляхів з довільної вершини s у деяку вершину j графа $G(V, E)$ рангу r .

Слід зазначити, що дерево всіх шляхів D може будуватися й від конкретної вершини i графа, у цьому випадку вершина $s = i$ та i -та горизонтальна лінійка виключається в D . Наприклад, при $i = 2$ дерево D буде мати вигляд як на рис. 7.8. Надалі стягнуто дерево шляхів, наведене на рис. 7.8, буде використовуватися для побудови однопрохідних алгоритмів розв'язання задачі (7.2), а дерево D (рис. 7.9) для побудови n -прохідних алгоритмів, суть побудови яких буде пояснена нижче.

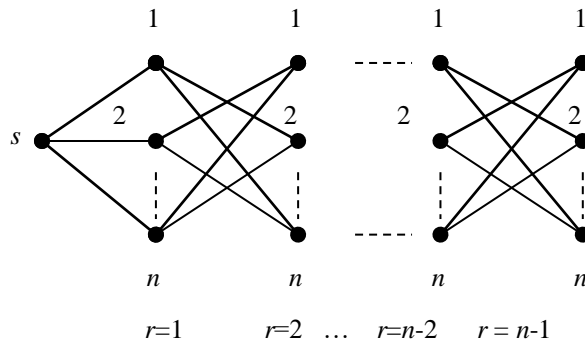


Рис. 7.8. Стягнуто дерево всіх шляхів D графа $G(V, E)$

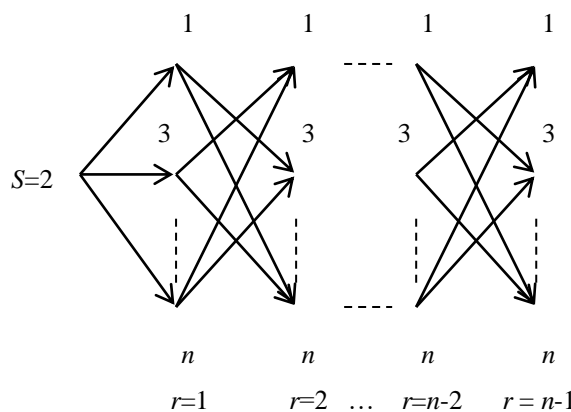


Рис.7.9. Стягнуто дерево всіх шляхів D графа від вершини $s=2$ $G(V, E)$

Таким чином, використовуючи граф D і ввівши правила формування шляхів наступного рангу, ми можемо з довільної вершини s поетапно будувати шляхи $\{\mu_{sj}^r\}$ довільного рангу аж до рангу $r = n - 1$. У нашій задачі під станом системи ми будемо мати на увазі різні способи об'єднання вершин графа D у кліці. Тоді кожному шляху $\{\mu_{sj}^r\}$ рангу r у графі D , що проходить через вершини (v_h, v_k, \dots, v_p) у вихідному графі G розв'язуваної задачі, відповідає кліка з $r - 1$ вершини (x_h, x_k, \dots, x_p) , що характеризується відповідною вагою за функціоналом $f(x_1, x_2, \dots, x_n)$ й не більш ніж m вагами за обмеженнями $g_j(x_1, x_2, \dots, x_n)$. Вагові характеристики $\{d_{sj}^{r-1}\}$ довільної кліки $Q^{r-1}(j)$, що складається з $r - 1$ -вершини й обумовленої одним зі шляхів $\mu_{sj}^r \in m_{sj}^r$ рангу r у

графі D , обчислюються за вагами функціонала шляхом підсумовування коефіцієнтів підмножини $L_f = \{C_{rj}\}$, що стоять при $S_{r-1}(C_n^{r-1}) \in P_f$, де P_f – всі підмножини $\{S_{r-1}(C_n^{r-1})\}_f$, що задовольняють умову

$$S_{r-1}(C_n^{r-1})_f \cap Q^{r-1}(j) \neq \emptyset,$$

а $S_{r-1}(C_n^{r-1})_f$ визначається функціоналом $f(X_1, X_2, \dots, X_n)$. Аналогічно визначаються вагові характеристики за вагами обмежень шляхом підсумовування коефіцієнтів підмножини $L_B = \{C_{rj}\}$, що стоять при $S_{r-1}(C_n^{r-1}) \in P_B$, де P_B – всі підмножини $\{S_{r-1}(C_n^{r-1})\}_B$, що задовольняють умову

$$S_{r-1}(C_n^{r-1})_B \cap Q^{r-1}(j) \neq \emptyset,$$

а $S_{r-1}(C_n^{r-1})_B$ визначається обмеженнями $g_j(X_1, X_2, \dots, X_n)$; $j = (\overline{1, m})$. Таким чином, вагові характеристики клік $Q^{r-1}(j)$ шляхів, що m_{sj}^r характеризуються множиною, за вагами функціонала й обмежень визначаються відповідно рівностями

$$d_{sj}^{f(r-1)} = \sum_{C_{rj} \in L_f} C_{rj}; \quad d_{sj}^{B(r-1)} = \sum_{C_{rj} \in L_B} C_{rj}. \quad (7.6)$$

Отже, у графі D кожний шлях має в загальному випадку $m + 1$ довжину, одну за вагами функціонала й m за вагами обмежень, і для розв'язання поставленої задачі нам у графі D потрібно побудувати шлях максимальної довжини за вагами функціонала від вершин $1, 2, \dots, n$ до всіх інших вершин графа й при цьому його довжини за вагами обмежень не повинні перевищувати відповідну величину b_j . Якщо на основі підмножин шляхів m_{sj}^{r-1} у графі D будувати підмножини m_{sj}^{r-2} й т. д. до m_{sj}^{r-n-1} , то ми змушені будемо побудувати $(n - 1)!$ шляхів, тому для формування шляхів уводиться процедура A , що дозволяє відсікати безперспективні шляхи. Для відсікання

безперспективних варіантів у процедурі A пропонується використовувати принцип оптимізації в напрямку до довільної вершини p , при формуванні шляхів наступного рангу m_{sp}^{r+1} на основі шляхів попереднього рангу m_{sj}^r , запропонованих у роботах [41, 49], що для розглянутої задачі визначається таким рекурентним співвідношенням:

$$\mu_{sp}^{r+1} = \max_j \{ \mu_{sj}^r \cup (j, p) \}; j = (\overline{1, n}); p = (\overline{1, n}); j \neq p, \quad (7.7)$$

де (j, p) – ребро графа D ;

n – число різних вершин у графі D .

Розглянемо можливість побудови n -прохідних і однопрохідних процедур розв'язання задачі (7.2) відповідно на стягнутих деревах, наведених на рис. 7.8 і 7.9.

Перед початком роботи процедури A_i змінної $i = 1$.

Процедура A_1 з n проходами

КРОК 1. Змінної $s := i$ та з вершини s будуються всі можливі шляхи рангу $r = 1$ до всіх вершин графа D (рис. 7.9), що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, при цьому довжини за вагами функціонала й обмеженнями обчислюються відповідно до співвідношень (7.6).

КРОК 2. Використовуючи шляхи поточного рангу r , будуються всі можливі шляхи рангу $r = r + 1$, що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, з використанням рекурентного співвідношення (7.6). При цьому перевірка обмежень і вибір шляху, максимального за вагами функціонала, здійснюється на основі обчислень довжин шляхів за вагами функціонала й обмежень відповідно до співвідношень (7.6). (Слід зазначити, що якщо в процесі застосування рекурентного співвідношення (7.7) виникають кілька шляхів однакової довжини, то необхідно їх усі продовжувати на наступному ранзі).

КРОК 3. Перевіряємо $m_{sj}^{r+1} = \emptyset$, якщо так, то шлях μ_{sj}^{*r} максимальної довжини, отриманий на ранзі r , є локальним екстремумом розв'язуваної задачі, інакше виконуємо наступний крок.

КРОК 4. Перевіряємо $i = n - 1$, якщо ні, то $i := i + 1$ і переходимо до виконання кроку 1, інакше процедура A_1 закінчує роботу, при цьому з множини локальних екстремумів $\{\mu_{sj}^{*r}\}$ вибирається глобальний μ_{sj}^{**r} , що відповідає оптимальному розв'язку задачі (7.2).

Для зниження тимчасової складності роботи алгоритму можливо використовувати однопрохідний варіант реалізації даної процедури на основі стягнутого дерева шляхів, наведеного на рис. 7.8. При цьому однопрохідна процедура A_2 має нижченаведений вигляд.

Процедура A_2 з 1 проходом

КРОК 1. З вершини s будуються всі можливі шляхи рангу $r = 1$ до всіх вершин графа D (рис. 7.8), що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, при цьому довжини за вагами функціонала й обмеженнями обчислюються відповідно до співвідношень (7.6).

КРОК 2. Використовуючи шляхи поточного рангу r , будуються всі можливі шляхи рангу $r = r + 1$, що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, з використанням рекурентного співвідношення (7.7). При цьому перевірка обмежень і вибір шляху, максимального за вагами функціонала, здійснюється на основі обчислень довжин шляхів за вагами функціонала й обмежень відповідно до співвідношень (7.6). (Слід зазначити, що якщо в процесі застосування рекурентного співвідношення (7.7) виникають кілька шляхів однакової довжини, то необхідно їх усі продовжувати на наступному ранзі).

КРОК 3. Перевіряємо $m_{sj}^{r+1} = \emptyset$, якщо так, то шлях μ_{sj}^{*r} максимальної довжини, отриманий на ранзі r , є локальним екстремумом розв'язуваної задачі, інакше виконуємо наступний крок.

КРОК 4. Перевіряємо ранг $r = n$, якщо ні, то переходимо до виконання кроку 2, інакше процедура A_2 закінчує роботу, і шлях μ_{sj}^{*r} максимальної довжини, отриманий на ранзі $r = n$, відповідає оптимальному розв'язку задачі.

Ще одним варіантом зменшення тимчасової складності алгоритмів на основі процедур A_1 і A_2 можуть бути процедури A' і A'' , що відрізняються від A_1 і A_2 тим, що на кожному ярусі формування шляхів на основі процедур A_1 і A_2 локальні екстремуми будуть виділятися не в кожній множині, а виділяються глобальні екстремуми на ярусі й на основі шляху, що відповідає глобальному екстремуму на ярусі, формуються шляхи наступного ярусу, що задовольняють обмеження, при цьому процедури A' і A'' будуть мати нижчеподаний вигляд.

Процедура A'

Перед початком роботи процедури A' змінної $i := 1$.

КРОК 1. Змінної $s = i$ та з вершини s будуються всі можливі шляхи рангу $r = 1$ до всіх вершин графа D (рис. 7.9), що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, при цьому довжини за вагами функціонала й обмеженнями обчислюються відповідно до співвідношень (7.6). Далі виділяється найдовший шлях на ярусі.

КРОК 2. Використовуючи найдовший шлях поточного рангу r , побудований на попередньому кроці, будуються всі можливі шляхи рангу $r = r + 1$, що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, з використанням рекурентного співвідношення (7.7). При цьому перевірка обмежень і вибір шляху, максимального за вагами функціонала, здійснюється на основі обчислень довжин шляхів за вагами функціонала й обмежень відповідно до співвідношень (7.6).

КРОК 3. Перевіряємо $m_{sj}^{r+1} = \emptyset$, якщо так, то шлях μ_{sj}^{*r} максимальної довжини, отриманий на ранзі r , є локальним екстремумом розв'язуваної задачі, інакше виконуємо наступний крок.

КРОК 4. Перевіряємо $i = n - 1$, якщо ні, то $i = i + 1$ і переходимо до виконання кроку 1, інакше процедура A' закінчує роботу, при цьому з множини локальних екстремумів (отриманих за один прохід) $\{\mu_{sj}^{*r}\}$ вибирається глобальний μ_{sj}^{**r} , що відповідає оптимальному розв'язку задачі.

Процедура A''

КРОК 1. З вершин s будуються всі можливі шляхи рангу $r = 1$ до всіх вершин графа D (рис. 7.8), що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, при цьому довжини за вагами функціонала й обмеженнями обчислюються відповідно до співвідношень (7.6). Далі виділяється найдовший шлях на ярусі.

КРОК 2. Використовуючи найдовший шлях поточного рангу r , побудований на попередньому кроці, будуються всі можливі шляхи рангу $r = r + 1$, що задовольняють обмеження $g_j(X_1, X_2, \dots, X_n); j = (\overline{1, m})$, з використанням рекурентного співвідношення (7.7). При цьому перевірка обмежень і вибір шляху, максимального за вагами функціонала, здійснюється на основі обчислень довжин шляхів за вагами функціонала й обмежень відповідно до співвідношень (7.6).

КРОК 3. Перевіряємо $m_{sj}^{r+1} = \emptyset$, якщо так, то шлях μ_{sj}^{*r} максимальної довжини, отриманий на ранзі r , є локальним екстремумом розв'язуваної задачі, інакше виконуємо наступний крок.

КРОК 4. Перевіряємо ранг $r = n$, якщо ні, то переходимо до виконання кроку 2, інакше процедура A'' закінчує роботу, і шлях μ_{sj}^{*r} максимальної довжини, отриманий на ранзі $r = n$, відповідає оптимальному розв'язку задачі.

Оцінка складності процедур $\{A\}$

Оскільки число шляхів, що будуються на довільному ранзі r , не може перевищити $(n-1)*(n-1)$, максимальний ранг r довільного шляху не перевищує $(n-1)$, а число циклів, виконване процедурою A_1 , дорівнює n , то після n циклів число шляхів, яке побудує

процедура A_1 , не може перевершити $(n-1)*(n-1)*(n-1)*n \approx n^4$, а число оброблених векторів n^5 . Таким чином, тимчасова складність алгоритму не перевищить у найгіршому разі $O(n^5 (m + 1))$. У випадку, коли розв'язання задачі здійснюється за один прохід процедури A_2 або за один прохід процедури A'' , але з виділенням найдовшого шляху на ярусі складність процедур A_2 і A'' не перевищить відповідно $O(n^4 k(m + 1))$ і $O(n^3 k(m + 1))$. Отже, алгоритми A_5 , A_4 , A_3 мають відповідно тимчасову складність, що не перевищує в найгіршому разі $O(n^5 k(m + 1))$, $O(n^4 k(m + 1))$ і $O(n^3 k(m + 1))$. Таким чином, запропонований підхід до розв'язання довільних задач булевого програмування дозволяє на основі запропонованих алгоритмів вирішувати з єдиних позицій будь-які задачі лінійного й нелінійного програмування за поліноміальний час з необхідною точністю. Якщо в співвідношення (7.1) увести позначення

$$f_k(x) = \sum_{j=1}^{p_k} C_{kj} S_k(C_n^k),$$

то воно набуде виду $F(x) = \sum_k f_k(x)$.

У загальному випадку, якщо функціонал і обмеження являють собою довільну нелінійну функцію від $f_k(x)$, то й така задача нелінійного програмування також може бути ефективно розв'язана за допомогою запропонованого підходу й при цьому будуть змінюватися тільки правила ваг вершин клік.

7.3. Метод розв'язання задачі визначення найкоротших гамільтонових шляхів

Ми розглядали задачі булевого програмування, у яких коефіцієнти C_{ij} в (1) не залежать від порядку проходження змінних, а тільки від їхніх сполучень. Покажемо на прикладі задачі про найкоротший гамільтонів шлях, що на основі запропонованої процедури $\{A\}$ можуть ефективно розв'язуватися й задачі булевого програмування, у яких довжина шляху в графі G залежить не тільки від сполучення змінних, але й від перестановок змінних.

Процедура *A* для визначення найкоротшого гамільтонового шляху в довільному графі

З вершини s будуються всі можливі шляхи рангу $r = 1$ до всіх вершин графа D , далі, використовуючи шляхи рангу $r = 1$, будуються всі можливі шляхи рангу $r = 2$ і на їхній основі формуються шляхи наступного рангу, з використанням рекурентного співвідношення, і т.д. до побудови шляхів рангу $r = n - 1$.

$$\mu_{sp}^{r+1} = \min_j \{ \mu_{sj}^r \} \cup (j, p); j = (\overline{1, n}); p = (\overline{1, n}); j \neq p. \quad (7.8)$$

Слід зазначити, що якщо в процесі застосування рекурентного співвідношення (7.8) виникають кілька найкоротших шляхів однакової довжини, то необхідно їх усі будувати на наступному ранзі.

Можна виділити нижченаведені особливості роботи процедури A . У процесі її роботи може виникати дві ситуації. Перша, коли процедура A на кожному кроці побудувала шляхи в множині m_{sj}^r , тобто принцип оптимальності роботи процедури не порушувався, й друга, коли до однієї з вершин жодного шляху побудувати не можна. Остання обставина можлива у двох випадках:

а) якщо аналізований граф неповний і до вершини p не існує шляху деякого рангу $r = k$;

б) коли деяка вершина p увійде в усі шляхи попереднього рангу.

Неважко бачити, що в першій ситуації процедура A не втрачає оптимального розв'язку, а в другій ситуації (випадок «б») оптимальний розв'язок може бути втрачений, оскільки принцип оптимальності роботи процедури порушується. Це означає, що на основі процедури A ми можемо побудувати тільки наближений алгоритм розв'язання задачі. Однак після порушення принципу оптимальності наступне продовження шляхів з використанням рекурентного співвідношення (7.8) дозволить мінімально відхилитися від оптимального розв'язку.

Алгоритм A_1 розв'язання задачі визначення найкоротших гамільтонових шляхів у довільних графах

КРОК 1. Привласнюємо значення $S := 1$ змінній S і в графі D , використовуючи процедуру A , визначаємо найкоротші шляхи рангу $r = n - 1$ від вершини S до всіх інших вершин графа й збільшуємо значення $S = S + 1$.

КРОК 2. Перевіряємо $S = n$, якщо ні, то переходимо до виконання кроку 1, інакше виконуємо наступний крок.

КРОК 3. Серед усіх побудованих найкоротших шляхів рангу $r = n - 1$ на кроках 1 і 2 вибираємо найкоротший, і алгоритм закінчує роботу.

Для розв'язання задачі за один прохід визначимо точку входу, починаючи з якої почне роботу алгоритм. Для цього введемо поняття мінімально вилученої вершини i у довільному графі $G(V, E)$

Нехай заданий граф $G(V, E)$ з n вершинами й множиною ребер E , кожному з яких присвоєна вага у вигляді довільного позитивного числа β_i . Поставимо кожній вершині s графа $G(V, E)$ у відповідність вектор X_s

$$X_s = \{ \beta_1, \beta_2, \dots, \beta_s, \emptyset, \beta_{s+1}, \dots, \beta_n \}; \quad s = (\overline{1, n}),$$

у якому на s -й позиції стоїть символ \emptyset , який означає, що елемент β_s у X_s відсутній. Множина $\{ \beta_i \}$ являє собою множину позитивних чисел, до яких будемо відносити й 0. Можна сказати, що деяка вершина i мінімально відділена від вершини s , обумовленої вектором X_s , якщо відстань $d(s, i)$ між вершинами s і i визначається співвідношенням

$$d(s, i) = \min_i \{ \beta_i \in X_s \}. \quad (7.9)$$

Якщо існує Q вершин i , що задовольняють співвідношення (7.9), то це означає, що існує Q мінімально відділених вершин від вершини s , обумовленої вектором X_s . Визначивши вершину i входу для графа можна розв'язувати задачу визначення найкоротшого гамільтонового шляху за один прохід процедури

A_1 , при цьому похибка у порівнянні з n -прохідною процедурою, як буде показано нижче, збільшується не більш ніж на один відсоток, а складність алгоритму знижується в $n - 1$ раз. Алгоритм з використанням мінімально відділеної точки позначимо A_2 .

7.4 Експериментальне дослідження алгоритмів розв'язання задач булевого програмування на основі теорії графів

При експериментальному дослідженні розроблених алгоритмів основна увага приділена оцінці тимчасової складності алгоритмів, оцінці похибки наближених алгоритмів і порівняльному аналізу розроблених алгоритмів з відомими алгоритмами.

Експериментальне дослідження алгоритмів на основі розроблених процедур для загальних методів розв'язання задач булевого програмування

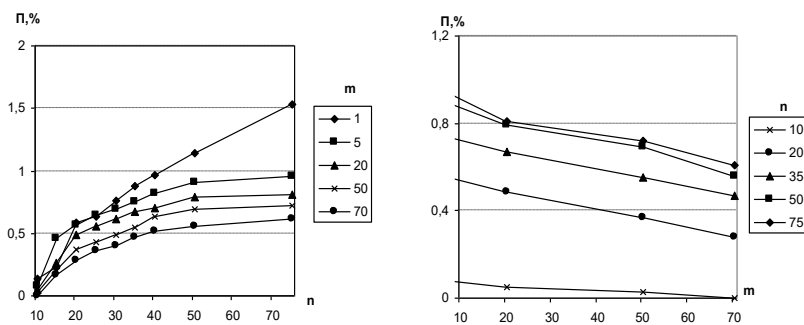
Досліджувалися такі алгоритми: алгоритм A_5 на основі багато прохідної процедури A_1 , алгоритм A_4 на основі однопрохідної процедури A_2 і алгоритм A_3 на основі однопрохідної процедури A'' . При дослідженні коефіцієнти у функціоналі й обмеженнях генерувалися за рівномірним законом розподілу у функціоналі в діапазоні від 0 до 10, а в обмеженнях від 0 до 20. На кожну точку при оцінці тимчасової складності алгоритмів у середньому й похибці алгоритмів розв'язувалося не менше 50 тестових задач, результати отримані з довірчою ймовірністю 0,95. Як точний алгоритм використовувався розроблений алгоритм для задачі квадратичного й лінійного програмування, скомбінованого на основі використання для прогнозування ідей рангового підходу, а для відсікання безперспективних шляхів методу галузей і границь, що дозволив зняти похибки для задач до розмірності, що не перевищує $n = 70$. Графіки залежності похибки від розмірності n розв'язуваних задач і від числа обмежень t у задачі (7.2) наведені на рис. 7.10-7.12, з яких видно, що похибка алгоритмів зі збільшенням числа обмежень t асимптотично зменшується, зі збільшенням n зростає

й при $m \geq 50$ похибка алгоритмів стабілізується та для задач лінійного програмування не перевищує 2 %, а для задач квадратичного програмування 5-10 %.

Розв'язання тестових задач показало, що збільшення діапазону зміни коефіцієнтів у функціоналі й обмеженнях приводить до різкого зниження похибок алгоритмів, перехід від нелінійності одного порядку до нелінійності більш високого порядку може приводити до незначного зростання похибки при невеликому числі обмежень, а зі зростанням числа обмежень зростання похибки дуже швидко компенсується. Експериментальне дослідження тимчасової складності показало (рис. 7.13), що число оброблених векторів не залежить від числа обмежень і в середньому для алгоритмів A_5 , A_4 , A_3 тимчасова складність не перевищує відповідно

$$O(0,1n^{4,9}), O(0,3n^{3,7}) \text{ і } O(0,4n^{2,8}).$$

а)



б)

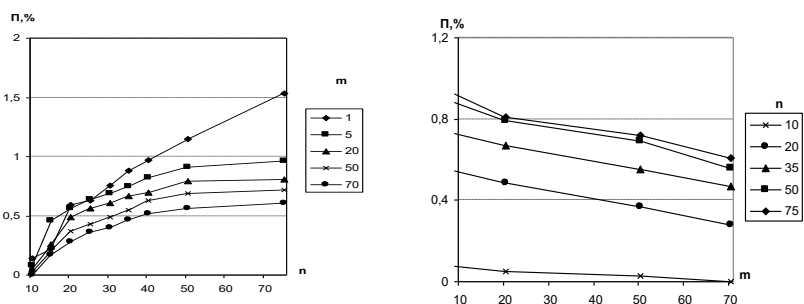
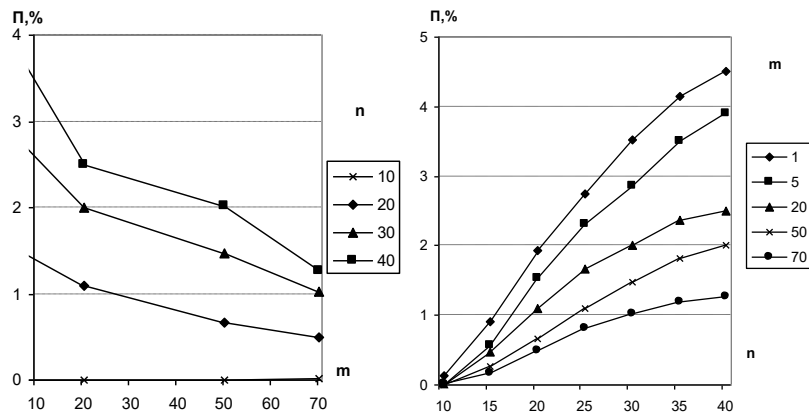


Рис.7.10 Залежність похибки алгоритму A_5 від розмірності розв'язуваної задачі:

- а) лінійного булевого програмування n при різному числі обмежень m ;
- б) квадратичного булевого програмування n при різному числі обмежень m

а)



б)

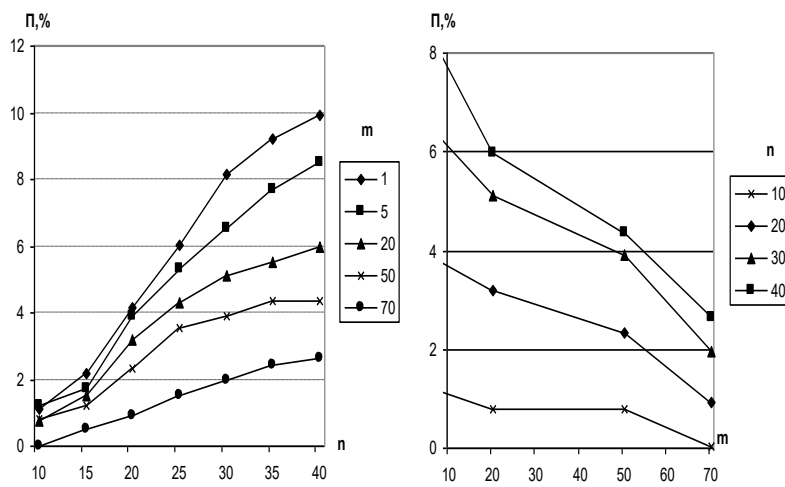
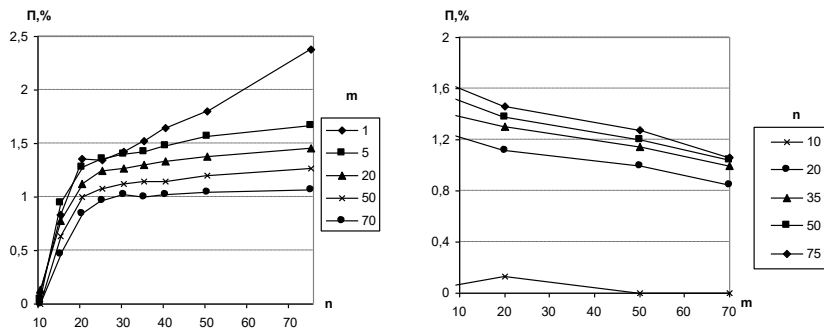


Рис. 7.11. Залежність похибки алгоритму A_4 від розмірності розв'язуваної задачі:

- а) лінійного булевого програмування n при різному числі обмежень m ;
- б) квадратичного булевого програмування n при різному числі обмежень m

а)



б)

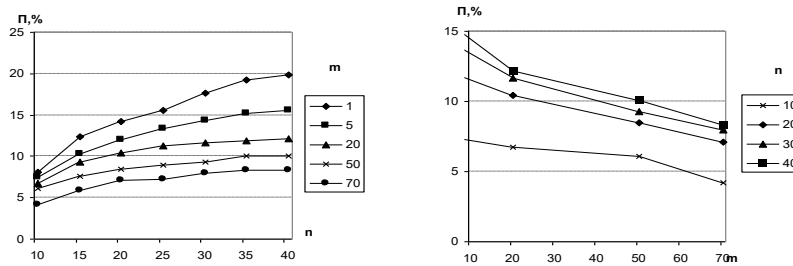


Рис. 7.12. Залежність похибки алгоритму A_3 від розмірності розв'язуваної задачі:

- а) лінійного булевого програмування n при різному числі обмежень m ;
- б) завдання квадратичного булевого програмування n при різному числі обмежень m

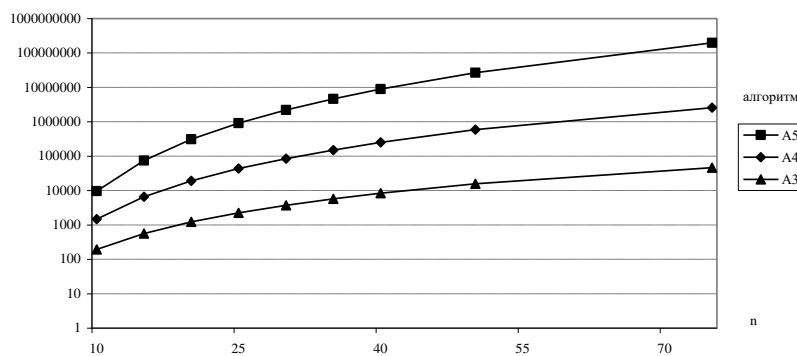
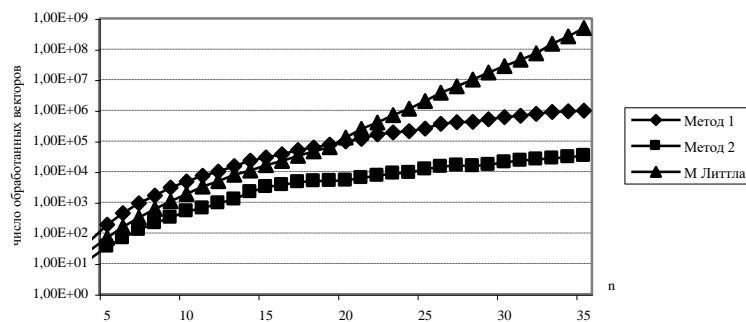


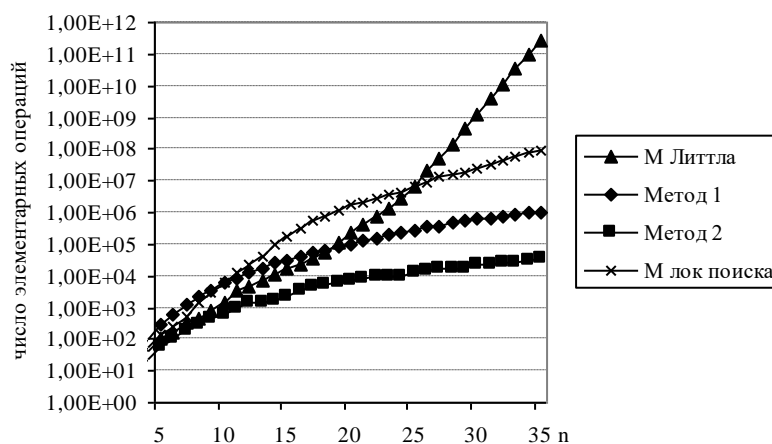
Рис. 7.13. Залежність числа оброблюваних векторів від розмірності розв'язуваної задачі лінійного й квадратичного програмування для алгоритмів A_5 , A_4 , A_3

Експериментальне дослідження алгоритмів розв'язуваної задачі визначення найкоротших гамільтонових шляхів

При експериментальному порівнянні розроблених алгоритмів з відомими вагами ребер графа генерувалися за рівномірним законом розподілу в діапазоні (0–50). Для одержання середнього значення кожної точки графіків усіх аналізованих характеристик розв'язувалося по 200 тестових задач, усі результати статистичного аналізу отримані з довірчою ймовірністю 0,95. На всіх рисунках розроблені алгоритми A_1 і A_2 позначені як метод 1 і метод 2. Як видно з графіків, наведених на рис. 7.14, при $n \geq 27$, алгоритми Літтла й локального пошуку мають істотно більш високу тимчасову складність у порівнянні з розробленими. При цьому, як показано у [85 – 88], алгоритми локального пошуку мають похибку, що лежить у діапазоні від 7 до 29 % і зростає зі збільшенням розмірності розв'язуваних задач.



а) Залежності числа оброблюваних векторів від розмірності розв'язуваної задачі



б) Залежності числа елементарних операцій від розмірності розв'язуваної задачі

Рис. 7.14

У розроблених алгоритмах, на основі ідей рангового підходу, похибка зі збільшенням розмірності задачі стабілізується й не перевищує 2% (рис. 7.15). З графіків, наведених на рис. 7.16, видно, що відсоток неточних рішень для алгоритму A_1 при $n \geq 27$ дуже низький. Експериментальне дослідження показало, що при $n \geq 27$ на 2000 тестових задач у середньому тільки 3 давали наблизений розв'язок і при цьому похибка не перевищувала 1–2%. Використання евристичного правила в алгоритмі A_2 дозволило істотно зменшити тимчасову складність цього алгоритму в порівнянні з A_1 , але при цьому зростає похибка і число неточних розв'язків у ньому зі збільшенням розмірності розв'язуваної задачі росте (рис. 7.16).

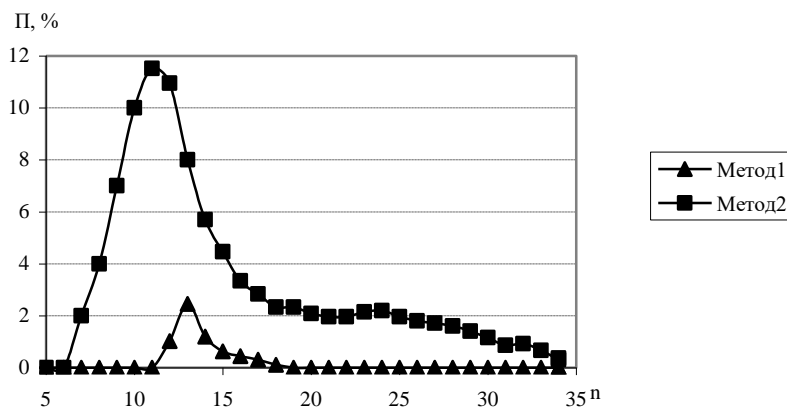


Рис.7.15. Залежність відносної похибки від розмірності розв'язуваної задачі

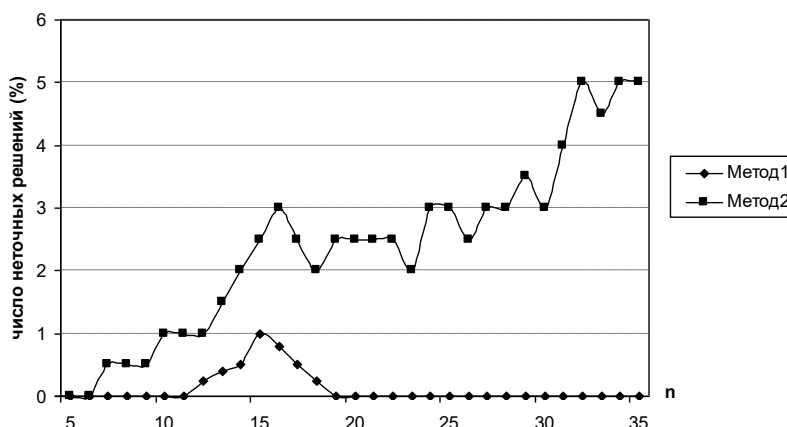


Рис. 7.16. Залежність числа неточних розв'язків, %, від розмірності, розв'язуваної задачі

Вправи

7.1. Поясніть, чому запропоновані в даному розділі узагальнені процедури можна розглядати як універсальну схему розв'язання довільних задач комбінаторної оптимізації.

7.2. Поясніть, як обчислюються довжини шляхів у графі при розв'язанні довільних задач нелінійного програмування.

7.3. Поясніть, чому зі збільшенням числа обмежень при розв'язанні задач нелінійного програмування похибка розв'язків зменшується.

7.4. Покажіть, що алгоритм визначення найкоротшого гамільтонового шляху, розглянутий у даному розділі, наближений.

7.5. Чому в узагальненій процедурі розв'язання довільних задач використовується стягнуте дерево всіх шляхів?

Висновки за розділами 6, 7

1. Важливою перевагою розглянутих алгоритмів на основі рангового підходу є той факт, що збільшення числа обмежень практично не впливає на похибку розв'язків алгоритмів, тоді як для методів розв'язання задач дискретної оптимізації, заснованих на ідеях методу галузей і меж, зростання числа обмежень до декількох сотень приводить фактично до неможливості їхнього практичного застосування.

2. Наближені алгоритми для розв'язання задач $(0,1)$ - рюкзак і ЗНП на основі рангового підходу мають властивість збіжності до точного розв'язку зі зростанням розмірності розв'язуваної задачі, що є важливою перевагою рангового підходу до розв'язання задач ЦЛП з булевими змінними у порівнянні з відомими алгоритмами.

3. Розглянуті алгоритми поліноміальної складності дозволяють одноманітно, з невеликою похибкою (яка зменшується з підвищенням числа обмежень) розв'язувати задачі лінійного та нелінійного булевого програмування й ефективно розпаралелюються.

БІБЛІОГРАФІЧНИЙ СПИСОК

1. Беллман Р. Динамическое программирование. – М.: ИИЛ, 1960. – 180 с.
2. Гуткин Л.С. Оптимизация радиоэлектронных устройств по совокупности показателей качества. – М.: Сов. радио, 1974. – 220 с.
3. Гуткин Л.С. Проблемы оптимизации радиосистем // Радиотехника. – 1971. – № 3. – С. 35-48.
4. Альошин Г.В. Оцінка якості інформаційно-вимірювальних систем. – Харків: УкрДАЗТ, 2008. – 288 с.
5. Алешин Г.В., Богданов Ю.А. Эффективность сложных радиотехнических систем. – К.: Наукова думка, 2008. – 288 с.
6. Крамер Г. Математические методы статистики. – М.: ИИЛ, 1948. – 120 с.
7. Амиантов И.Н. Теория связи. – М.: Сов. радио, 1971. – 210 с.
8. Бакут П.А., Большаков И.А. Вопросы статистической теории радиолокации. – М.: Сов. радио, 1964. – Т. 1, 2. – 334 с.
9. Зюко А.Г. Помехоустойчивость и эффективность систем связи. – М.: Связь, 1972.
10. Космические траекторные измерения/ Под ред. П.А. Агаджанова и В.Е. Дулевича. – М.: Сов. радио; 1970. – 190 с.
11. Вакман Д.Е., Седлецкий Р.М. Вопросы синтеза радиолокационных сигналов. – М.: Сов. радио, 1973.
12. Винокуров В.И., Ваккер Р.А. Вопросы обработки сложных сигналов в корреляционных системах. – М.: Сов. радио, 1972. – 276 с.
13. Лезин Ю.С. Оптимальные фильтры и накопители импульсных сигналов. – М.: Сов. радио, 1969. – 153 с.
14. Окунев Ю.В., Плотников В.И. Принципы системного подхода к проектированию в технике связи. – М.: Связь, 1976. – 128 с.
15. Зуховицкий С.Н., Авдеева Л.И. Линейное и выпуклое программирование. – М.: Наука, 1967. – 212 с.

16. Свириденко С.С. Об искажениях псевдослучайного сигнала с ограниченным спектром на выходе фільтра // Радиотехника. – 1969. – № 8. – С. 48-51.

17. Евтянов С.Н. Переходные процессы в приемо-усилительных схемах. – М.: Связьиздат, 1948. – 135 с.

18. Об оптимальных параметрах и экономической эффективности многостанционной системы спутниковой связи / Н.В. Талызин, Л.Я. Кантор, Е.А. Манякин, Ю.М. Паянский // Радиотехника. – 1969. – № 11. – С. 23-34.

19. Чупик С.И. О распределении требований к точности измерительных устройств многопараметрических систем при учете их стоимости и надежности // Эффективность обработки информации в системах траекторных измерений. – М.: МО СССР, 1968. – 87с.

20. Консон А.С. Экономические расчеты в приборостроении. – М.: Высшая школа, 1973. – 134 с.

21. Вудворд Ф.М. Теория вероятности и теория информации с применением в радиолокации. – М.: Сов. радио, 1968. – 348с.

22. Вагнер Г. Основы исследования операций. – М.: Мир, 1973. – Т.2. – 231 с.

23. Зайченко Ю.П. Исследование операций. – К.: Вища школа, 1988. – 552 с.

24. Калинин В.Н., Резников Б.А., Варакин Е.И. Теория систем и оптимального управления. Ч. 2. Понятия, модели и алгоритмы оптимального выбора. – М.: МО СССР, 1987. – 590 с.

25. Кофман А., Анри-Лабродер А. Методы и модели исследования операций. Целочисленное программирование. – М.: Мир, 1977. – 236 с.

26. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность: Пер. с англ. – М.: Мир, 1985. – 512 с.

27. Рейнгольд Э., Нивергельт Ю., Део Н. Комбинаторные алгоритмы. Теория и практика. – М.: Мир, 1980. – 476 с.

28. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления. – М.: Наука, 1969. – 112 с.

29. Емеличев В.А., Комлик В.И. Метод построения последовательности планов для решения задач дискретной оптимизации. – М.: Наука, 1981. – 208 с.

30. Журавлев Ю.И., Финкельштейн Ю.Ю. Сфера применения методов дискретного программирования // Применение исследования операций в экономике. – М.: Экономика, 1977. – С. 26–59.

31. Михалевич В.С., Кукса А.И. Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов. – М.: Наука, 1983. – 208 с.

32. Solution method on the basis of rank approach for integer linear problems with boolean variables // S.V.Listrovoy, D.Yu.Golubnichiy, E.S. Listrovaya Engineering Simulation. – 1999. – vol.16. – P. 707–725.

33. Parallel algorithms of calculation process optimization for the boolean programming problems // S.V.Listrovoy, V.F.Tretjak, A.S.Listrovaya - Engineering Simulation. – 1999. – vol.16. – P. 569 – 579.

34. Листровой С.В. Параллельный алгоритм для решения задачи линейного программирования с булевыми переменными // Электронное моделирование. – 1991. – Т 13, № 3. – С. 29–32.

35. Листровой С.В., Гуль А.Ю. Метод решения задачи о минимальном покрытии на основе рангового подхода // Электрон. моделирование. – 1999. – № 1. – С. 58–70.

36. Танаев В.С., Гордон В.С., Шафранский Я.М. Теория расписаний одностадийные системы. – М :Наука, 1984. – 381 с.

37. Ху Т. Целочисленное программирование и потоки в сетях. – М: Мир, 1974. – 365 с.

38. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов: Пер. с англ. – М.: Мир, 1979. – 536 с.

39. Филлипс Д., Гарсиа-Диас А. Методы анализа сетей: Пер. с англ. – М.: Мир, 1984. – 496 с.

40. Листровой С.В., Симашкевич О.Н. Об использовании гарантированных прогнозов в методах решения задач булевого программирования на основе рангового подхода // Электрон. моделирование. – 2003. – Т. 25, № 4. – С. 89–103.