Я.М. ГРИГОРЕНКО
Н.Д. ПАНКРАТОВА

ОБЧИСЛЮВАЛЬНІ МЕТОДИ

В ЗАДАЧАХ ПРИКЛАДНОЇ МАТЕМАТИКИ

Я.М. ГРИГОРЕНКО
Н.Д. ПАНКРАТОВА

Обчислювальні методи в задачах прикладної математики

Затверджено Міністерством освіти України як навчальний посібник для студентів вищих навчальних закладів, які навчаються за спеціальністю «Прикладна математика»

КИЇВ
«ЛИБІДЬ»
1995

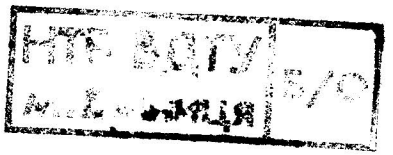
ББК 22.19я73
Г83
УДК 519.95

Рецензенти: д-р фіз.-мат. наук, проф. А.Т.Василенко (Інститут механіки НАН України), д-р фіз.-мат. наук, проф. А.А.Глущенко (Київський університет)

Головна редакція літератури з природничих та технічних наук

Головний редактор Е.О.Вавілова

Редактор О.М.Миронець



Г83 Григоренко Я.М., Панкратова Н.Д.
Обчислювальні методи в задачах прикладної математики: Навч. посібник. — К.: Либідь, 1995. — 280 с.
ISBN 5-325-00486-7.

У навчальному посібнику викладено обчислювальні методи, які використовуються при розв'язанні на ЕОМ задач прикладної математики. Послідовно розглядаються деякі методи інтерполювання, розв'язання систем лінійних та нелінійних рівнянь, обчислення власних значень матриць, задач Коші, лінійних та нелінійних крайових задач для звичайних диференціальних рівнянь та диференціальних рівнянь в частинних похідних. Методи розв'язання задач прикладної математики супроводжуються конкретними прикладами, які ілюструють особливості їх застосування до розв'язання конкретних задач.

Для студентів технічних вузів та університетів, а також спеціалістів з прикладної математики, які займаються розв'язанням науково-технічних задач.

Г $\frac{1602120000-039}{224-95}$ Без оголошення

ББК 22.19я73

ISBN 5-325-00486-7

ЗМІСТ

Передмова 5

Глава 1. Інтерполяція функцій

§ 1.1. Постановка задачі інтерполяції 7
§ 1.2. Скінченні та поділені різниці 9
§ 1.3. Інтерполяційна формула Ньютона 12
§ 1.4. Інтерполяційні поліноми Гаусса, Стірлінга та Бесселя 19
§ 1.5. Інтерполяційна формула Лагранжа 23
§ 1.6. Інтерполяційний многочлен Ерміта 27
§ 1.7. Збіжність інтерполяції 31
§ 1.8. Інтерполяція сплайнами 32

Глава 2. Методи розв'язання систем лінійних алгебраїчних рівнянь

§ 2.1. Загальна характеристика методів 40
§ 2.2. Метод Гаусса та його модифікації 41
§ 2.3. Зв'язок методу Гаусса з розкладанням матриці на множники 49
§ 2.4. Метод прогонки 52
§ 2.5. Схема Халецького 53
§ 2.6. Метод квадратного кореня 55
§ 2.7. Обчислювання визначника та оберненої матриці 58
§ 2.8. Загальні зауваження про побудову ітераційних процесів 60
§ 2.9. Метод простої ітерації 62
§ 2.10. Метод Зейделя 67

Глава 3. Розв'язання нелінійних систем рівнянь

§ 3.1. Вступні зауваження. Поняття про принцип стискаючих відображень 71
§ 3.2. Метод простої ітерації. Метод Зейделя 74
§ 3.3. Метод Ньютона розв'язання систем нелінійних рівнянь 77
§ 3.4. Методи квазіньютонівського типу 82
§ 3.5. Метод покоординатного спуску 86
§ 3.6. Метод продовження розв'язання по параметру 90

Глава 4. Обчислення власних значень та власних векторів матриць 94

§ 4.1. Постановка задачі. Загальна характеристика методів 94
§ 4.2. Метод Данилевського 96
§ 4.3. Метод Левер'є 102
§ 4.4. Зведення симетричної матриці до тридіагонального вигляду 104
§ 4.5. Ітераційний метод Якобі для симетричних матриць 108
§ 4.6. QR-алгоритм 114

§ 4.7. Метод обернених ітерацій 118
§ 4.8. Розв'язання часткової проблеми власних значень 120

Глава 5. Методи розв'язання задач Коші для звичайних диференціальних рівнянь 126

§ 5.1. Вступні зауваження. Постановка задачі 126
§ 5.2. Загальні зауваження щодо оцінки похибки при розв'язанні задач Коші 129
§ 5.3. Метод Ейлера і його модифікації 131
§ 5.4. Метод Рунге-Кутта 135
§ 5.5. Багатокрокові методи 141
§ 5.6. Поняття стійкості й жорстких рівнянь при розв'язанні задач 153

Глава 6. Методи розв'язання лінійних крайових задач для звичайних диференціальних рівнянь 162

§ 6.1. Постановка задачі. Зведення до задачі Коші 162
§ 6.2. Метод диференціальної прогонки 167
§ 6.3. Метод різницевої прогонки 174
§ 6.4. Метод дискретної ортогоналізації 178
§ 6.5. Методи Рітца і Бубнова-Гальоркіна 183
§ 6.6. Метод сплайн-колокації 187

Глава 7. Методи розв'язання нелінійних крайових задач для звичайних диференціальних рівнянь 194

§ 7.1. Вступні зауваження. Постановка задачі 194
§ 7.2. Метод зведення нелінійної крайової задачі до системи нелінійних рівнянь і задачі Коші 195
§ 7.3. Метод лінеаризації 201
§ 7.4. Метод продовження розв'язання по параметру 207
§ 7.5. Метод скінченних різниць 214
§ 7.6. Метод Рітца 217

Глава 8. Методи розв'язання крайових задач для диференціальних рівнянь у частинних похідних 220

§ 8.1. Вступні зауваження. Постановка основних задач 220
§ 8.2. Рівняння параболічного типу. Явні та неявні скінченнорізницеві методи 222
§ 8.3. Метод прогонки для рівняння теплопровідності 235
§ 8.4. Рівняння гіперболічного типу. Метод сіток 237
§ 8.5. Скінченнорізницевий метод розв'язання рівнянь еліптичного типу 242
§ 8.6. Поняття про метод прямих розв'язання граничних задач 250
§ 8.7. Метод Рітца 252
§ 8.8. Метод Бубнова—Гальоркіна 254
§ 8.9. Метод Власова—Канторовича 256
§ 8.10. Метод, що базується на сплайн-апроксимації функцій в одному напрямі 260
§ 8.11. Метод, що базується на апроксимації дискретними рядами Фур'е 272

Контрольні запитання та завдання 277

Список рекомендованої літератури 278

ПЕРЕДМОВА

При вирішенні багатьох проблем математичної фізики та сучасної техніки виникають задачі, пов'язані з прикладною математикою. Широке застосування комп'ютерів у практиці розв'язання прикладних задач викликало необхідність у розробці більш ефективних та удосконалених обчислювальних методів. Викладенню підходів використання обчислювальних методів у задачах прикладної математики, що описуються лінійними та нелінійними алгебраїчними і диференціальними рівняннями в частинах або звичайних похідних, і присвячена ця книга. Основна увага приділяється не теоретичним, а обчислювальним аспектам цих методів.

Книга написана на фізичному рівні строгості, що, на думку авторів, повністю відповідає точності, яка потрібна при розв'язанні прикладних задач. У відповідності з цим деякі оцінки точності одержуваних розв'язків даються за допомогою індуктивних способів. Майже всі методи, що викладаються, проілюстровано багатьма прикладами.

У першій главі розглядається задача інтерполяції функцій. Наводяться загальновідомі інтерполяційні формули Ньютона, Бесселя, Гаусса, Стірлінга, Лагранжа, Ерміта з означеннями областей їх застосування. Значна увага приділяється інтерполюванню сплайнами, які інтенсивно використовуються в задачах прикладної математики як більш природний апарат наближення функцій.

Розв'язанню систем лінійних алгебраїчних рівнянь із залученням прямих та ітераційних методів присвячена друга глава, в якій наведені прямі методи Гаусса та його модифікації, метод квадратного кореня, схема Халецького. Показано можливість обчислення визначника та оберненої матриці на основі використання прямих методів. Обговорюються загальні зауваження щодо побудови ітераційних методів розв'язання систем лінійних алгебраїчних рівнянь. Подано алгоритми чисельної реалізації методів простої ітерації і Зейделя.

Методи розв'язання нелінійних систем рівнянь розглядаються в третій главі, де дається поняття про принцип стиска-

ючих відображень, методи простої ітерації, Зейделя, Ньютона. Значна увага приділяється методам квазіньютонівського типу, покоординатного спуску та методу продовження розв'язку по параметру.

Обчислення власних значень і власних векторів матриць подано в четвертій главі. Показано можливість розв'язання повної проблеми власних значень за допомогою методів Данилевського, Лавер'є, прямого та ітераційного методів Якобі для симетричної матриці, методу QR-алгоритму для несиметричної матриці. Проблема часткових власних значень розв'язується з використанням класичного степеневого методу і методу скалярних добутків.

П'ята глава присвячена викладенню методів розв'язання задач Коші для звичайних диференціальних рівнянь. Тут поряд з традиційними методами розв'язання задач Коші, такими як методи Ейлера, Рунге—Кутта, Адамса, також розглядаються методи розв'язання задачі Коші для жорстких рівнянь.

У шостій главі наведені методи розв'язання лінійних крайових задач для систем звичайних диференціальних рівнянь, методи диференціальної та різницевої прогонки на базі зведення до задачі Коші, метод дискретної ортогоналізації, методи Рітца та Бубнова—Гальоркіна і метод сплайн-колокації.

Методи розв'язання нелінійних крайових задач для звичайних диференціальних рівнянь розглядаються в сьомій главі. Тут дається постановка задачі для системи нелінійних звичайних диференціальних рівнянь. Для розв'язання розглядуваного класу задач пропонуються методи зведення нелінійної крайової задачі до систем рівнянь і задачі Коші, лінеаризації, продовження розв'язання по параметру, скінченних різниць і Рітца.

Методи розв'язання диференціальних рівнянь в частинних похідних викладено у восьмій главі. Тут наводяться явні та неявні скінченнорізницеві методи розв'язання задач для рівнянь параболічного, гіперболічного та еліптичного типів. Показано, що розв'язання диференціального рівняння еліптичного типу можна знайти за допомогою методу прямих, методів Рітца, Бубнова—Гальоркіна, Власова—Канторовича. Розглядається метод, що базується на сплайн-апроксимації функцій в одному напрямі, який дозволяє враховувати довільні граничні умови, а також метод, що базується на апроксимації функцій дискретними рядами Фур'є.

При написанні навчального посібника використано багаторічний досвід викладання авторами відповідних курсів у Київському університеті та Київському політехнічному інституті та результати власних розробок.

ГЛАВА 1

ІНТЕРПОЛЯЦІЯ ФУНКЦІЙ

§ 1.1. ПОСТАНОВКА ЗАДАЧІ ІНТЕРПОЛЯЦІЇ

Інтерполяція функцій належить до задач обчислювальної математики. Часто потрібно побудувати функцію $f(x)$ для всіх значень x на інтервалі $a \le x \le b$, якщо відомі її значення у деякому скінченному числі точок цього інтервалу. Ці значення можуть бути знайдені на підставі реального експерименту або шляхом обчислень. Крім того, може статись, що функція задається формулою, і обчислення її значень за нею дуже трудомістке. Тому бажано мати для функції більш просту формулу, яка надала б можливість знаходити її приблизне значення з потрібною точністю у будь-якій точці відрізка.

Задача наближення функцій виникає під час створення стандартних програм обчислення елементарних та спеціальних функцій, властивості яких дозволяють значно зменшити кількість обчислень. При цьому розглядаються всі функції $g(x)$, програма обчислення яких розміщується у k комірках пам'яті ЕОМ, таких що деяка норма похибки $\|f - g\|$ не перевищує ε. Серед усіх таких функцій потрібно вибрати ту, обчислення якої потребує мінімальних витрат часу ЕОМ. Залежно від конкретної постановки задачі норма може бути вибрана по-різному, зокрема у випадках для відрізка $[a, b]$, на якому наближається функція, вибираємо

$$\|f\| = \sup_{[a,b]} |f|.$$

Для підвищення потрібної точності в окремих точках слід ввести множник $p(x)$, що забезпечує мализну відносної похибки. Наприклад, одна із стандартних функцій $\sin(x)$ забезпечує мализну похибки у нормі

$$\|f\| = \sup |p(x)f(x)|;$$

$$p(x) = \min (10^{19}, x^{-1}).$$

Вигляд функції, що наближається, суттєво залежить від мети наближення. Нехай із потрібною точністю функція може бути подана многочленом десятого степеня, або виразом

$$a_1 \cos \beta_1 x + a_2 \cos \beta_2 x.$$

Якщо знайдене наближення використовується у теоретичних дослідженнях, для розв'язання задачі на моделюючому пристрої або у технічному процесі, то друга форма запису більш зручна. У випадку обчислення

значень функції на ЕОМ друга форма потребує більшої кількості арифметичних операцій.

Апарат інтерполяції поліномами широко використовується у чисельному аналізі, на основі якого будується багато методів розв'язання ряду задач, зокрема лінійних та нелінійних крайових задач математичної фізики. Застосування процесу інтерполяції надає можливість розширити розв'язання ряду лінійних крайових задач на випадки з більш складними граничними умовами. При розв'язанні нелінійних задач процес інтерполяції використовується під час передачі інформації від попереднього наближення до наступного.

Математична задача інтерполяції функцій формулюється таким чином: нехай на відрізку $[a,b]$ задано точки $x_0, x_1, .., x_n$, які називають *вузлами інтерполяції*, і значення деякої функції $f(x)$ у цих точках

$$f(x_0) = y_0, f(x_1) = y_1, \dots, f(x_n) = y_n.$$

Потрібно побудувати функцію $F(x)$ (інтерполююча функція), що належить деякому класу і набуває у вузлах інтерполяції тих самих значень, що й $f(x)$, тобто таку, що

$$F(x_0) = y_0, F(x_1) = y_1, \dots, F(x_n) = y_n. \qquad (1.1)$$

Геометрично це означає, що потрібно знайти криву $y = F(x)$, яка проходить через задану систему точок $M_i(x_i, y_i)$.

У такій загальній постановці задача може мати нескінченну кількість розв'язань, або зовсім не мати їх. У деяких випадках, якщо функція $f(x)$, що інтерполюється, періодична, то за клас $F(x)$ можна взяти тригонометричні многочлени; якщо ж вона обертається на нескінченність у заданих точках або поблизу них, то за клас $F(x)$ доцільно взяти клас раціональних функцій. Задача інтерполяції стає однозначною, якщо замість довільної функції $F(x)$ шукати поліном $P_n(x)$ степеня n, що задовольняє умови (1.1), тобто такий, що

$$P_n(x_0) = y_0, \ P_n(x_1) = y_1, \dots, P_n(x_n) = y_n. \qquad (1.2)$$

Отриману інтерполяційну формулу, реалізовану у економічному алгоритмі, звичайно використовують для наближеного обчислення значень даної функції $f(x)$ для значень аргументу x, відмінних від вузлів інтерполяції. Таку операцію називають *інтерполяцією функції* $f(x)$. При цьому розрізняють саме інтерполяцію, коли $x \in [x_0, x_n]$, та екстраполяцію, коли $x \notin [x_0, x_n]$. У подальшому під терміном *інтерполювання* будемо розуміти обидві операції.

Таким чином, загальна задача полягає у побудові інтерполяційного многочлена

$$f(x) = P_n(x) \qquad (1.3)$$

та в оцінці похибки загальної інтерполяційної формули (1.3).

Зазначимо, що двох різних інтерполяційних поліномів степеня n існувати не може. Насправді, припустивши зворотне, дійдемо висновку, що різниця двох таких поліномів, що є поліномом n-го степеня, має $n + 1$ нулів, отже, тотожно дорівнює нулеві.

Існує велика кількість різних засобів розв'язання задачі інтерполяції. Різноманітність методів зумовлюється різноманітністю різних постановок проблеми, що виникає під час розв'язання практичних задач.

§ 1.2. СКІНЧЕННІ ТА ПОДІЛЕНІ РІЗНИЦІ

Скінченні різниці використовуються при вивченні функцій, що задаються таблицею значень у рівновіддалених вузлах, та при обчисленні таких функцій.

Нехай $y = f(x)$ — задана функція. Позначимо через $\Delta x = h$ фіксовану величину приросту аргументу (крок). Тоді вираз

$$\Delta y \equiv \Delta f(x) = f(x + \Delta x) - f(x) \tag{1.4}$$

називається *скінченною різницею першого порядку* функції y. Аналогічно визначаються скінченні різниці вищих порядків

$$\Delta^n y = \Delta(\Delta^{n-1} y) \quad (n = 2, 3, \ldots).$$

Наприклад,

$$\Delta^2 y = \Delta[f(x + \Delta x) - f(x)] = [f(x + 2\Delta x) - f(x + \Delta x)] -$$

$$- [f(x + \Delta x) - f(x)] = f(x + 2\Delta x) - 2f(x + \Delta x) + f(x).$$

Має місце твердження: якщо

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

— поліном n-го степеня, то

$$\Delta^n P_n(x) = n! a_0 h^n = \text{const}. \tag{1.5}$$

Як висновок із (1.5) дістанемо

$$\Delta^s P_n(x) = 0 \text{ при } s > n.$$

Символ Δ можна розглядати як оператор, що зіставляє функції $y = f(x)$ функцію $\Delta y = f(x + \Delta x) - f(x)$.

Нехай функція $f(x)$ має неперервну похідну $f^{(n)}(x)$ на відрізку $[x, x + n\,\Delta x]$. Тоді правильна формула

$$\Delta^n f(x) = (\Delta x)^n f^{(n)}(x + \theta\, n\, \Delta x), \tag{1.6}$$

де $0 < \theta < 1$. Перейдемо до границі при $\Delta x \to 0$; отримаємо

$$f^{(n)}(x) = \lim_{\Delta x \to 0} \frac{\Delta^n f(x)}{(\Delta x)^n}.$$

Таким чином, при малих Δx правильна наближена формула

$$f^{(n)}(x) \approx \frac{\Delta^n f(x)}{(\Delta x)^n}. \tag{1.7}$$

Наведемо деякі властивості скінченних різниць.

1. Скінченна різниця довільного порядку може бути виражена через значення функції; зокрема, для скінченної різниці другого порядку маємо

$$\Delta^2 f_k = f_{k+2} - 2f_{k+1} + f_k.$$

2. Властивості скінченних різниць аналогічні властивостям похідних, наприклад, лінійність

$$\Delta^m(c_1 f \pm c_2 g)_k = c_1 \Delta^m f_k \pm c_2 \Delta^m g_k.$$

3. Скінченні різниці першого порядку від многочлена степеня n є многочленами $(n-1)$-го степеня, тобто $\Delta(P_n) = P_{n-1}$, а скінченні різниці n-го порядку від многочлена степеня n постійні всличини

$$\Delta^n(P_n) = \text{const}, \quad \Delta^{n+1}(P_n) = 0.$$

Скінченні різниці гладкої функції змінюються плавно і наявність при цьому різких відхилень вказує на помилку при обчисленні значень функцій.

При обчисленні різниць з використанням наближених значень функцій похибка різниць збільшується. Так, якщо похибка табличних значень дорівнює половині одиниці останнього розряду, похибка скінченної різниці першого порядку дорівнюватиме одиниці останнього розряду, похибка скінченної різниці другого порядку — двом одиницям останнього розряду і, нарешті, похибка скінченної різниці m-го порядку — 2^m одиницям останнього розряду. Із цього випливає, що при $m \to \infty$ абсолютна похибка скінченних різниць збільшується, і тому скінченні різниці вищих порядків при обчисленні не використовуються.

Різницеві відношення, які також називають *поділеними різницями* функції, використовуються при обчисленні та при вивченні функцій у випадку, коли останні задаються на довільній системі значень аргументу. Припустимо, що для деяких різних між собою значень аргументу $x_0, x_1, x_2, \ldots$ задано значення функції $f(x_0), f(x_1), f(x_2), \ldots$. *Поділеними різницями першого порядку* називають величини

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \quad f(x_1, x_2) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \quad \ldots$$

Поділені різниці першого порядку мають сенс середніх швидкостей зміни функції f на відрізках (x_0, x_1), (x_1, x_2), ... Вони використовуються у побудові поділених різниць другого порядку

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0};$$

$$f(x_1, x_2, x_3) = \frac{f(x_2, x_3) - f(x_1, x_2)}{x_3 - x_1}, \ldots$$

Поділені різниці порядку $n + 1$ $(n = 2,...)$ визначаються за допомогою поділених різниць попереднього порядку n за формулою

$$f(x_0, x_1, \ldots, x_n, x_{n+1}) = \frac{f(x_1, x_2, \ldots, x_{n+1}) - f(x_0, x_1, \ldots, x_n)}{x_{n+1} - x_0}.$$

Неважко отримати прості вирази поділених різниць усіх порядків через значення функції. Дійсно, за визначенням поділених різниць першого порядку

$$f(x_0, x_1) = \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0}.$$

Тому для поділених різниць другого порядку дістанемо

$$f(x_0, x_1, x_2) = \frac{1}{x_2 - x_0}[f(x_1, x_2) - f(x_0, x_1)] =$$

$$= \frac{1}{x_2 - x_0}\left[\frac{f(x_1)}{x_1 - x_2} + \frac{f(x_2)}{x_2 - x_1} - \frac{f(x_0)}{x_0 - x_1} - \frac{f(x_1)}{x_1 - x_0}\right] =$$

$$= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)}.$$

За індукцією можна довести, що при будь-якому n правильна рівність

$$f(x_0, x_1, \ldots, x_n) =$$

$$= \sum_{i=0}^{n} \frac{f_i(x)}{(x_i - x_0)(x_i - x_1)\ldots(x_i - x_{i-1})(x_i - x_{i+1})\ldots(x_i - x_n)} =$$

$$= \sum_{i=0}^{n} \frac{f_i(x)}{\omega'(x_i)}, \ \omega(x_i) = \prod_{j=0}^{n} (x - x_j).$$

Неважко зробити висновок, що поділені різниці симетричні відносно своїх вузлів.

Якщо функція $f(x)$ неперервна зі своїми похідними до n-го порядку на інтервалі (a, b), то виконується рівність

$$f(x_0, x_1, \dots, x_n) = \frac{1}{n!} f^{(n)}(\xi), \quad \xi \in [a, b]. \tag{1.9}$$

У випадку рівновіддалених вузлів, тобто коли $x_k = x_0 + kh$ та $f(x_k) = f(x_0 + kh)$, можна простежити зв'язок між скінченними та поділеними різницями. Для різниць першого порядку

$$f(x_0, x_1) = f(x_0, x_0 + h) = \frac{f(x_0 + h) - f(x_0)}{x_0 + h - x_0} = \frac{\Delta y_0}{1!h}.$$

Аналогічно для різниць другого порядку

$$f(x_0, x_1, x_2) = f(x_0, x_0 + h, x_0 + 2h) = \frac{1}{2!h}\left[\frac{\Delta y_1}{1!h} - \frac{\Delta y_0}{1!h}\right] = \frac{\Delta^2 y_0}{2!h^2}$$

та для n-го порядку

$$f(x_0, x_1, \dots, x_n) = \frac{\Delta^n y_0}{n!h^n}. \tag{1.10}$$

Із формул (1.9) та (1.10) випливає

$$f^{(n)}(\xi)h^n = \Delta^n y_0; \tag{1.11}$$

Отже, можна говорити про мализну скінченної різниці через мализну h.

§ 1.3. ІНТЕРПОЛЯЦІЙНА ФОРМУЛА НЬЮТОНА

Нехай для функцій $y = f(x)$ задані значення $y_i = f(x_i)$ для рівновіддалених значень незалежної змінної величини $x_i = x_0 + ih$ $(i = 0, 1, \dots, n)$, де h — крок інтерполяції. Потрібно підібрати многочлен $P_n(x)$ степеня не вище n, який у точках x_i набуває значення $P_n(x_i) = y_i$ $(i = 0, 1, \dots, n)$. Ці умови еквівалентні тому, що

$$\Delta^m P_n(x) = \Delta^m y_0, \quad m = 0, 1, 2, \dots, n.$$

Будуємо першу поділену різницю

$$f(x, x_0) = \frac{f(x) - f(x_0)}{x - x_0}. \tag{1.12}$$

Звідки $f(x) = f(x_0) + f(x, x_0)(x - x_0)$.

Будуємо другу поділену різницю

$$f(x, x_0, x_1) = \frac{f(x, x_0) - f(x_0, x_1)}{x - x_1},$$

і, отже, $f(x, x_0) = f(x_0, x_1) + (x - x_1)f(x, x_0, x_1)$. Підставимо цей вираз у (1.12) і дістанемо

$$f(x) = y_0 + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x, x_0, x_1). \qquad (1.13)$$

Тепер із рівності

$$f(x, x_0, x_1, x_2) = \frac{f(x, x_0, x_1) - f(x_0, x_1, x_2)}{x - x_2}$$

знаходимо

$$f(x, x_0, x_1) = f(x_0, x_1, x_2) + (x - x_2)f(x, x_0, x_1, x_2).$$

Підставляючи цей вираз у (1.13), матимемо

$$f(x) = y_0 + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) +$$
$$+ (x - x_0)(x - x_1)(x - x_2)f(x, x_0, x_1, x_2).$$

Якщо ми продовжуватимемо цей процес підстановки, то прийдемо до рівності

$$f(x) = y_0 + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + \dots$$
$$+ (x - x_0)(x - x_1) \dots (x - x_{n-1})f(x_0, x_1, \dots, x_n) +$$
$$+ (x - x_0)(x - x_1) \dots (x - x_n)f(x, x_0, \dots, x_n). \qquad (1.14)$$

Сукупність членів правої частини формули (1.14) без останнього члена є многочленом n-го степеня. Запишемо формулу (1.14) у вигляді

$$f(x) = P_n(x) + (x - x_0)(x - x_1)(x - x_n)f(x, x_0, \dots, x_n)$$

і вважатимемо у ній послідовно $x = x_0, x_1, \dots, x_n$, дістанемо $n + 1$ рівність

$$f(x_k) = P_n(x_k) = y_k, \quad k = 0, 1, 2, \dots, n.$$

Із цього випливає, що вираз

$$P_n(x) = y_0 + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + \dots$$
$$+ (x - x_0)(x - x_1) \dots (x - x_{n-1})f(x_0, x_1, \dots, x_n) \qquad (1.15)$$

є інтерполяційним многочленом для функції $f(x)$, побудованим по $n + 1$ вузлах $x_0, x_1, \dots, x_n$. Це й є інтерполяційний многочлен Ньютона.

Поділені різниці, які стоять у правій частині рівності (1.15), можна виразити через скінченні різниці функції $y = f(x)$ за допомогою співвідношення (1.10). Тоді вираз (1.15) запишеться у вигляді

$$P_n(x) = y_0 + \frac{\Delta y_0}{1!h}(x - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1) + \dots +$$

$$+\frac{\Delta^n y_0}{n!h^n}(x-x_0)(x-x_1)\dots(x-x_{n-1}). \qquad (1.16)$$

Отже, ми дістали перший інтерполяційний многочлен Ньютона, який повністю задовольняє вимоги поставленої задачі, тобто степінь многочлена не перевищує степеня n, і також $P_n(x_i) = y_i$. При $h \to 0$ вираз (1.16) перетворюється на формулу Тейлора для функції y.

Дійсно,

$$\lim_{h\to 0}\frac{\Delta^k y_0}{h^k} = \left(\frac{d^k y}{dx^k}\right)_{x=x_0} = y^{(k)}(x_0);$$

$$\lim_{h\to 0}(x-x_0)\dots(x-x_n) = (x-x_0)^n.$$

Звідси при $h \to 0$ формула (1.16) набуває вигляду многочлена Тейлора

$$P_n(x) = y(x_0) + y'(x_0)(x-x_0) + \dots + \frac{y^{(n)}(x_0)}{n!}(x-x_0)^n.$$

Для практичного використання інтерполяційну формулу Ньютона (1.16) записують у дещо перетвореному вигляді. Для цього введемо нову змінну величину t, поклавши $x = x_0 + th$, де $t = \dfrac{x-x_0}{h}$ — кількість кроків h, необхідних для досягнення точки x із точки x_0:

$$x - x_0 = th, \quad (x-x_0)(x-x_0-h) = t(t-1)h^2, \dots$$

Після внесення вказаних величин у вираз для $P_n(x)$ отримаємо першу інтерполяційну формулу Ньютона для інтерполювання вперед, тобто поблизу початку таблиці:

$$P_n(x) = y_0 + \frac{t}{1!}\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + \dots + \frac{t(t-1)\dots(t-n+1)}{n!}\Delta^n y_0. \qquad (1.17)$$

Якщо у (1.17) покласти $n = 1$, то дістанемо формулу лінійного інтерполювання

$$P_1(x) = y_0 + t\Delta y_0.$$

При $n = 2$ будемо мати формулу параболічного або квадратичного інтерполювання

$$P_2(x) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2}\Delta^2 y_0.$$

Якщо надана нескінченна таблиця функції y, то число n у інтерполяційній формулі (1.17) може бути будь-яким. Практично у даному випадку число n вибирають так, щоб $\Delta^n y_i$ була сталою із заданою точністю.

За початкове значення x_0 можна приймати будь-яке табличне значення аргументу x.

Якщо надана скінченна таблиця значень функції y, то число n обмежено, а саме: n не може бути більше числа значення функції y, зменшеного на одиницю.

Припустимо , що точка інтерполяції розташована поблизу скінченної точки x_n таблиці. У цьому випадку вузли інтерполяції слід брати у порядку x_n, $x_n - h$, $x_n - 2h$, ... Формула Ньютона інтерполювання назад тоді матиме вигляд

$$P_n(x) = f(x_n) + f(x_n, x_{n-1})(x - x_n) + f(x_n, x_{n-1}, x_{n-2})(x - x_n)(x - x_{n-1}) + \\ + \ldots + f(x_n, x_{n-1}, \ldots, x_0)(x - x_n)(x - x_{n-1}) \ldots (x - x_1). \qquad (1.18)$$

Поділені різниці можна виразити через скінченні різниці, якщо скористатися можливістю переставляти у них аргументи та співвідношенням (1.10)

$$f(x_n) = y_n, f(x_n, x_{n-1}) = f(x_{n-1}, x_n) = \frac{\Delta y_{n-1}}{1!h};$$

$$f(x_n, x_{n-1}, x_{n-2}) = f(x_{n-2}, x_{n-1}, x_n) = \frac{\Delta^2 y_{n-2}}{2!h^2}, \ldots$$

Введемо змінну t, поклавши $x = x_n + th$, дістанемо для $f(x) = y(x)$ другу інтерполяційну формулу Ньютона для інтерполювання у кінці таблиці

$$P_n(x_n + th) = y_n + \frac{t}{1!}\Delta y_{n-1} + \frac{t(t+1)}{2!}\Delta^2 y_{n-2} + \ldots + \\ + \frac{t(t+1)\ldots(t+n-1)}{n!}\Delta^n y_0. \qquad (1.19)$$

Як перша, так і друга інтерполяційні формули Ньютона можуть бути використані для екстраполяції функції, тобто для знаходження значень функції y, значення аргументів x якої лежать поза таблицею. Якщо $x < x_0$ і значення x близьке до x_0, то вигідно використовувати перший інтерполяційний поліном Ньютона, причому тоді $t = \frac{x - x_n}{h}$ і $t > 0$. Таким чином, перша інтерполяційна формула Ньютона застосовується для інтерполювання вперед та екстраполювання назад, а друга — навпаки, для інтерполювання назад та екстраполювання вперед.

Зазначимо, що операція екстраполювання, взагалі кажучи, менш точна, ніж операція інтерполювання.

Інтерполяційні формули Ньютона вигідні, оскільки при додаванні m нових вузлів інтерполяції потрібні додаткові обчислення тільки для m нових членів, без зміни старих.

Замінюючи функцію $f(x)$ інтерполяційним многочленом $P_n(x)$, ми припускаємо похибку $R_n(x) = f(x) - P_n(x)$, яка називається *похибкою*

інтерполяції або, що те ж саме, *залишковим членом інтерполяційної формули*. Ясно, що у вузлах інтерполяції ця похибка дорівнює нулю.

Похибку інтерполяції можна подати у вигляді

$$R_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(r+1)!}(x - x_0)(x - x_1) \dots (x - x_n), \quad (1.20)$$

де $\xi \in [a, b]$ і залежить від x.

Звідси випливає оцінка

$$R_n(x) = |f(x) - P_n(x)| \le \frac{M_{n+1}}{(n+1)!}|\omega(x)|, \quad (1.21)$$

де $M_{n+1} = \sup\limits_{x \in [a,b]} |f^{(n+1)}(x)|$, $\omega(x) = (x - x_0)(x - x_1) \dots (x - x_n)$.

Для інтерполяційних многочленів Ньютона (1.17), (1.19) залишковий член записується відповідно у вигляді

$$R_n(x) = h^{n+1} \frac{t(t-1)(t-2) \dots (t-n)}{(n+1)!} y^{(n+1)}(\xi); \quad (1.22)$$

$$R_n(x) = h^{n+1} \frac{t(t+1)(t+2) \dots (t+n)}{(n+1)!} y^{(n+1)}(\xi), \quad (1.23)$$

де ξ — внутрішня точка найменшого інтервалу, що містить у собі всі вузли x_i $(i = 0, 1, \dots, n)$ і точку x.

Але оцінити абсолютне значення похідної високого порядку часто важко, а інколи й неможливо, наприклад, коли функція $f(x)$ задана таблицею і її аналітичний вираз невідомий. При практичних обчисленнях інтерполяційна формула Ньютона обривається на членах, що містять у собі такі різниці, які у границях заданої точності можна вважати сталими. При цьому похибка інтерполяції не перевищує одиниці молодшого розряду табличних значень.

Припускаючи, що Δ^{n+1} майже стала для функції $y = f(x)$ і величина h достатньо мала, і враховуючи, що

$$f^{(n+1)}(x) = \lim_{h \to 0} \frac{\Delta^{n+1} y}{h^{n+1}},$$

наближено можна покласти

$$f^{(n+1)}(\xi) \approx \frac{\Delta^{n+1} y_0}{h^{n+1}}.$$

У цьому випадку залишковий член (1.22) першої інтерполяційної формули Ньютона дорівнює

$$R_n(x) \approx \frac{t(t-1)(t-2) \dots (t-n)}{(n+1)!} \Delta^{n+1} y_0.$$

За цих же вимог для залишкового члена (1.23) другої інтерполяційної формули Ньютона дістанемо вираз

$$R_n(x) \approx \frac{t(t+1)(t+2)\dots(t+n)}{(n+1)!}\,\Delta^{n+1}y_n.$$

При побудові інтерполяційних многочленів часто виникає питання щодо найвигіднішого вибору вузлів інтерполяції. При невдалому розташуванні вузлів інтерполяції x_i верхня грань модуля похибки $R_n(x)$ (1.20) може бути надто великою. Оскільки у цій формулі $f^{(n+1)}(\xi)$ залежить від властивості функції $f(x)$ й не піддається регулюванню, постає задача про раціональний вибір вузлів інтерполяції x_i так, щоб поліном $\omega(x) = (x - x_0)(x - x_1)\dots(x - x_n)$ мав найменше максимальне значення за абсолютною величиною на інтервалі $[a, b]$. Ця задача розв'язується за допомогою многочлена Чебишева

$$T_{n+1}(x) = \frac{(b-a)^{n+1}}{2^{2n+1}} \cos\left[(n+1)\arccos\left(\frac{2x-(b+a)}{b-a}\right)\right], \qquad (1.24)$$

причому за вузли інтерполяції потрібно брати корені многочлена (1.24), тобто точки

$$x_k = \frac{a+b}{2} + \frac{b-a}{2}\cos\frac{(2k+1)\pi}{2(n+1)}, \quad k = 0, 1, \dots, n.$$

При цьому

$$\max_{x\in[a,b]} |\omega(x)| = \frac{(b-a)^{n+1}}{2^{2n+1}}$$

і оцінка (1.21) набуде вигляду

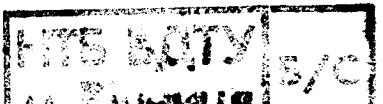
$$|f(x) - P_n(x)| \le \frac{M_{n+1}}{(n+1)!}\,\frac{(b-a)^{n+1}}{2^{2n+1}}. \qquad (1.25)$$

Приклад 1. За даною таблицею значень функції $y = \lg(x)$ (табл.1) знайти $\lg(1001)$.

Таблиця 1

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
1000	3,0000000	43214	− 426	8
1010	3,0043214	42788	− 418	9
1020	3,0086002	42370	− 409	8
1030	3,0128372	41961	− 401	
1040	3,0170333	41560		
1050	3,021893			

Розв'язання. Складаємо таблицю різниць (табл. 1). Зазначимо, що у стовпцях різниць не вказуються десяткові розряди. Записуючи значення функції у одиницях сьомого розряду, зазначаємо, що треті різниці практично сталі. Тому у формулі (1.17) достатньо покласти



$n = 3$. При цьому використовується верхній горизонтальний рядок таблиці різниць. Для $x = 1001$ маємо $t = \dfrac{x - x_0}{h}$ і $t = 0{,}1$. Таким чином,

$$\lg 1001 = 3{,}0000000 + 0{,}1 \cdot 0{,}0043214 + \frac{0{,}1 \cdot 0{,}9}{2} \cdot 0{,}0000426 +$$

$$+ \frac{0{,}1 \cdot 0{,}9 \cdot 1{,}9}{6} \cdot 0{,}0000008 = 3{,}0004341.$$

Оцінимо залишковий член. За формулою (1.22) при $n = 3$ маємо

$$R_3(x) = \frac{h^4 t(t-1)(t-2)(t-3)}{4!} f^{(4)}(\xi),$$

де $1000 < \xi < 1050$.

Оскільки $f(x) = \lg(x)$, то $f^{(4)}(x) = -\dfrac{3!}{x^4} \lg e$, через що $|f^{(4)}(\xi)| < \dfrac{3!}{(1000)^4} \lg e$. При $h = 10$ та $t = 0{,}1$ отримаємо

$$|R_3(1001)| < \frac{0{,}1 \cdot 0{,}9 \cdot 1{,}9 \cdot 2{,}9 \cdot 10^4 \cdot \lg e}{4 \cdot (1000)^4} \approx 0{,}5 \cdot 10^{-9}.$$

Таким чином, залишковий член може вплинути лише на дев'ятий десятковий знак. Зазначимо, що отримане значення $\lg 1001$ збігається із значенням у семизначній таблиці значень.

Приклад 2. Використовуючи таблицю значень функції $y = \sin x$ (табл. 2), знайти $\sin 54°$ та $\sin 56°$, вказати похибку результатів.

Розв'язання. Склавши таблицю різниць (табл. 2), бачимо, що треті різниці практично сталі. Тому у формулі (1.19) достатньо взяти чотири члени. Для обчислення $\sin 54°$ маємо $t = \dfrac{54° - 55°}{5°} = -0{,}2$. За формулою (1.19) дістанемо

Таблиця 2

x, град	y	Δy	$\Delta^2 y$	$\Delta^3 y$
30	0,5000	736	−44	−5
35	0,5736	692	−49	−5
40	0,6428	643	−54	−3
45	0,7071	589	−57	
50	0,7660	532		
55	0,8192			

$$\sin 54° = 0{,}8192 +$$

$$+ (-0{,}2) \cdot 0{,}0532 -$$

$$- \frac{(-0{,}2) \cdot 0{,}8}{2} \cdot 0{,}0057 -$$

$$- \frac{(-0{,}2) \cdot 0{,}8 \cdot 1{,}8}{6} \cdot 0{,}0003 = 0{,}80903.$$

Залишковий член при $n = 3$ в узгодженні із формулою (1.23) має вигляд

$$R_3(x) = h^4 \frac{t(t+1)(t+2)(t+3)}{4!} \cdot f^{(4)}(\xi).$$

У нашому випадку $h = 5^\circ = 0{,}0873$; $t = -0{,}2$; $f^{(4)}(\xi) = \sin \xi \leq 1$. Таким чином,

$$|R_3(54^\circ)| \leq \frac{(0{,}0873)^4 \cdot 0{,}2 \cdot 0{,}8 \cdot 1{,}8 \cdot 2{,}8}{24} \approx 0{,}2 \cdot 10^{-5}.$$

Звідси видно, що залишковий член може вплинути тільки на п'ятий десятковий знак. Тому остаточний результат записуємо у вигляді $\sin 54^\circ = 0{,}8090$. Отримане значення цілком збігається із табличним.

Знайдемо тепер $\sin 56^\circ$. У цьому випадку маємо

$$t = \frac{56^\circ - 55^\circ}{5^\circ} = 0{,}2,$$

і за формулою (1.19) отримаємо

$$\sin 56^\circ = 0{,}8192 + 0{,}2 \cdot 0{,}0532 - \frac{0{,}2 \cdot 1{,}2}{2} \cdot 0{,}0057 -$$

$$- \frac{0{,}2 \cdot 1{,}2 \cdot 2{,}2}{3!} 0{,}0003 = 0{,}8294.$$

Залишковий член при $t = 0{,}2$, $h = 0{,}0873$ оцінюється таким чином:

$$|R_3(56^\circ)| \leq \frac{(0{,}0873)^4 \cdot 0{,}2 \cdot 1{,}2 \cdot 2{,}2 \cdot 3{,}2}{24} \approx 0{,}4 \cdot 10^{-5}.$$

Таким чином, залишковий член може вплинути тільки на п'ятий десятковий знак. Тому приймаємо $\sin 56^\circ = 0{,}8291$.

§ 1.4 ІНТЕРПОЛЯЦІЙНІ ПОЛІНОМИ ГАУССА, СТІРЛІНГА ТА БЕССЕЛЯ

При побудові інтерполяційних формул Ньютона використовуються лише значення функції, що розташовані по один бік від вибраного початкового значення, тобто ці формули носять однобічний характер. У багатьох випадках виявляються корисними інтерполяційні формули, що містять у собі як наступні, так і попередні значення функції по відношенню до її початкового значення. Найбільш використовуються з різницями, розташованими у горизонтальному рядку діагональної таблиці різниць даної функції, що відповідають початковим значенням x_0 та y_0, або у

рядках, які безпосередньо примикають до неї. Ці різниці Δy_{-1}, Δy_0, Δy^2_{-1}, ... називають *центральними різницями* (табл. 3), де

$$x_i = x_0 + ih \quad (i = 0, \pm 1, \pm 2, \ldots), \quad y_i = f(x_i);$$

$$\Delta y_i = y_{i+1} - y_i, \quad \Delta^2 y_i = \Delta y_{i+1} - \Delta y_i \ldots$$

Таблиця 3

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$	$\Delta^6 y$
x_{-4}	y_{-4}						
		Δy_{-4}					
x_{-3}	y_{-3}		$\Delta^2 y_{-4}$				
		Δy_{-3}		$\Delta^3 y_{-4}$			
x_{-2}	y_{-2}		$\Delta^2 y_{-3}$		$\Delta^4 y_{-4}$		
		Δy_{-2}		$\Delta^3 y_{-3}$		$\Delta^5 y_{-4}$	
x_{-1}	y_{-1}		$\Delta^2 y_{-2}$		$\Delta^4 y_{-3}$		$\Delta^6 y_{-4}$
		Δy_{-1}		$\Delta^3 y_{-2}$		$\Delta^5 y_{-3}$	
x_0	y_0		$\Delta^2 y_{-1}$		$\Delta^4 y_{-2}$		$\Delta^6 y_{-3}$
		Δy_0		$\Delta^3 y_{-1}$		$\Delta^5 y_{-2}$	
x_1	y_1		$\Delta^2 y_0$		$\Delta^4 y_{-1}$		$\Delta^6 y_{-2}$
		Δy_1		$\Delta^3 y_0$		$\Delta^5 y_{-1}$	
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$		
		Δy_2		$\Delta^3 y_1$			
x_3	y_3		$\Delta^2 y_2$				
		Δy_3					
x_4	y_4						

Відповідні інтерполяційні формули носять назву *інтерполяційних формул із центральними різницями*. До них відносяться формули Гаусса, Стірлінга, Бесселя.

Нехай є $2n + 1$ рівновіддалених вузлів інтерполювання x_{-n}, $x_{-(n-1)}$, ..., x_{-1}, x_0, x_1, ..., x_{n-1}, x_n, де $\Delta x_i = x_{i+1} - x_i = h = \text{const}$ ($i = -n, -(n-1), \ldots, n-1$)) і для функцій $y = f(x)$ відомі її значення у цих вузлах $y_i = f(x_i)$ ($i = 0, \pm 1, \ldots, \pm n$). Потрібно побудувати многочлен степеня не вище $2n$ такий, що $P(x_i) = y_i$ при $i = 0, \pm 1, \ldots, \pm n$.

Наведемо інтерполяційні формули, опускаючи їх виведення. Перша інтерполяційна формула Гаусса:

$$P(x) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_{-1} + \frac{(t+1)t(t-1)}{3!}\Delta^3 y_{-1} +$$

$$+ \frac{(t+1)t(t-1)(t-2)}{4!}\Delta^4 y_{-2} + \frac{(t+2)(t+1)t(t-1)(t-2)}{5!}\Delta^5 y_{-2} + \ldots +$$

$$+ \frac{(t+n-1)\ldots(t-n+1)}{(2n-1)!}\Delta^{2n-1} y_{-(n-1)} + \frac{(t+n-1)\ldots(t-n)}{(2n)!}\Delta^{2n} y_{-n}. \tag{1.26}$$

Тут $t = \frac{x - x_0}{h}$, а $\Delta y_0, \Delta^2 y_{-1}, \Delta^3 y_{-1}, \Delta^4 y_{-2}, \Delta^5 y_{-2}, \Delta^6 y_{-3}, \ldots$ — центральні різниці, що утворюють у таблиці 3 нижній ламаний рядок.

Друга інтерполяційна формула Гаусса:

$$P(x) = y_0 + t\Delta y_{-1} + \frac{t(t+1)}{2!}\Delta^2 y_{-1} + \frac{(t+1)t(t-1)}{3!}\Delta^3 y_{-2} +$$

$$+ \frac{(t+2)(t+1)t(t-1)}{4!}\Delta^4 y_{-2} + \ldots + \frac{(t+n-1)\ldots(t-n+1)}{(2n-1)!}\Delta^{2n-1} y_{-n} +$$

$$+ \frac{(t+n)(t+n-1)\ldots(t-n+1)}{(2n!)}\Delta^{2n} y_{-n}, \tag{1.27}$$

де $\Delta y_{-1}, \Delta^2 y_{-1}, \Delta^3 y_{-2}, \Delta^4 y_{-2}, \Delta^5 y_{-3}, \Delta^6 y_{-3} \ldots$ — центральні різниці, що утворюють у табл. 3 верхній ламаний рядок.

Формули Гаусса використовуються для інтерполяції у середині таблиці поблизу x_0: перша при $x > x_0$, друга — для $x < x_0$.

Інтерполяційна формула Стірлінга утворюється, якщо взяти середнє арифметичне інтерполяційних формул Гаусса (1.26) і (1.27):

$$P(x) = y_0 + t\frac{\Delta y_{-1} + \Delta y_0}{2} + \frac{t^2}{2}\Delta^2 y_{-1} + \frac{t(t^2-1^2)}{3!}\cdot\frac{\Delta^3 y_{-2} + \Delta^3 y_{-1}}{2} +$$

$$+ \frac{t^2(t^2-1^2)}{4!}\Delta^4 y_{-2} + \frac{t(t^2-1^2)(t^2-2^2)}{5!}\cdot\frac{\Delta^5 y_{-3} + \Delta^5 y_{-2}}{2} +$$

$$+ \frac{t^2(t^2-1^2)(t^2-2^2)}{6!}\Delta^6 y_{-3} + \ldots + \frac{t^2(t^2-1^2)(t^2-2^2)\ldots[t^2-(n-1)^2]}{(2n-1)!} \times$$

$$\times \frac{\Delta^{2n-1} y_{-n} + \Delta^{2n-1} y_{-(n-1)}}{2} + \ldots + \frac{t^2(t^2-1^2)\ldots[t^2-(n-1)^2]}{(2n)!}\Delta^{2n} y_{-n}. \tag{1.28}$$

Із другої інтерполяційної формули Гаусса можна одержати інтерполяційну формулу Бесселя у вигляді

$$P(x) = \frac{y_0 + y_1}{2} + (t - 0{,}5)\Delta y_0 + \frac{t(t-1)}{2} \cdot \frac{\Delta^2 y_{-1} + \Delta^2 y_0}{2} +$$

$$+ \frac{(t-0{,}5)t(t-1)}{3!}\Delta y_{-1}^3 + \frac{t(t-1)(t+1)(t-2)}{4!} \cdot \frac{\Delta^4 y_{-2} + \Delta^4 y_{-1}}{2} +$$

$$+ \frac{t(t-0{,}5)(t-1)(t+1)(t-2)}{5!}\Delta^5 y_{-2} +$$

$$+ \frac{t(t-1)(t+1)(t-2)(t+2)(t-3)}{6!} \cdot \frac{\Delta^6 y_{-3} + \Delta^6 y_{-2}}{2} + \ldots +$$

$$+ \frac{t(t-1)(t+1)(t-2)(t+2)\ldots(t-n)(t+n-1)}{(2n)!} \cdot \frac{\Delta^{2n} y_{n-2} + \Delta^{2n} y_{-n+1}}{2} +$$

$$+ \frac{(t-0{,}5)t(t-1)(t+1)\ldots(t-n)(t+n-1)}{(2n+1)!}\Delta^{2n+1} y_{-n}, \qquad (1.29)$$

де, як і у (1.28), $t = \dfrac{x - x_0}{h}$.

Більш детальний розгляд інтерполяційних формул виявляє, що при $|t| \le 0{,}25$ доцільно використовувати формулу Стірлінга, а при $0{,}25 \le t \le 0{,}75$ — формулу Бесселя.

Залишковий член інтерполяційних формул Гаусса та Стірлінга для порядку $2n$ максимальних використаних різниць із таблиці і при $x \in [x_0 - nh, x_0 + nh]$ має вигляд

$$R_n(x) = \frac{h^{2n+1} f^{(2n+1)}(\xi)}{(2n+1)!} t(t^2 - 1^2)(t^2 - 2^2)(t^2 - 3^2)\ldots(t^2 - n^2),$$

де $t = \dfrac{x - x_0}{h}$, $\xi \in [x_0 - nh, x_0 + nh]$.

Якщо ж аналітичний вираз функції невідомий, то при малому h

$$R_n(x) \approx \frac{\Delta^{2n+1} y_{-n-1} + \Delta^{2n+1} y_{-n}}{2(2n+1)!} t(t^2 - 1^2)(t^2 - 2^2)(t^2 - 3^2)\ldots(t^2 - n^2).$$

Якщо $(2n+1)$ — порядок максимально використаної різниці із таблиці і $x \in [x_0 - nh, x_0 + (n+1)h]$, то залишковий член інтерполяційної формули Бесселя запишеться у вигляді

$$R_n(x) = \frac{h^{2n+2} f^{(2n+2)}(\xi)}{(2n+2)!} t(t^2 - 1^2)(t^2 - 2^2)(t^2 - 3^2)\ldots(t^2 - n^2)(t - (n+1)),$$

де $t = \dfrac{x - x_0}{h}$, $\xi \in [x_0 - nh, x_0 + (n+1)h]$.

Якщо ж функція $f(x)$ задана таблично і крок h малий, то

$$R_n(x) \approx \frac{\Delta^{2n+2}y_{-n-1} + \Delta^{2n+2}y_{-n}}{2(2n+2)!}\, t(t^2-1^2)(t^2-2^2)\dots(t^2-n^2)(t-(n+1)).$$

Зокрема, при $t = 0{,}5$ дістанемо похибку інтерполяції на середину

$$R_n(x) = \frac{h^{2n+2}f^{(2n+2)}(\xi)}{(2n+2)!}(-1)^{n+1} \cdot \frac{[1\cdot 3\cdot 5\cdot \dots \cdot (2n+1)]^2}{2^{2n+2}}$$

або

$$R_n(x) \approx \frac{\Delta^{2n+2}y_{-n-1} + \Delta^{2n+2}y_{-n}}{2(2n+2)!}(-1)^{n+1} \cdot \frac{[1\cdot 3\cdot 5\cdot \dots \cdot (2n+1)]^2}{2^{2n+2}}.$$

§ 1.5 ІНТЕРПОЛЯЦІЙНА ФОРМУЛА ЛАГРАНЖА

Наведені у попередніх параграфах інтерполяційні формули використовуються лише у випадку рівновіддалених вузлів інтерполяції. Для довільно заданих вузлів користуються більш загальною формулою — інтерполяційною формулою Лагранжа.

Нехай на відрізку $[a, b]$ задано $n+1$ значення аргументу $x_0, x_1, \dots x_n$ і відомі для функції $y = f(x)$ відповідні значення $f(x_0) = y_0, f(x_1) = y_1, \dots, f(x_n) = y_n$. Потрібно побудувати многочлен $L_n(x)$ степеня не вище n, який набуває у заданих вузлах $x_0, x_1, \dots x_n$ ті ж самі значення, що й функція $f(x)$, тобто такий, що

$$L_n(x_i) = f(x_i), \quad i = 0, 1, \dots, n. \tag{1.30}$$

Інтерполяційна формула Лагранжа надає можливість подати многочлен $L_n(x)$ у вигляді лінійної комбінації значень функції у вузлах інтерполяції

$$L_n(x) = \sum_{k=0}^{n} C_k(x) f(x_k). \tag{1.31}$$

Знайдемо явний вираз для коефіцієнтів $C_k(x)$. Із умови (1.30) матимемо

$$\sum_{k=0}^{n} C_k(x_i) f(x_k) = f(x_i), \quad i = 0, 1, 2, \dots, n.$$

Ці співвідношення будуть виконуватись, якщо на функції $C_k(x)$ накласти умови

$$C_k(x_i) = \begin{cases} 0, & i \neq k, \\ 1, & i = k, \end{cases} \quad i = 0, 1, 2, \dots, n,$$

які означають, що кожна із функцій $C_k(x)$ $(k = 0, 1, \ldots, n)$ має не менше n нулів на $[a, b]$. Оскільки $L_n(x)$ — многочлен степеня n, коефіцієнти $C_k(x)$ природно шукати також у вигляді многочленів степеня n у вигляді

$$C_k(x) = \lambda_k(x - x_0)(x - x_1) \ldots (x - x_{k-1})(x - x_{k+1}) \ldots (x - x_n).$$

Із умови $C_k(x_k) = 1$ знаходимо

$$\lambda_k^{-1} = (x_k - x_0)(x_k - x_1) \ldots (x_k - x_{k-1})(x_k - x_{k+1}) \ldots (x_k - x_n).$$

Таким чином, коефіцієнти $C_k(x)$ інтерполяційного многочлена (1.31) знаходяться за формулою

$$C_k(x) = \prod_{j \neq k} (x - x_j) \Big/ \prod_{j \neq k} (x_k - x_j). \tag{1.32}$$

Часто коефіцієнти C_k записують у іншому вигляді. Введемо многочлен $\omega(x)$ степеня $(n + 1)$

$$\omega(x) = (x - x_0)(x - x_1) \ldots (x - x_{k-1})(x - x_k)(x - x_{k+1}) \ldots (x - x_n)$$

і обчислимо його похідну у точці x_k

$$\omega'(x_k) = (x_k - x_0)(x_k - x_1) \ldots (x_k - x_{k-1})(x_k - x_{k+1}) \ldots (x_k - x_n).$$

Дістанемо

$$C_k(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)}.$$

Інтерполяційний многочлен Лагранжа запишеться у вигляді

$$L_n(x) = \sum_{k=0}^{n} \frac{\omega(x)}{(x - x_k)\omega'(x_k)} f(x_k) \tag{1.33}$$

або

$$L_n(x) = \sum_{k=0}^{n} f(x_k) \prod_{\substack{j=0 \\ j \neq k}}^{n} (x - x_j) \Big/ \prod_{\substack{j=0 \\ j \neq k}}^{n} (x_k - x_j). \tag{1.34}$$

Вирази

$$l_i^{(n)}(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)} \tag{1.35}$$

називаються *коефіцієнтами Лагранжа.*

Виведено, що оцінювати точність інтерполяційного многочлена Лагранжа можна як і інтерполяційну формулу Ньютона за (1.21). Інколи корисно для спрощення обчислень використовувати інваріантність коефіцієнтів Лагранжа щодо лінійної підстановки. Якщо $x = at + h$, $x_j = at_j + h$ $(j = 0, 1, 2, \ldots, n)$, то

$$L_i^{(n)}(x) = L_i^{(n)}(t).$$

Запишемо формулу Лагранжа для випадку рівновіддалених вузлів. Нехай $x - x_0 = x_2 - x_1 = \ldots = x_n - x_{n-1} = h$. Для спрощення зробимо заміну $x = ht + x_0$, тоді $t_0 = 0$, $t_1 = 1$, $t_2 = 2$, $\ldots$, $t_n = n$, $x - x_k = h(t - k)$

$$\omega(x) = h^{n+1}t(t-1)(t-2)\ldots(t-n), \quad \omega'(x) = (-1)^{n-k}k!(n-k)!h^n.$$

Тоді інтерполяційний многочлен Лагранжа (1.33) запишеться у вигляді

$$L_n(x) = L_n(x + th) = t(t-1)(t-2)\ldots(t-n)\sum_{k=0}^{n}\frac{(-1)^{n-k}f(x_k)}{(t-k)k!(n-k)!}.$$

У випадку рівновіддалених вузлів існують таблиці для коефіцієнтів Лагранжа і процес їх обчислення значно спрощується.

Якщо потрібно знайти не загальний вираз, а лише його значення при конкретних x, і при цьому значення функції подані у досить великій кількості вузлів, то зручно користуватися інтерполяційною схемою Ейткена.

Згідно з цією схемою послідовно обчислюються многочлени

$$L_{i,i+1}(x) = \frac{1}{x_{i+1} - x_i}\begin{vmatrix} y_i & x_i - x \\ y_{i+1} & x_{i+1} - x \end{vmatrix};$$

$$L_{i,i+1,i+2}(x) = \frac{1}{x_{i+2} - x_i}\begin{vmatrix} L_{i,i+1}(x) & x_i - x \\ L_{i+1,i+2}(x) & x_{i+2} - x \end{vmatrix};$$

$$L_{i,i+1,i+2,i+3}(x) = \frac{1}{x_{i+3} - x_i}\begin{vmatrix} L_{i,i+1,i+2}(x) & x_i - x \\ L_{i+1,i+2,i+3}(x) & x_{i+3} - x \end{vmatrix}, \ldots \quad (1.36)$$

Інтерполяційний многочлен n-го степеня, який набуває у точках x_i значення y_i $(i = 0, 1, \ldots, n)$ запишеться у вигляді

$$L_{012\ldots n}(x) = \frac{1}{x_n - x_0}\begin{vmatrix} L_{01\ldots(n-1)}(x) & x_0 - x \\ L_{12\ldots n}(x) & x_n - x \end{vmatrix}. \quad (1.37)$$

Обчислення за схемою Ейткена звичайно проводять доти, поки послідовні значення $L_{01\ldots n}(x)$ і $L_{01\ldots n,(n+1)}(x)$ будуть збігатися у границях заданої точності. Схема Ейткена легко реалізується на ЕОМ і забезпечує можливість автоматичного контролю.

Приклад 1. Побудувати інтерполяційний многочлен Лагранжа для функції $f(x)$, значення якої подано в табл. 4.

Таблиця 4.

i	0	1	2	3
x_i	0	0,1	0,3	0,5
y_i	−0,5	0	0,2	1

Розв'язання. За формулою (1.35) для $n = 3$ дістанемо вираз $L_0^{(3)}(x)$ при $k = 0; 2; 3$

$$L_0^{(3)}(x) =$$

$$= \frac{(x-0,1)\cdot(x-0,3)\cdot(x-0,5)}{(-0,1)\cdot(-0,3)\cdot(-0,5)} = -\frac{x^3-0,9x^2+0,23x-0,015}{0,015};$$

$$L_2^{(3)}(x) = \frac{x\cdot(x-0,1)\cdot(x-0,5)}{0,3\cdot 0,2\cdot(-0,2)} = -\frac{x^3-0,6x^2+0,05x}{0,012};$$

$$L_3^{(3)}(x) = \frac{x(x-0,1)(x-0,3)}{0,5\cdot 0,4\cdot 0,2} = -\frac{x^3-0,4x^2+0,03x}{0,04}$$

($L_1^{(3)}$ у даному випадку знаходити не слід, оскільки $y_1 = 0$). Тоді потрібний многочлен буде мати вигляд

$$L_3(x) = L_0^{(3)}(x)y_0 + L_1^{(3)}(x)y_1 + L_2^{(3)}(x)y_2 + L_3^{(3)}(x)y_3 =$$

$$= \frac{125}{3}x^3 - 30x^2 + \frac{73}{12}x - 0,5.$$

Приклад 2. З якою точністю можна обчислити за формулою Лагранжа $\ln 100,5$ за відомими значеннями $\ln 100$, $\ln 101$, $\ln 102$, $\ln 103$?

Розв'язання. Залишковий член формули Лагранжа із (1.21) при $n = 3$ має вигляд

$$R_3(x) = \frac{f^{(4)}(\xi)}{4!}(x-x_0)(x-x_1)(x-x_2)(x-x_3).$$

У нашому випадку $x_0 = 100$, $x_1 = 101$, $x_2 = 102$, $x_3 = 103$, $x = 100,5$, $100 < \xi < 103$. Оскільки $f(x) = \ln x$, то $f^{(4)}(x) = -\frac{6}{x^4}$.

Таким чином,

$$|R_3(100,5)| \le \frac{6}{(100)^4 4!}\cdot 0,5\cdot 0,5\cdot 1,5\cdot 2,5 = 0,23\cdot 10^{-8}.$$

Приклад 3. Функцію $y = \sqrt[3]{x}$ подано в табл. 5.

Таблиця 5

x	1,0	1,1	1,3	1,5	1,6
y	1,000	1,032	1,091	1,145	1,170

Використовуючи схему Ейткена, знайти $\sqrt[3]{1,15}$.

Розв'язання. Послідовно, згідно з (1.36) знаходимо

$$L_{01}(x) = \frac{1}{x_1-x_0}\begin{vmatrix} y_0 & x_0-x \\ y_1 & x_1-x \end{vmatrix} = \frac{1}{0,1}\begin{vmatrix} 1 & -0,15 \\ 1,032 & -0,05 \end{vmatrix} = 1,048;$$

$$L_{12}(x) = \frac{1}{x_2-x_1}\begin{vmatrix} y_1 & x_1-x \\ y_2 & x_2-x \end{vmatrix} = \frac{1}{0,2}\begin{vmatrix} 1,032 & -0,05 \\ 1,091 & -0,15 \end{vmatrix} = 1,047;$$

$$L_{23}(x) = \frac{1}{x_3 - x_2}\begin{vmatrix} y_2 & x_2 - x \\ y_3 & x_3 - x \end{vmatrix} = \frac{1}{0{,}2}\begin{vmatrix} 1{,}091 & 0{,}15 \\ 1{,}145 & 0{,}35 \end{vmatrix} = 1{,}050;$$

$$L_{34}(x) = \frac{1}{x_4 - x_3}\begin{vmatrix} y_3 & x_3 - x \\ y_4 & x_4 - x \end{vmatrix} = \frac{1}{0{,}1}\begin{vmatrix} 1{,}145 & 0{,}35 \\ 1{,}170 & 0{,}45 \end{vmatrix} = 1{,}057;$$

$$L_{012}(x) = \frac{1}{x_2 - x_0}\begin{vmatrix} L_{01} & x_0 - x \\ L_{12} & x_2 - x \end{vmatrix} = \frac{1}{0{,}3}\begin{vmatrix} 1{,}048 & -0{,}15 \\ 1{,}047 & 0{,}15 \end{vmatrix} = 1{,}048.$$

Значення L_{01} і L_{012} збігаються до третього знака. На цьому обчислення можна припинити і із точністю до 10^{-3} записати $\sqrt[3]{1{,}15} \approx 1{,}048$.

§ 1.6. ІНТЕРПОЛЯЦІЙНИЙ МНОГОЧЛЕН ЕРМІТА

Розглянемо більш загальну постановку задачі інтерполяції. Нехай у вузлах $x_k \in [a, b]$, $k = 0, 1, \ldots, m$, серед яких немає збіжних, задані значення функції $f(x_k)$ та її похідних $f^{(i)}(x_k)$ до порядку $N_k - 1$ включно, $i = 1, 2, \ldots, N_k - 1$. Таким чином, у кожній точці x_k, $k = 0, 1, \ldots, m$ відомі $f(x_k), f'(x_k), \ldots, f^{(N_k-1)}(x_k)$, і отже, всього відомо $N_0 + N_1 + \ldots + N_m$ величин. Потрібно побудувати алгебраїчний многочлен $H_n(x)$ степеня $n = N_0 + \ldots + N_m - 1$, для якого

$$H_n^{(i)}(x_k) = f^{(i)}(x_k), \quad k = 0, 1, \ldots, m; \quad i = 0, 1, \ldots, N_{k-1}. \tag{1.38}$$

Многочлен $H_n(x)$, що задовольняє умови (1.38), називається *інтерполяційним многочленом Ерміта* для функції $f(x)$. Число N_k називається *кратністю вузла* x_k.

Покажемо, що інтерполяційний многочлен Ерміта існує і єдиний. Умови інтерполяції (1.38) являють собою систему лінійних алгебраїчних рівнянь відносно коефіцієнтів $a_0, a_1, \ldots, a_n$ многочлена $H_n(x) = a_0 + a_1 x + \ldots + a_n x^n$.

Кількість рівнянь цієї системи дорівнює числу невідомих і є $N_0 + N_1 + \ldots + N_m$. Тому достатньо показати, що однорідна система

$$H_n^{(i)}(x_k) = 0, \quad k = 0, 1, \ldots, m; \quad i = 0, 1, \ldots, N_{k-1} \tag{1.39}$$

має тільки тривіальний розв'язок $a_0 = a_1 = \ldots = a_n = 0$. Група умов (1.39) при фіксованому k та $i = 0, 1, \ldots, N_k - 1$ означає, що число x_k буде коренем кратності N_k многочлена $H_n(x)$. Таким чином, многочлен $H_n(x)$ має із врахуванням кратності не менш $N_0 + N_1 + \ldots + N_m = n + 1$ коренів на $[a, b]$. Оскільки степінь $H_n(x)$ дорівнює n, цей многочлен тотожно дорівнює 0, а отже, дорівнюють 0 і всі його коефіцієнти, і однорідна система рівнянь (1.39) має єдиний розв'язок $a_0 = a_1 = \ldots = a_n = 0$. Неоднорідна система (1.38) має однозначний розв'язок при будь-яких правих

частинах. Оскільки значення $f^{(i)}(x_k)$, $k = 0, 1, \ldots, m$, $i = 0, 1, \ldots, N_k - 1$ входять лише у праву частину системи (1.38), коефіцієнти a_i многочлена $H_m(x)$ лінійно виражаються через значення $f^{(i)}(x_k)$, і цей многочлен можна подати у вигляді лінійної комбінації

$$H_n(x) = \sum_{k=0}^{m} \sum_{i=0}^{N_k - 1} C_{ki}(x) f^{(i)}(x_k),$$

де C_{ki} — многочлен степеня n. Зважаючи на громіздкість виразів для C_{ki}, вони тут не наводяться.

Одержимо похибку інтерполювання $R_n(x) = f(x) - H_n(x)$.

Для цього розглянемо допоміжну функцію

$$g(s) = f(s) - H_n(s) - k\omega(s), \qquad (1.40)$$

де $s \in [a, b]$, k — стала і

$$\omega(s) = (s - x_0)^{N_0}(s - x_1)^{N_1} \ldots (s - x_m)^{N_m}. \qquad (1.41)$$

Сталу k вибираємо так, щоб у точці інтерполювання x виконувалась умова $g(x) = 0$, тобто призначимо $k = \dfrac{f(x) - H_n(x)}{\omega(x)}$.

Вузли x_k будуть коренями кратності N_k функції $g(s)$, $k = 1, 2, \ldots, m$. Крім того, точка $x \in [a, b]$ є коренем $g(s)$. Таким чином, функція $g(s)$ має із урахуванням кратності $N_0 + N_1 + \ldots + N_m + 1 = n + 2$ кореня на відрізку $[a, b]$.

За теоремою Ролля похідна $g'(s)$ має хоча б один 0 між двома сусідніми коренями функції $g(s)$. Тобто $g'(s)$ має не менш, як $m + 1$ коренів на $[a, b]$ у точках, не збіжних ні з однією із точок $x_0, x_1, \ldots, x_m, x$. Крім того, $g'(s)$ має у точці x_k корінь кратності $N_k - 1$, $k = 0, 1, \ldots, m$. Таким чином, $g'(s)$ із урахуванням кратності має не менш

$$(N_0 - 1) + \ldots + (N_m - 1) + (m + 1) = N_0 + N_1 + \ldots + N_m = n + 1$$

коренів на $[a, b]$. Аналогічно $g''(s)$ має не менш як n коренів тощо. Похідна $g^{(n+1)}(s)$ хоча б один раз обертається на 0 на $[a, b]$, тобто існує точка $\xi \in [a, b]$, у якій $g^{(n+1)}(\xi) = 0$. Із (1.40) маємо

$$g^{(n+1)}(s) = f^{(n+1)}(s) - k\omega^{(n+1)}(s).$$

Оскільки $\omega(s)$ — многочлен степеня $n + 1$ із старшим коефіцієнтом 1, маємо $\omega^{(n+1)}(s) = (n + 1)!$. Тому із умови $g^{(n+1)}(\xi) = 0$ дістанемо, враховуючи вираз для k, вигляд для похибки інтерполювання

$$f(x) - H_n(x) = \frac{f^{(n+1)}(\xi)}{(n + 1)!}(x - x_0)^{N_0}(x - x_1)^{N_1} \ldots (x - x_m)^{N_m}. \qquad (1.42)$$

Зазначимо, що звичайний многочлен Ньютона з такою ж кількістю коефіцієнтів, тобто того ж степеня, також має похибку $o(h^{n+1})$. Але на одній і тій самій сітці величина похибки многочлена Ньютона буде більша, ніж для полінома Ерміта: його допоміжний многочлен $\omega(x)$ (1.21) містить більше вузлів, ніж $\omega(s)$ (1.41), і тому до нього належать більші співмножники. Очевидно також, що чим більш високі похідні використовуються при побудові інтерполяційного полінома Ерміта заданого степеня, тим менше потрібно число вузлів і тим менша буде чисельна величина його похибки, незважаючи на те що порядок точності залишається незмінним.

Приклад. Нехай $x_0 < x_1 < x_2$ — точки, у яких задані значення $f(x_0) = f_0$, $f(x_1) = f_1$, $f(x_2) = f_2$. Потрібно побудувати многочлен третього степеня $H_3(x)$ такий, що

$$H_3(x_0) = f_0,\ H_3(x_1) = f_1,\ H_3'(x_1) = f_1',\ H_3(x_2) = f_2.$$

Розв'язання. Будемо шукати його у вигляді

$$H_3(x) = c_0(x)f_0 + c_1(x)f_1 + c_2(x)f_2 + b_1(x)f_1',$$

де $c_0(x)$, $c_1(x)$, $c_2(x)$ — многочлени третього степеня. Очевидно, що $H_3(x)$ і буде шуканим інтерполяційним многочленом, якщо покласти

$$c_0(x_0) = 1,\ c_1(x_0) = 0,\ c_2(x_0) = 0,\ b_1(x_0) = 0;$$

$$c_0(x_1) = 0,\ c_1(x_1) = 1,\ c_2(x_1) = 0,\ b_1(x_1) = 0;$$

$$c_0(x_2) = 0,\ c_1(x_2) = 0,\ c_2(x_2) = 1,\ b_1(x_2) = 0;$$

$$c_0'(x_1) = 0,\ c_1'(x_1) = 0,\ c_2'(x_1) = 0,\ b_1'(x_1) = 1.$$

Знайдемо многочлени третього степеня, що задовольняють перечислені вимоги. Оскільки многочлен $c_0(x)$ має кратний корінь у точці x_1 і простий корінь у точці x_2, його можна шукати у вигляді

$$c_0(x) = k(x - x_1)^2(x - x_2).$$

Із умови $c_0(x_0) = 1$ знаходимо

$$k = \frac{1}{(x_0 - x_1)^2(x_0 - x_2)}.$$

Таким чином,

$$c_0(x) = \frac{(x - x_1)^2(x - x_2)}{(x_0 - x_1)^2(x_0 - x_2)}.$$

Аналогічно дістанемо

$$c_2(x) = \frac{(x - x_0)(x - x_1)^2}{(x_2 - x_0)(x_2 - x_1)^2};$$

$$b_1(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_1 - x_2)(x_1 - x_0)}.$$

Многочлен $c_1(x)$ шукатимемо у формі

$$c_1(x) = (x - x_0)(x - x_2)(\alpha x + \beta),$$

де α та β — сталі, що потребують визначення. Із умови $c_1(x_1) = 1$ знаходимо

$$\alpha x_1 + \beta = \frac{1}{(x_1 - x_0)(x_1 - x_2)}. \tag{1.43}$$

Умова $c_1'(x_1)$ веде до рівняння

$$(x_1 - x_0)(x_1 - x_2)\alpha + (\alpha x_1 + \beta)(2x_1 - x_0 - x_2) = 0. \tag{1.44}$$

Із рівнянь (1.43) та (1.44) знаходимо

$$\alpha = -\frac{2x_1 - x_0 - x_2}{(x_1 - x_0)^2(x_1 - x_2)^2};$$

$$\beta = \frac{1}{(x_1 - x_0)(x_1 - x_2)}\left(1 + \frac{(2x_1 - x_0 - x_2)x_1}{(x_1 - x_0)(x_1 - x_2)}\right).$$

Таким чином,

$$c_1 = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}\left(1 - \frac{(2x_1 - x_0 - x_2)(x - x_1)}{(x_1 - x_0)(x_1 - x_2)}\right).$$

Шуканий інтерполяційний многочлен $H_3(x)$ має вигляд

$$H_3(x) = \frac{(x - x_2)(x - x_1)^2}{(x_0 - x_2)(x_0 - x_1)^2} \cdot f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} \times$$

$$\times \left(1 - \frac{(2x_1 - x_0 - x_2)(x - x_1)}{(x_1 - x_0)(x_1 - x_2)}\right) \cdot f(x_1) + \frac{(x - x_0)(x - x_1)^2}{(x_2 - x_0)(x_2 - x_1)^2} \times$$

$$\times f(x_2) + \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_1 - x_2)(x_1 - x_0)} \cdot f'(x_1).$$

Згідно з (1.42) похибку інтерполювання у випадку одержаного многочлена $H_3(x)$ можна записати у вигляді

$$f(x) - H_3(x) = \frac{f^{IV}(\xi)}{24}(x - x_0)(x - x_1)^2(x - x_2),$$

де $\xi \in (x_0, x_2)$.

§ 1.7. ЗБІЖНІСТЬ ІНТЕРПОЛЯЦІЇ

При розв'язанні задач інтерполяції функції завжди постає питання: за яких умов похибка методу прямує до нуля, тобто коли і як інтерполяційний многочлен прямує до $f(x)$? На практиці використовують два шляхи переходу до границі. Перший полягає у тому, щоб, зберігаючи степінь інтерполяційного многочлена, зменшити крок сітки, тобто скористатися більш докладними таблицями. Другий — зберігаючи крок сітки, збільшити кількість вузлів інтерполяції, тобто збільшити степінь многочлена.

Розглянемо перший випадок. Якщо $f(x)$ має неперервні похідні до $n+1$-го степеня, то при інтерполюванні многочленом $P_m(x)$ $(m \le n)$ похибка методу буде $o(h^{m+1})$. Отже, при фіксованому степені многочлена і зменшенні кроку сітки похибка $|f(x) - P_n(x)|$ необмежено спадає. Якщо похідна, що входить до оцінки похибки, обмежена, то інтерполяційний поліном рівномірно прямує до $f(x)$ на обмеженому відрізку $a \le x \le b$.

Точно кажучи, для кожного значення x вибирають свої вузли інтерполяції, найближчі (на даній сітці) до точки x, тобто складають свій многочлен $P_m(x)$. При цьому точка x заздалегідь лежить між крайніми вузлами інтерполяції, що використовуються у даному многочлені. Тому поліном $\omega(x)$, що входить у оцінку похибки (1.21), обмежений рівномірно по x:

$$|\omega_m(x)| < \max_i |x - x_i|^{m+1} \le (mh)^{m+1},$$

де h — крок сітки (для нерівномірних сіток — максимальний крок). Для заданої точності ε визначимо крок сітки із умови $M_{m+1}(mh)^{m+1} \le \varepsilon(m+1)!$, де $M_{m+1} = \sup_{x \in [a,b]} |f^{(m+1)}(x)|$. Тоді для всіх сіток із даним або більш дрібним кроком і будь-якої точки відрізку $a \le x \le b$ похибка інтерполяційного многочлена $P_m(x)$, вузли якого були вибрані вказаним вище способом, буде не більше ε.

Аналогічні твердження справедливі для інтерполяційного многочлена Ерміта.

Розглянемо другий випадок. Збільшувати степінь інтерполяційного многочлена зовсім не завжди доцільно: по-перше, не відомо, як швидко зростає максимум похідної M_m із збільшенням її порядку; по-друге, у функції може бути лише скінченна кількість похідних. Тому на практиці інтерполювати поліномами високого степеня не доцільно. Якщо три-п'ять вузлів не забезпечують потрібної точності, тоді потрібно зменшувати крок таблиці, а не збільшувати кількість вузлів.

При цьому потрібно зазначити, що при інтерполюванні степеневими рядами потрібна більша обережність. Труднощі полягають у тому, що при табуляції функцій у міру збільшення порядку різниці зменшуються так

швидко, що через декілька етапів вони зникають. Але це не означає, що інтерполяційна формула, якщо взяти досить велику кількість членів, зменшить помилку до довільно малої величини. У багатьох випадках помилка зменшується завдяки коригуючому впливу більш високих різниць, але потім досягається мінімум, після чого похибка знову починає зростати і стає довільно великою. Тут беруться до уваги математично задані функції, які можуть бути обчислені із будь-якою точністю. Це явище можна простежити на інтерполяційних формулах із центральними різницями. Як відомо, центральні різниці, що стоять вздовж конкретного рядка таблиці різниць визначаються лише сусідніми значеннями, що безпосередньо примикають до даної точки на початку рядка. У процесі переходу до різниць вищого порядку це сусідство розширюється все більше і більше. Тому використання невеликого числа членів інтерполяційної формули відрізняється від використання великого числа членів наступним. У даному випадку припускаємо придатність поліноміального наближення у малому, у зворотному ж випадку наближення вважається дійсним у цілому. Має місце той факт, що наближення многочленами у достатньо малому околі точки завжди надійне й справджується, а у цілому воно не завжди надійне і потребує обережності. О.Рунге у 1901 р. та Е.Борель у 1903 р. встановили, що для простих аналітичних функцій можна отримувати зовсім невірні результати при рівновіддаленому інтерполюванні. Коли ми розміщуємо задані точки щільніше одна до одної, інтерполюючий многочлен, який містить всі наші точки, дійсно необмежено наближується до заданої функції $f(x)$ на більшій частині заданого інтервалу. Але після деякої точки, яку можна заздалегідь визначити, інтерполюючий поліном не прямує ні до якої границі і фактично збільшується необмежено у кожній точці цієї частини інтервалу. Труднощі, досліджені Рунге, зумовлені тільки рівновіддаленим характером заданих значень. Якщо дані розміщуються не еквідистантно, а розподілені по нульових точках многочлена Чебишова $(2n + 1)$ порядку $T_{2n+1}(x)$, то труднощі зникають. Помилки інтерполяції тоді коливаються біля величини одного й того ж порядку протягом усього інтервалу і у кожній точці прямують до нуля при необмеженому збільшенні кількості початкових точок.

§ 1.8. ІНТЕРПОЛЯЦІЯ СПЛАЙНАМИ

Теорія сплайнів або сплайн-апроксимацій являє собою дуже важливий розділ теорії наближення функцій, що інтенсивно розвивається. У багатьох задачах прикладної математики сплайни, тобто кусково-поліноміальні функції, є більш природним апаратом наближення функцій, ніж многочлени.

Як уже зазначалось у попередніх параграфах, практичні можливості застосування інтерполяційних многочленів при рівновіддалених вузлах обмежені. На відміну від цих многочленів, зокрема, послідовності інтер-

поляційних кубічних сплайнів, які дали поштовх розвиненню усієї теорії сплайнів, на рівномірній сітці вузлів завжди збігаються до інтерпольованих функцій, і збіжність підвищується зі збільшенням числа вузлів з покращанням диференційних властивостей функції. При цьому алгоритми обчислення кубічних сплайнів дуже прості і ефективно реалізуються на ЕОМ.

Сплайни характеризуються, в першу чергу, тим, що вони мають добрі апроксиматичні властивості та певну перевагу відносно обчислювальної реалізації. У цьому параграфі наведемо деякі відомості про інтерполяційні поліноміальні сплайни, і зокрема, розглянемо інтерполювання кубічними сплайнами.

Спочатку зупинимось на інтерполяційних поліноміальних сплайнах.

Нехай на відрізку $[a, b]$ задана сітка Δ: $a = x_0 < x_1 < \ldots < x_N = b$. Для цілого $k \geq 0$ позначимо через $C^{(k)}[a, b]$ множину k раз неперервно диференційовних на $[a, b]$ функцій, а через P_m — множину поліномів степеня не більше m.

Функцію $S_{m,k}(x)$ будемо називати *поліноміальним сплайном* степеня m дефекту k (k — ціле число, $1 \leq k \leq m$) з вузлами на сітці Δ, якщо:

1) $S_{m,k}(x) \in P_m$, $x \in [x_i, x_{i+1}]$, $i = 0, 1, 2, \ldots, N-1$;

2) $S_{m,k}(x) \in C^{(m-k)}[a, b]$.

Точки $\{x_i\}$ називаються *вузлами сплайна*.

З наведеного випливає, що сплайн $S_{m,k}(x)$ має неперервні похідні до $m-k$-го порядку, $m-k+1$ похідна від сплайна може бути розривною на $[a, b]$. Множину сплайнів позначимо через $S_{m,k}(\Delta)$. Далі будемо розглядати лише сплайни дефекту 1, тобто $k = 1$, і позначати їх просто через $S_m(x)$.

На основі поліноміальних сплайнів розглянемо інтерполяційні сплайни.

Сплайн $S_m(x)$ називають *інтерполяційним поліноміальним сплайном* на сітці Δ, що інтерполює функцію $f(x)$, якщо:

1) $S_m(x) \in P_m$, $x \in [x_i, x_{i+1}]$, $i = 0, 1, 2, \ldots, N-1$;

2) $S_m(x) \in C^{(m-1)}[a, b]$;

3) $S_m(x_i) = f(x_i) = y_i$, $i = 0, 1, \ldots, N$.

Тут розглядаємо сплайни непарного степеня.

Сплайн $S_m(x)$ спряжений в $N-1$ вузлах з N алгебраїчних поліномів степеня m, і тому він має $N(m+1) - (N-1)m = m + N$ вільних параметрів. Задовольняючи інтерполяційній умові 3), використовуємо $N+1$ вільних параметрів. Крім цього треба ще реалізувати $(N+m) - (N+1) = m - 1$ параметр. Для цього використовують граничні умови, які задаються на кінцях інтервалу в точках $x = a$ і $x = b$. Таким чином, крім умов 1)—3) для визначення інтерполяційного сплайна $S_m(x)$ потрібно ще задати на гра-

ниці інтервалу $[a, b]$ $m-1$ граничних умов у вигляді лінійних комбінацій похідних від $S_m(x)$.

Якщо сплайн $S_m(x)$, що інтерполює функцію $f(x)$ при певних граничних умовах, існує, то його можна подати у вигляді лінійної комбінації базисних функцій. Найбільш простий вигляд мають базисні функції цього зображення через фундаментальні сплайни, тобто у вигляді, аналогічному відомій формулі Лагранжа. Кожній з інтерполяційних умов відповідає сплайн, що інтерполює в цій точці одиницю, а у всіх інших — нуль. Маємо

$$S_m(x) = \sum_{i=0}^{N} f(x_i) F_m^i(x). \tag{1.45}$$

При цьому фундаментальні сплайни мають вигляд $F_m^i(x_p) = \delta_{ip}$ для $x_p \in \Delta$ при $i = 0, 1, \ldots, N$, де δ_{ip} — символ Кронекера, тобто

$$\delta_{ip} = \begin{cases} 1, & i = p; \\ 0, & i \neq p. \end{cases}$$

Крім того, в точках $x = x_0$ і $x = x_N$ фундаментальні сплайни повинні ще задовольняти граничним умовам.

Зображення сплайна $S_m(x)$ у вигляді (1.45) називається *інтерполяційною формулою Лагранжа для сплайнів.* При $N - m = 0$ $F_m^{(i)}(x)$ — фундаментальні поліноми, а $S_m(x)$ — інтерполяційний поліном Лагранжа.

У різних розділах прикладної математики найбільш часто застосовуються кубічні інтерполяційні сплайни. У зв'язку з цим розглянемо більш детально кубічні інтерполяційні сплайни, які випливають із інтерполяційних сплайнів $S_m(x)$ при $m = 3$.

Функція $S(x) = S_3(x)$ називається *кубічним сплайном*, що інтерполює функцію $f(x)$ у вузлах сітки Δ, якщо:

1) $S(x) \in P_3$, $x \in [x_i, x_{i+1}]$, $i = 0, 1, 2, \ldots, N-1$;

2) $S(x) \in C^2[a, b]$;

3) $S(x_i) = f(x_i) = y_i$, $i = 0, 1, \ldots, N$. (1.46)

Кубічний сплайн має два вільних параметри, і тому для побудови інтерполяційного сплайна треба сформулювати дві допоміжні граничні умови, наприклад:

а) $S'(a) = f'(a)$, $S'(b) = f'(b)$;

б) $S''(a) = f''(a)$, $S''(b) = f''(b)$;

с) $S'''(z-0) = S'''(z+0)$, $(z = x_1, \; z = x_{N-1})$.
(1.47)

Розглянемо нерівномірну сітку Δ. Позначимо $h_i = x_{i+1} - x_i$, $y_i = f(x_i)$, $m_i = S'(x_i)$, $M_i = S''(x_i)$ $(i = 0, 1, 2, \ldots, N)$. Запишемо

$$S(x) = M_i \cdot \frac{(x_{i+1} - x)^3}{6h_i} + M_{i+1} \cdot \frac{(x - x_i)^3}{6h_i} + \left(y_i - \frac{M_i h_i^2}{6}\right) \cdot \frac{x_{i+1} - x}{h_i} +$$

$$+ \left(y_{i+1} - \frac{M_{i+1} h_i^2}{6}\right) \cdot \frac{x - x_i}{h_i} \qquad (1.48)$$

та

$$S(x) = m_i \cdot \frac{(x_{i+1} - x)^2 (x - x_i)}{h_i^2} - m_{i+1} \cdot \frac{(x - x_i)^2 \cdot (x_{i+1} - x)}{h_i^2} + y_i \times$$

$$\times \frac{(x_{i+1} - x)^2 [2(x - x_i) + h_i]}{h_i^3} + y_{i+1} \cdot \frac{(x - x_i)^2 [2(x_{i+1} - x) + h_i]}{h_i^3}, \qquad (1.49)$$

де $x \in [x_i, x_{i+1}]$, $i = 0, 1, \ldots, N-1$.

Якщо тепер замість змінної x ввести t таким чином, що $t = (x - x_i)/h_i$, де $t \in [0, 1]$ на кожному інтервалі $[x_i, x_{i+1}]$, то вирази (1.48), (1.49) перетворяться відповідно в такі:

$$S(x) = y_i(1 - t) + y_{i+1} t - \frac{h_i^2 t(t-1)[(2-t)M_i + (1+t)M_{i+1}]}{6}; \qquad (1.50)$$

$$S(x) = y_i(1-t)^2(1+2t) + y_{i+1} t^2(3-2t) + m_i h_i t(1-t)^2 - m_{i+1} h_i t^2(1-t);$$

$$x \in [x_i, x_{i+1}], \quad i = 0, 1, \ldots, N-1. \qquad (1.51)$$

З формули (1.50) випливає, що

$$S'(x) = \frac{y_{i+1} - y_i}{h_i} + h_i/6[(2 - 6t + 3t^2)M_i + (1 - 3t^2)M_{i+1}];$$

$$S''(x) = -M_i(1-t) - M_{i+1} t. \qquad (1.52)$$

З виразів (1.48) і (1.52) маємо, що функції $S(x)$ і $S''(x)$ неперервні в точках x_i, $i = 1, 2, \ldots, N-1$. Задовольняючи вимогу неперервності першої похідної $S'(x)$, із (1.52) отримуємо

$$\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = \frac{6}{h_{i-1} + h_i}\left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}\right), \qquad (1.53)$$

$$i = 1, 2, \ldots, N-1.$$

Рівняння (1.53) разом із граничними умовами (1.47) утворюють систему відносно невідомих M_i.

Для граничних умов типу а) і б) ця система набуває вигляду

$$2M_0 + \lambda_0^* M_1 = d_0^*;$$

$$\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = d_i, \quad i = 1, 2, \ldots, N-1;$$

$$\mu_N^* M_{N-1} + 2M_N = d_N^*, \tag{1.54}$$

де $\lambda_i = \dfrac{h_i}{h_i + h_{i-1}}$; $\mu_i = 1 - \lambda_i$; $d_i = \dfrac{6}{h_{i-1} + h_i}\left(\dfrac{y_{i+1} - y_i}{h_i} - \dfrac{y_i - y_{i-1}}{h_{i-1}}\right)$.

У випадку граничних умов типу а)

$$\lambda_0^* = \mu_N^* = 1, \quad d_0^* = 6/h_0\left(\frac{y_1 - y_0}{h_0} - y_0'\right);$$

$$d_N^* = 6/h_{N-1}\left(y_N' - \frac{y_N - y_{N-1}}{h_{N-1}}\right). \tag{1.55}$$

а для граничних умов типу б) маємо

$$\lambda_0^* = \mu_N^* = 0, \quad d_0^* = 2y_0'', \quad d_N^* = 2y_N''. \tag{1.56}$$

При граничних умовах типу в) система рівнянь (1.54) має вигляд

$$(1 + \lambda_1)M_1 + (\lambda_1 - \mu_1)M_2 = \lambda_1 d_1;$$

$$\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = d_i, \quad i = 2, 3, \ldots, N-2;$$

$$(\mu_{N-1} - \lambda_{N-1})M_{N-2} + (1 + \mu_{N-1})M_{N-1} = \mu_{N-1}d_{N-1};$$

$$M_0 = \lambda_1^{-1}(M_1 - \mu_1 M_2), \quad M_N = \mu_{N-1}^{-1}(M_{N-1} - \lambda_{N-1}M_{N-2}). \tag{1.57}$$

При цьому для всіх типів граничних умов матриці системи рівнянь мають стрічкову структуру, для розв'язання яких існує ефективний метод прогонки (глава 2).

З метою економії обчислень в системах (1.54), (1.57) доцільно ввести $\overline{M}_i = M_i/6$ і замість формули (1.50) використовувати

$$S(x) = y_i + \frac{x - x_i}{h_i}\left((y_{i+1} - y_i) -\right.$$

$$\left. - (x_{i+1} - x)\left[(x_{i+1} - x + h_i)\overline{M}_i + (h_i + x - x_i)\overline{M}_{i+1}\right]\right). \tag{1.58}$$

У випадку, коли сплайн виражається через m_i, задовольняючи умови (1.46), отримуємо

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_i = 3\left(\mu_i \frac{y_{i+1} - y_i}{h_i} + \lambda_i \frac{y_i - y_{i-1}}{h_{i-1}}\right), \quad i = 1, \ldots, N-1, \tag{1.59}$$

де $\mu_i = h_{i-1}(h_{i-1} + h_i)^{-1}$, $\lambda_i = 1 - \mu_i$.

Додаючи до рівнянь (1.59) граничні умови типу а) і б), маємо

$$2m_0 + \mu_0^* m_1 = c_0^0;$$

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = c_i, \quad i = 1, 2, \ldots, N-1;$$

$$\lambda_N^* m_{N-1} + 2m_N = c_N^*, \tag{1.60}$$

де $c_i = 3\left(\mu_i \frac{y_{i+1} - y_i}{h_i} + \lambda_i \frac{y_i - y_{i-1}}{h_{i-1}}\right)$. Тут для умов типу а)

$$\mu_0^* = \lambda_N^* = 0, \quad c_0^* = 2y_0', \quad c_N^* = 2y_N', \tag{1.61}$$

а для умов типу б)

$$\mu_0^* = \lambda_N^* = 1, \quad c_0^* = 3\frac{y_1 - y_0}{h_0} - (h_0/2)y_0'';$$

$$c_N^* = 3\frac{y_N - y_{N-1}}{h_{N-1}} + (h_{N-1}/2)y_N''. \tag{1.62}$$

Для умов типу в) маємо систему рівнянь:

$$(1 + \gamma_0)m_1 + \gamma_0 m_2 = c_1^*;$$

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = c_i, \quad i = 2, \ldots, N-2;$$

$$\gamma_N m_{N-2} + (1 + \gamma_N)m_{N-1} = c_{N-1}^*, \tag{1.63}$$

де

$$c_1^* = (1/3)c_1 + 2\gamma_0 \frac{y_2 - y_1}{h_1};$$

$$c_{N-1}^* = (1/3)c_{N-1} + 2\gamma_N \frac{y_{N-1} - y_{N-2}}{h_{N-2}};$$

$$m_0 + (1 + \gamma_0^2)m_1 - \gamma_0^2 m_2 = 2\left(\frac{y_1 - y_0}{h_0} - \gamma_0^2 \frac{y_2 - y_1}{h_1}\right);$$

$$-\gamma_N^2 m_{N-2} + (1 + \gamma_N^2)m_{N-1} + m_N = 2\left(\frac{y_N - y_{N-1}}{h_{N-1}} - \gamma_N^2 \frac{y_{N-1} - y_{N-2}}{h_{N-2}}\right);$$

$$\gamma_0 = h_0/h_1, \quad \gamma_N = h_{N-1}/h_{N-2}.$$

У системах рівнянь (1.60), (1.63) матриці також мають стрічкову структуру, що дозволяє до розв'язання системи застосувати метод прогонки. Має місце твердження, що інтерполяційний кубічний сплайн $S(x)$, що задовольняє граничні умови (1.47), існує і єдиний.

Крім того, кубічні сплайни мають цікаві екстремальні властивості, які полягають в тому, що інтерполяційний сплайн $S(x)$ на сітці Δ з граничними умовами $S'(a) = f'(a)$; $S'(b) = f'(b)$ або $S''(a) = S''(b) = 0$ мінімізує функціонал $\Phi(f) = \int_a^b [f''(x)]^2 dx$.

Цей факт називають *властивістю мінімальної кривизни*, оскільки $f''(x)$ наближено описує кривизну кривої. Цю властивість кубічного сплайну також пов'язують з мінімізацією потенціальної енергії, яка витрачається на деформацію пружної балки. Наведемо деякі приклади.

Приклад 1. Побудуємо кубічний сплайн $S(x)$, якщо задано значення функції $f(x_i) = y_i$, $i = 1, 2, 3, 4$:

$$x_0 = 0,\ x_1 = 1/4,\ x_2 = 1/2,\ x_3 = 3/4,\ x_4 = 1;$$

$$y_0 = 1,\ y_1 = 2,\ y_2 = 1,\ y_3 = 0,\ y_4 = 1.$$

На кінцях інтервалу $[\,0,\,1\,]$ задані граничні умови $y_0' = 0$, $y_4' = 0$, тобто маємо граничні умови типу а) у формулі (1.47).

Розв'язання. Використовуючи співвідношення (1.55), знаходимо, що

$$\lambda_0^* = \mu_4^* = 1;\ \ \lambda_i = \mu_i = 1/2,\ \ i = 1, 2, 3;$$

$$d_0^* = 96;\ \ d_4^* = -96.$$

Враховуючи одержані значення, з (1.54) отримуємо систему рівнянь відносно невідомих $\overline{M}_i = M_i/6$, $i = 0, 1, \ldots, 4$:

$$2\overline{M}_0 + \overline{M}_1 = 16;$$

$$(1/2)\overline{M}_0 + 2\overline{M}_1 + (1/2)\overline{M}_2 = -16;$$

$$(1/2)\overline{M}_1 + 2\overline{M}_2 + (1/2)\overline{M}_3 = 0;$$

$$(1/2)\overline{M}_2 + 2\overline{M}_3 + (1/2)\overline{M}_4 = 16;$$

$$(1)\overline{M}_3 + 2\overline{M}_4 = -16.$$

Розв'язок цієї системи набуває таких значень: $\overline{M}_0 = 96/7$, $\overline{M}_1 = -80/7$, $\overline{M}_2 = 0$, $\overline{M}_3 = 80/7$, $\overline{M}_4 = -96/7$.

Використовуючи вихідні й одержані дані, за формулою (1.58) знаходимо вирази для шуканого кубічного сплайна:

$$S(x) = \begin{cases} 1 - (32/7)x^2(22x - 9), & 0 \le x \le 1/4; \\ 2 + (1/7)(4x - 1)[-7 + 10(1 - 2x)(3 - 4x)], & 1/4 \le x \le 1/2; \\ 1 - (2/7)(2x - 1)[7 + 5(9 - 16x^2)], & 1/2 \le x \le 3/4; \\ (1/7)(4x - 3)[7 - 4(1 - x)(37 - 44x)], & 3/4 \le x \le 1. \end{cases}$$

Приклад 2. Побудуємо кубічний сплайн $S(x)$ при тих же значеннях вузлів і заданих в них функцій, як і в попередньому прикладі, але при граничних умовах типу б), тобто при $y_0'' = 0$, $y_4'' = 0$.

Розв'язання. Використовуючи вирази (1.56), знаходимо

$\lambda_0^* = 0, \quad \mu_4^* = 0, \quad \lambda_i = \mu_i = 1/2, \quad i = 1, 2, 3; \quad d_0^* = 0, \quad d_4^* = 0.$

З (1.54) отримуємо систему рівнянь:

$$2\overline{M}_0 = 0;$$

$$(1/2)\overline{M}_0 + 2\overline{M}_1 + (1/2)\overline{M}_2 = -16;$$

$$(1/2)\overline{M}_1 + 2\overline{M}_2 + (1/2)\overline{M}_3 = 0;$$

$$(1/2)\overline{M}_2 + 2\overline{M}_3 + (1/2)\overline{M}_4 = 16;$$

$$2\overline{M}_4 = 0.$$

Дістанемо: $\overline{M}_0 = 0, \quad \overline{M}_1 = -8, \quad \overline{M}_2 = 0, \quad \overline{M}_3 = 8, \quad \overline{M}_4 = 0.$

За всіма даними можна побудувати інтерполяційний сплайн $S(x)$:

$$S(x) = \begin{cases} 1 - 2x(16x^2 - 3), & 0 \le x \le 1/4; \\ 2 + 32(x - 1)(x - 1/4)^2, & 1/4 \le x \le 1/2; \\ 1 + 32(x - 1/2)[(x - 1/2)^2 - 3/16], & 1/2 \le x \le 3/4; \\ -4(x - 3/4)(8x^2 - 18x + 9), & 3/4 \le x \le 1. \end{cases}$$

Якщо в прикладі знайти значення похідних від сплайна $S(x)$ на кінцях інтервалу, які дорівнюють $y'(0) = 6$, $y'(1) = 6$, і розглянути задачу прикладу 1 при цих значеннях похідних, то приходимо до тієї ж самої системи рівнянь, яка отримана в прикладі 2. Тобто маємо раніше одержаний сплайн $S(x)$ при значеннях похідних на кінцях інтервалу $y''(0) = y''(1) = 0$.

Цей факт ще раз свідчить про єдиність побудованого кубічного сплайна.

ГЛАВА 2

МЕТОДИ РОЗВ'ЯЗАННЯ СИСТЕМ ЛІНІЙНИХ АЛГЕБРАЇЧНИХ РІВНЯНЬ

§ 2.1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА МЕТОДІВ

Чисельні методи лінійної алгебри відіграють особливу роль у дискретному аналізі, що обумовлено двома причинами.

По-перше, багато лінійних задач математичного аналізу, диференціальних та інтегральних рівнянь після дискретизації зводяться до розв'язування задач лінійної алгебри. Таким чином, чисельні методи лінійної алгебри виявляються інструментом чисельного розв'язання великого кола математичних, а також, науково-технічних задач. По-друге, можна сформулювати таке твердження: більшість нелінійних задач «за малим» лінійні, тобто нелінійні моделі в малому околі деякого розв'язання можуть бути описані лінійними. В основі скінченновимірних лінійних моделей лежить лінійна алгебра, отже, першим кроком розв'язання нелінійних задач є дослідження лінеаризованих моделей, їх дискретизація та чисельні методи лінійної алгебри.

У поданій главі розглядаються чисельні методи розв'язання систем лінійних алгебраїчних рівнянь (СЛАР)

$$A\bar{x} = \bar{f}, \tag{2.1}$$

де $A = \{a_{ij}\}$ $(i, j = 1, \ldots, n)$ — матриця вимірності $n \times n$, $\bar{x} = (x_1, x_2, \ldots, x_n)^T$ — шуканий вектор; $\bar{f} = (f_1, f_2, \ldots, f_n)$ — заданий вектор. Припускається, що визначник матриці A відмінний від нуля, тому розв'язок існує єдиний. Для більшості обчислювальних задач характерним є великий порядок матриці A. Із курсу алгебри відомо, що систему (2.1) можна розв'язати двома способами: за формулами Крамера або методом послідовного виключення невідомих (методом Гаусса). При великих m перший спосіб, побудований на обчислюванні визначників, вимагає $0(m!)$ арифметичних дій, в той час як метод Гаусса — тільки $0(m^3)$ дій. Тому метод Гаусса у різних модифікаціях широко використовується при розв'язуванні на ЕОМ задач лінійної алгебри. Незважаючи на те, що в даний час є достатньо велика кількість методів розв'язання СЛАР і цьому питанню присвячено багато наукових творів, вказати найбільш ефективний метод складно, оскільки потрібні великі експериментальні й теоретичні дослідження.

Методи чисельного розв'язування системи (2.1) підрозділяються на дві групи: прямі та ітераційні. У *прямих* (або точних) методах розв'язок

$\bar{x}$ системи (2.1) визначається за скінченне число арифметичних дій. Якщо ці дії виконуються у цілих числах, дістанемо точні розв'язки системи. Зазначимо, що внаслідок похибок заокруглення при розв'язанні задач на ЕОМ прямі методи не зводять до точного розв'язку системи (2.1), і називати їх *точними* можна лише абстрагуючись від похибок заокруглення. При цьому припускається, що коефіцієнти та праві частини системи (2.1) відомі точно. Частіше над усе ці методи мають два етапи: на першому етапі перетворюють систему до одного або іншого простого вигляду, на другому — розв'язують спрощену систему рівнянь і дістають значення невідомих.

Зіставлення різних прямих методів проводиться звичайно за числом арифметичних дій (а ще частіше — за асимптотикою при великих m числа арифметичних дій), необхідних для здобуття розв'язку. При інших рівних умовах перевага віддається методу з найменшим числом дій.

У *ітераційних методах* (їх також називають методами послідовних наближень) розв'язок $\bar{x}$ системи визначається як границя при $n \to \infty$ послідовних наближень $\bar{x}^{(n)}$, де n — номер ітерації. Як правило, за певне число ітерацій потрібне значення не досягається. Звичайно задається деяке число $\varepsilon > 0$ (точність) й обчислення проводяться до здійснення оцінки

$$\| \bar{x}^{(n)} - \bar{x} \| \leq \varepsilon. \tag{2.2}$$

Число ітерацій $n = n(\varepsilon)$, яке необхідно провести для отримання заданої точності ε, для великої кількості методів можна знайти із теоретичних передумов. Якість різних ітераційних процесів порівнюється з необхідним числом ітерацій $n(\varepsilon)$.

§ 2.2. МЕТОД ГАУССА ТА ЙОГО МОДИФІКАЦІЇ

До точних методів розв'язування СЛАР відноситься метод Гаусса, який широко використовується в обчислювальній практиці.

Основна ідея методу Гаусса полягає у зведенні початкової системи рівнянь до еквівалентної їй системи з трикутною матрицею — прямий хід методу виключення. Із одержаної таким чином системи невідомі знаходяться шляхом послідовної підстановки невідомих — зворотний хід.

Запишемо систему (2.1) у розгорнутому вигляді

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= f_1;\\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= f_2;\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= f_n,
\end{aligned} \tag{2.3}$$

де

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}; \quad \bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \\ x_n \end{pmatrix}; \quad \bar{f} = \begin{pmatrix} f_1 \\ f_2 \\ \\ f_n \end{pmatrix}.$$

Найпростішим варіантом методу виключення Гаусса є *схема єдиного поділу*. Відповідно до неї для побудови алгоритму виключення невідомої x_i i-те рівняння ($i = 1, 2, \dots, n$), яке називається *головним*, поділяється на провідний коефіцієнт a_{ii}, потім отримане таким чином рівняння множиться на $a_{i+1,i}, a_{i+2,i}, \dots, a_{ni}$ і віднімається із відповідних рядків матриці A. В результаті приходимо до системи з трикутною матрицею з одиничними діагональними елементами, еквівалентної початковій системі, з перетвореним вектором правих частин, який є розв'язком системи. Якщо серед провідних елементів який-небудь на i-му кроці перетворюється на нуль, то у відповідній системі для продовження процесу виключення достатньо перенумерувати рівняння.

Розглянемо цей процес детальніше. Припустимо, що $a_{11} \neq 0$. Поділивши перше рівняння на a_{11}, дістанемо

$$x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1n}x_n = y_1, \tag{2.4}$$

де $c_{1j} = \dfrac{a_{ij}}{a_{11}}$, $j = 2, \dots, n$, $y_1 = \dfrac{f_1}{a_{11}}$.

Зупинимось на рівняннях системи (2.3), що залишилися

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = f_n, \quad i = 2, 3, \dots, n. \tag{2.5}$$

Помножимо (2.4) на a_{i1} і віднімемо одержане із i-го рівняння системи (2.5), $i = 2, 3, \dots, n$. У результаті отримаємо систему рівнянь:

$$\begin{gathered} x_1 + c_{12}x_2 + \dots + c_{1j}x_j + \dots + c_{1n}x_n = y_1; \\ a_{22}^{(1)}x_2 + \dots + a_{2j}^{(1)}x_j + \dots + a_{2n}^{(1)}x_n = f_2^{(1)}; \\ \dots\dots\dots\dots\dots\dots \\ a_{n2}^{(1)}x_2 + \dots + a_{nj}^{(1)}x_j + \dots + a_{nn}^{(1)}x_n = f_n^{(1)}. \end{gathered} \tag{2.6}$$

Тут

$$a_{ij}^{(1)} = a_{ij} - c_{1j}a_{i1}, \quad f_i^{(1)} = f_i - y_1a_{i1}, \quad i, j = 2, 3, \dots, n. \tag{2.7}$$

У системі (2.6) невідома x_1 міститься тільки у першому рівнянні, тому надалі достатньо мати діло з укороченою системою рівнянь

$$\begin{gathered} a_{22}^{(1)}x_2 + \dots + a_{2j}^{(1)}x_j + \dots + a_{2n}^{(1)}x_n = f_2^{(1)}; \\ \dots\dots\dots\dots\dots\dots \\ a_{n2}^{(1)}x_2 + \dots + a_{nj}^{(1)}x_j + \dots + a_{nn}^{(1)}x_n = f_n^{(1)}. \end{gathered} \tag{2.8}$$

Цим ми здійснили перший крок методу Гаусса. Якщо $a_{22}^{(1)} \neq 0$, то з системи (2.8) аналогічно можна вилучити невідому x_2 і прийти до системи, еквівалентній (2.3). Виключаючи таким же чином невідомі $x_3, x_4, \ldots, x_n$, прийдемо остаточно до системи рівнянь вигляду

$$\begin{aligned} x_1 + c_{12}x_2 + \ldots + c_{1n}x_n &= y_1; \\ x_2 + \ldots + c_{2n}x_n &= y_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_{n-1} + c_{n-1,n}x_n &= y_{n-1}, \\ x_n &= y_n, \end{aligned} \tag{2.9}$$

еквівалентній початковій системі (2.3).

Одержання системи (2.9) являє собою прямий хід методу Гаусса. Зворотний хід полягає в знаходженні невідомих $x_1, x_2, \ldots, x_n$ з системи (2.9). Оскільки матриця системи має трикутний вигляд, можна послідовно, починаючи з x_n, знайти усі невідомі. Дійсно, $x_n = y_n$, $x_{n-1} = y_{n-1} - c_{n-1,n}x_n$ і т. д. Загальні формули зворотного ходу мають вигляд

$$x_i = y_i - \sum_{j=i+1}^{n} c_{ij}x_j, \quad i = n-1, \ldots, 1; \quad x_n = y_n. \tag{2.10}$$

При реалізації на ЕОМ прямого ходу методу Гаусса немає необхідності виконувати дії зі змінними $x_1, x_2, \ldots, x_n$. Достатньо вказати алгоритм, згідно з яким матриця перетворюється на трикутну, та відповідні перетворення правих частин системи. Дістанемо ці загальні формули. Нехай здійснено перші $k-1$ кроків, тобто вже вилучені змінні $x_1, x_2, \ldots, x_{k-1}$. Тоді маємо систему

$$\begin{aligned} x_1 + c_{12}x_2 + \ldots + c_{1k}x_k + \ldots + c_{1n}x_n &= y_1; \\ x_2 + \ldots + c_{2k}x_k + \ldots + c_{2n}x_n &= y_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_{k-1} + \ldots + c_{k-1k}x_k + \ldots + c_{k-1n}x_n &= y_{k-1}; \\ a_{kk}^{(k-1)}x_k + \ldots + a_{kn}^{(k-1)}x_n &= f_k^{(k-1)}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{nk}^{(k-1)}x_k + \ldots + a_{nn}^{(k-1)}x_n &= f_n^{(k-1)}. \end{aligned} \tag{2.11}$$

Розглянемо k-те рівняння цієї системи

$$a_{kk}^{(k-1)}x_k + \ldots + a_{kn}^{(k-1)}x_n = f_k^{(k-1)}$$

і припустимо, що $a_{kk}^{(k-1)} \neq 0$. Поділивши обидві частини цього рівняння на $a_{kk}^{(k-1)}$, дістанемо

$$x_k + c_{k,k+1}x_{k+1} + \ldots + c_{kn}x_n = y_k, \qquad (2.12)$$

де

$$c_{kj} = a_{kj}^{(k-1)} / a_{kk}^{(k-1)}, \quad j = k+1, k+2, \ldots, m;$$

$$y_k = f_k^{(k-1)} / a_{kk}^{(k-1)}.$$

Помножимо рівняння (2.12) на $a_{ik}^{(k-1)}$ і віднімемо співвідношення з i-го рівняння (2.11), де $i = k+1, k+2, \ldots, n$. У результаті остання група рівнянь системи (2.11) набере вигляду

$$x_k + c_{kk+1}x_{k+1} + \ldots + c_{kn}x_n = y_k;$$

$$a_{k+1k+1}^{(k)}x_{k+1} + \ldots + a_{k+1n}^{(k)}x_n = f_{k+1}^{(k)};$$

. . . . . .

$$a_{nk+1}^{(k)}x_{k+1} + \ldots + a_{nn}^{(k)}x_n = f_n^{(k)},$$

де $a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)}c_{kj}$, $i, j = k+1, k+2, \ldots, n$; $f_i^{(k)} = f_i^{(k-1)} - a_{ik}^{(k-1)}y_k$, $i = k+1, k+2, \ldots, m$.

Таким чином, в прямому ході методу Гаусса коефіцієнти рівнянь перетворюються за таким правилом:

$$a_{kj}^{(0)} = a_{kj}, \quad k, j = 1, 2, \ldots, n; \qquad (2.13)$$

$$c_{kj} = a_{kj}^{(k-1)} / a_{kk}^{(k-1)}, \quad j = k+1, k+2, \ldots, n, \quad k = 1, 2, \ldots, n;$$

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)}c_{kj}, \quad i, j = k+1, k+2, \ldots, n, \quad k = 1, 2, \ldots, n. \qquad (2.14)$$

Обчислення правих частин системи (2.9) здійснюється за формулами

$$f_k^{(0)} = f_k; \quad y_k = f_k^{(k-1)} / a_{kk}^{(k-1)}, \quad k = 1, 2, \ldots, n; \qquad (2.15)$$

$$f_i^{(k)} = f_i^{(k-1)} - a_{ik}^{(k-1)}y_k, \quad i = k+1, k+2, \ldots, n. \qquad (2.16)$$

Коефіцієнти c_{ij} і праві частини y_i, $i = 1, 2, \ldots, n$, $j = i+1, i+2, \ldots, n$ зберігаються у пам'яті ЕОМ та використовуються при зворотному ході за формулами (2.10).

Основним обмеженням методу є припущення про те, що усі провідні елементи $a_{ii}^{(k-1)}$, на які проводиться поділ, відмінні від нуля. Навіть якщо який-небудь провідний елемент не дорівнює нулю, а просто близький до нього, в процесі обчислення може трапитися сильне нагромадження похибок. У такому разі у відповідній системі достатньо буде перенумерувати рівняння для продовження процесу вилучення.

Для великих n, як відомо, число операцій множення та поділу в методі Гаусса близьке до $n^3/3$. Це значить, що на обчислення однієї невідомої витрачається в середньому $n^3/3$ дій.

Однією з модифікацій алгоритму методу вилучення Гаусса є *схема вилучення*, яка полягає в зведенні початкової матриці до трикутної, на діагоналі у якої стоять провідні елементи. Для вилучення невідомої x_1 з усіх рівнянь системи (2.3), крім першого, на першому кроці покладемо, приймаючи до уваги перше рівняння

$$x_1 = \frac{(-1)}{a_{11}}(a_{12}x_2 + \ldots + a_{1n}x_n) + f_1, \qquad (2.17)$$

і підставимо (2.17) у 2, 3, ... , n-е рівняння системи (2.3).

Після підстановки дістаємо перетворену початкову систему вигляду

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= f_1; \\ a_{22}^{(1)}x_2 + \ldots + a_{2n}^{(1)}x_n &= f_2^{(1)}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{2n}^{(1)}x_2 + \ldots + a_{nn}^{(1)}x_n &= f_n^{(1)}, \end{aligned} \qquad (2.18)$$

де елементи матриці $a_{ij}^{(1)}$, $2 \le i, j \le n$, і вектора $b_i^{(1)}$, $2 \le i \le n$ після першого кроку можна отримати за формулами

$$a_{ij}^{(1)} = a_{ij} - \frac{1}{a_{11}} a_{1j}a_{i1};$$

$$f_i^{(1)} = f_i - \frac{1}{a_{11}} a_{i1}f_1.$$

На другому кроці розв'язування невідома x_2 вилучається з усіх рівнянь системи (2.18), крім першого та другого. Нехай $a_{22}^{(1)} \neq 0$. Тоді з другого рівняння виражаємо x_2

$$x_2 = \frac{(-1)}{a_{22}^{(1)}}(a_{23}^{(1)}x_3 + \ldots + a_{2n}^{(1)}x_n) + f_2^{(1)} \qquad (2.19)$$

через другі невідомі та підставляємо у (2.19) 3, 4, ... , n-е рівняння (2.18). Після підстановки дістаємо перетворену початкову систему

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= f_1; \\ a_{22}^{(1)}x_2 + \ldots + a_{2n}^{(1)}x_n &= f_2^{(1)}; \\ a_{33}^{(2)}x_3 + \ldots + a_{3n}^{(2)}x_n &= f_3^{(2)}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n3}^{(2)}x_3 + \ldots + a_{nn}^{(2)}x_n &= f_n^{(2)}, \end{aligned}$$

де елементи матриці $a_{ij}^{(2)}$, $3 \le i, j \le n$ та векторів $f^{(2)}$, $3 \le i \le n$ після другого кроку можна отримати за формулами

$$a_{ij}^{(2)} = a_{ij}^{(1)} - \frac{1}{a_{22}^{(1)}} a_{i2}^{(1)} a_{2j}^{(1)};$$

$$f_i^{(2)} = f_i^{(1)} - \frac{1}{a_{22}^{(1)}} a_{i2}^{(1)} f_2^{(1)}.$$

Продовжуючи цей процес вилучення за умовою, що

$$a_{33}^{(2)} \neq 0, \ldots, a_{n-1n-1}^{(n-2)} \neq 0,$$

після $(n-1)$ кроку вилучення дістанемо перетворену початкову систему

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = f_1;$$

$$a_{22}^{(1)}x_2 + \ldots + a_{2n}^{(1)}x_n = f_2^{(1)};$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_{nn}^{(n-1)}x_n = f_n^{(n-1)} \qquad (2.20)$$

з верхньою трикутною матрицею U

$$U = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ & a_{22}^{(1)} & & \\ & 0 & \ddots & a_{nn}^{(n-1)} \end{bmatrix} \qquad (2.21)$$

і перетвореним вектором правих частин

$$g^{(n)} = (f_1, f_2^{(1)}, \ldots, f_n^{(n-1)}).$$

У матричному запису система (2.20) набуває вигляду

$$U\bar{x} = \bar{g}. \qquad (2.22)$$

Перетворення початкової СЛАР до системи (2.22) з трикутною матрицею це прямий хід вилучення. На зворотному ході обчислюються компоненти вектора розв'язку у зворотному порядку

$$x_n = \frac{f_n^{(n-1)}}{a_{nn}^{(n-1)}};$$

$$x_{n-1} = \frac{1}{a_{n-1n-1}^{(n-2)}} (f_{n-1}^{(n-2)} - a_{n-1,n}^{(n-2)} x_n);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_1 = \frac{1}{a_{11}} (f_1 - a_{1n}x_n - \ldots - a_{12}x_2). \qquad (2.23)$$

Як приклад, розв'яжемо систему рівнянь

$$x_1 - 2x_2 + 3x_3 = 1;$$

$$2x_1 + 3x_2 - x_3 = 2;$$

$$-x_1 - x_2 + x_3 = 3. \qquad (2.24)$$

Перший крок прямого ходу вилучення

$$x_1 - 2x_2 + 3x_3 = 1;$$

$$x_2 - x_3 = 0;$$

$$-3x_2 + 4x_3 = 4.$$

Другий крок

$$x_1 - 2x_2 + 3x_3 = 1;$$

$$x_2 - x_3 = 0;$$

$$x_3 = 4.$$

Зворотний хід

$$x_3 = 4;$$

$$x_2 = 4;$$

$$x_1 = -3.$$

Розглянемо *схему вибору головного елемента.* При необхідності розв'язку системи лінійних рівнянь з більш високою точністю за провідний елемент вибирають найбільший за модулем з елементів матриці A. Вилученням відповідної невідомої дістають систему рівнянь, аналогічну (2.6). Із цієї системи, вибираючи як провідний найбільший за модулем з коефіцієнтів, дістають наступну систему. Продовжуючи цей процес, отримують шукану систему, аналогічну (2.9), з трикутною матрицею.

Таким чином, переставленням рядків та стовпців матриці досягають того, щоб провідний елемент на k-му кроці вилучення був найбільшим за модулем з коефіцієнтів, які враховуються у вилученні

$$|a_{kk}^{(k-1)}| = \max_{k <= i,j <= n} |a_{ij}^{(k-1)}|, 1 \le x \le n,$$

де $a_{ij}^{(0)} = a_{ij}$.

Повне упорядкування вимагає більшої додаткової роботи, тому часто зупиняються на методі вилучення з частковим упорядкуванням по рядках (стовпцях). Вибір провідних елементів здійснюється відповідно до формули

$$|a_{kk}^{(k-1)}| = \max_{k \le i \le n} |a_{ik}^{(k-1)}|.$$

У системах з діагональним переважанням

$$|a_{kk}^{(k-1)}| > \sum_{j=1}^{n} |a_{kj}^{(k-1)}|, \quad k = 1, \ldots, n$$

похибки заокруглення практично не нагромаджуються.

Різні варіанти методу Гаусса з вибором провідного елемента проілюструємо на прикладі системи двох рівнянь

$$a_{11}x_1 + a_{12}x_2 = f_1;$$

$$a_{21}x_1 + a_{22}x_2 = f_2. \qquad (2.25)$$

Припустимо, що $|a_{12}| > |a_{11}|$, тоді на першому кроці будемо вилучати невідому x_2. Такий спосіб еквівалентний тому, що система (2.25) переписується у вигляді

$$a_{12}x_2 + a_{11}x_1 = f_1;$$

$$a_{22}x_2 + a_{21}x_1 = f_2 \qquad (2.26)$$

і до (2.26) застосовується перший крок звичайного методу Гаусса.

Даний спосіб вилучення називається *методом Гаусса з вибором провідного елемента по рядку*. Він еквівалентний застосуванню звичайного методу Гаусса до системи, в якій на кожному кроці вилучення відповідно перенумеровуються невідомі.

Застосовується також метод вилучення Гаусса з вибором провідного елемента по стовпцю. Припустимо, що $|a_{21}| > |a_{11}|$.

Перепишемо систему (2.25) у вигляді

$$a_{21}x_1 + a_{22}x_2 = f_2;$$

$$a_{11}x_1 + a_{12}x_2 = f_1$$

і застосуємо до неї на першому кроці звичайний метод Гаусса. Таким чином, метод Гаусса з вибором провідного елемента по стовпцю еквівалентний застосуванню звичайного методу Гаусса до системи, в якій на кожному кроці вилучення відповідно перенумеровуються рівняння.

При побудові алгоритму методом Гаусса на ЕОМ слід врахувати, що операція ділення на ЕОМ більш трудомістка і тому при побудові провідного рівняння необхідно ділити не на провідний елемент, а множити на елемент, зворотний провідному.

Обчислювальну похибку методу Гаусса можна усунути в арифметиці цілих чисел. При цьому, якщо ділення на провідні елементи вилучення проводиться з остачею, то ці дії не виконуються, а запам'ятовуються чисельник та знаменник. Компоненти вектора розв'язку визначаються у вигляді дробів. Метод Гаусса в арифметиці цілих чисел для загальних матриць A вимагає певного мистецтва масштабування. Справа в тому, що відчутними до похибок заокруглення виявляються системи з великим розкидом за абсолютною величиною коефіцієнтів. Процес масштабування дозволяє «вирівнювати» коефіцієнти системи.

Розглянемо систему рівнянь

$$\begin{cases} 7{,}000x_1 + 4144x_2 = 3104; \\ 592{,}0x_1 + 4308x_2 = 2876. \end{cases}$$

Точний розв'язок: $x_1 = -0{,}6$; $x_2 = 0{,}75$; за схемою єдиного ділення: $x_1 = -0{,}5000$; $x_2 = 0{,}74999$.

Проводимо вирівнювання коефіцієнтів:

$$\begin{cases} 1{,}000 * 7x_1 + 69{,}07 * 60x_2 = 3104; \\ 84{,}57 * 7x_1 + 71{,}80 * 60x_2 = 2876; \end{cases}$$

$$\begin{cases} \bar{x}_1 + 69{,}07\bar{x}_2 = 3104; \\ 84{,}57\bar{x}_1 + 71{,}80\bar{x}_2 = 2876. \end{cases}$$

За схемою єдиного ділення:

$$\begin{cases} \bar{x}_1 + 69{,}07\bar{x}_2 = 3104; \\ \qquad 5769\bar{x}_2 = 2596 * 10^{-2}; \end{cases}$$

$\bar{x}_2 = 45{,}000$; $\bar{x}_1 = -4{,}000$; $7x_1 = -4{,}000$; $x_1 = -0{,}5714$; $60x_2 = 45{,}00$; $x_2 = 0{,}7500$.

§ 2.3. ЗВ'ЯЗОК МЕТОДУ ГАУССА З РОЗКЛАДАННЯМ МАТРИЦІ НА МНОЖНИКИ

Як вже було показано, метод Гаусса перетворює початкову систему рівнянь $A\bar{x} = \bar{f}$ (2.1) в еквівалентну їй систему $C\bar{x} = \bar{y}$ (2.9) (або $U\bar{x} = \bar{g}$ (2.22)).

З'ясуємо, яким чином зв'язані між собою вектори правих частин $\bar{f}$ початкової та $\bar{y}$ (або $\bar{g}$) перетвореної матриць. Для цього звернемося до формул (2.16), із яких послідовно дістанемо

$$f_1 = a_{11}y_1, \quad f_2 = a_{21}y_1 + a_{22}^{(1)}y_2, \ \ldots \tag{2.27}$$

і взагалі

$$f_j = b_{j1}y_1 + b_{j2}y_2 + \ldots + b_{jj}y_j, \quad j = 1, 2, \ldots, m,$$

де b_{j1} — чисельні коефіцієнти, причому $b_{jj} = a_{jj}^{(j-1)}$. Співвідношення (2.27) можна записати у матричному вигляді $\bar{f} = B\bar{y}$, де B — нижня трикутна матриця з елементами $a_{jj}^{(j-1)}$, $j = 1, 2, \ldots, m$ ($a_{11}^{(0)} = a_{11}$) на головній діагоналі. Оскільки основне припущення при формулюванні методу Гаусса полягало в тому, що усі $a_{jj}^{(j-1)} \neq 0$, тому на діагоналі матриці B стоять ненульові елементи, і, отже, матриця B має обернену. Підставляючи в

рівняння (2.9) вираз для $\bar{y}$ у вигляді $\bar{y} = B^{-1}\bar{f}$, приходимо до рівняння $C\bar{x} = B^{-1}\bar{f}$ або, що те ж саме, до

$$BC\bar{x} = \bar{f}. \tag{2.28}$$

Зіставляючи (2.28) з рівнянням (2.1), приходимо до висновку, що в результаті застосування методу Гаусса отримано розкладання початкової матриці в добуток $A = BC$, де B — нижня трикутна матриця з нульовими елементами на головній діагоналі і C — верхня трикутна матриця з одиничною головною діагоналлю.

Аналогічні міркування можна провести і щодо системи рівнянь (2.22), одержаної в результаті застосування відмінного попереднього обчислювального алгоритму. Тут вектор $\bar{g}$ буде мати вигляд $\bar{g} = L^{-1}f^{(i)}$ $(i = 1, 2, \ldots, n-1)$.

Розглянемо схему LU-розкладання. Визначимо нижню трикутну матрицю L з одиничною головною діагоналлю:

$$L = \begin{bmatrix} 1 & & & & \\ \frac{a_{21}}{a_{11}} & 1 & & & 0 \\ \frac{a_{31}}{a_{11}} & \frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{a_{n1}}{a_{11}} & \frac{a_{n2}^{(1)}}{a_{22}^{(1)}} & \cdots & \frac{a_{nn-1}^{(n-2)}}{a_{n-1n-1}^{(n-2)}} & 1 \end{bmatrix}.$$

Прямою підстановкою можна перевірити, що розкладання початкової матриці A в добуток

$$A = LU \tag{2.29}$$

правильне (матриця U має вигляд (2.21)).

Зображення матриці A у вигляді (2.29) називається *LU-розкладанням матриці*. Очевидно, що розв'язання початкової системи рівнянь еквівалентне розв'язанню двох систем з трикутними матрицями

$$L\bar{g} = \bar{f}; \tag{2.30}$$

$$U\bar{x} = \bar{g}. \tag{2.31}$$

Таким чином, розглянутий варіант методу вилучення Гаусса — це визначення вектора $\bar{g}$ розв'язком (2.30) і потім визначення вектора $\bar{x}$ — розв'язком (2.31).

Твердження 1. Для існування LU-розкладання матриці A необхідно й достатньо, щоб у матриці A усі головні мінори були відмінні від нуля.

У довільної невиродженої матриці A головні мінори, тобто

$$|a_{11}|, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & \dots & a_{1n} \\ a_{n1} & \dots & a_{nn} \end{vmatrix},$$

можуть перетворюватися на нуль.

Для застосування методу Гаусса до таких матриць або, що те ж саме, одержання LU-розкладання, необхідно провести переставлення рядків (або стовпців) так, щоб головні мінори стали відмінні від нуля.

Звичайно, переставлення рядків (стовпців) не проводять окремо від процедури вилучення, а з'єднують ці два процеси в один. Якщо $a_{11} = 0$, переставимо рядки матриці A так, щоб у лівому верхньому куті виявився ненульовий елемент. У першому стовпці такий елемент завжди знайдеться, інакше $\det A = 0$. Якщо після першого кроку дістанемо $a_{22}^{(1)} = 0$, то виконаємо, як і вище, переставлення: у другому стовпці завжди знайдеться ненульовий елемент, інакше два перших стовпці були б лінійно залежні і $\det A = 0$. Помістимо рядок з ненульовим елементом у другому стовпці на місце другого рядка, тоді $a_{22}^{(1)} \neq 0$. Продовжуючи цей процес вилучення й переставлення рядків (якщо елемент $a_{kk}^{(k-1)} = 0$) до $k = n$, дістанемо LU-розкладання матриці A з додатковою матрицею P переставлень рядків:

$$PA = LU.$$

Матриця A одержується з одиничної матриці E переставленням тих же рядків. Наприклад, переставлення другого та четвертого рядків матриці відповідає

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Розв'язуючи систему рівнянь (2.24) за схемою LU-розкладання, отримаємо матриці L і U у вигляді

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3/7 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 7 & -7 \\ 0 & 0 & 1 \end{bmatrix}.$$

Відповідно до (2.30) маємо

$$\begin{cases} g_1 & = 1, \\ 2g_1 + g_2 & = 2, \\ -g_1 - 3/7g_2 + g_3 & = 3. \end{cases}$$

Розв'язуючи цю систему, дістаємо

$$\bar{g} = \{1; 0; 4\}.$$

Відповідно до (2.31) маємо

$$\begin{cases} x_1 - 2x_2 + 3x_3 = 1, \\ 7x_2 - 7x_3 = 0, \\ x_3 = 4. \end{cases}$$

Зворотний хід:

$$x_3 = 4;\ x_2 = 4;\ x_1 = -3.$$

§ 2.4. МЕТОД ПРОГОНКИ

Метод вилучення Гаусса для тридіагональних систем або, як його часто називають, *метод прогонки* доцільно використовувати для матриць з діагональним переважанням, які часто виникають при розв'язуванні диференціальних рівнянь різницевими методами.

Для випадку, коли матриця A системи рівнянь (2.1) має тридіагональний вигляд ($|j - i| > 1,\ a_{ij} = 0$), систему лінійних алгебраїчних рівнянь можна записати у вигляді

$$-u_1 + b_1u_2 = -f_1;$$

$$a_iu_{i-1} - c_iu_i + b_iu_{i+1} = -f_i,\quad i = 2, 3, \ldots, n-1;$$

$$a_nu_{n-1} - u_n = -f_n.$$

Виконуючи прямий хід методу Гаусса, дістанемо СЛАР з дводіагональною матрицею. Позначимо елемент кодіагоналі цієї матриці через α_i, $i = 1, \ldots, n-1$, а елементи правої частини одержаної СЛАР через β_i, $i = 1, \ldots, n$. Шуканий розв'язок (зворотний хід) знаходиться за рекурентною формулою

$$u_n = \beta_n;$$

$$u_i = \alpha_iu_{i+1} + \beta_i;$$

$$i = n-1, n-2, \ldots, 1.$$

Взявши до уваги тридіагональну структуру початкової СЛАР, прогінні коефіцієнти можна обчислити за рекурентними формулами

$$\alpha_1 = b_1,\quad \beta_1 = f_1;$$

$$\alpha_i = b_i/(c_i - a_i\alpha_{i-1}),\quad i = 2, \ldots, n-1;$$

$$\beta_i = (f_i + a_i\beta_{i-1})/(c_i - a_i\alpha_{i-1}),\quad i = 2, \ldots, n.$$

При цьому необхідно зберігати в пам'яті ЕОМ інформацію прямого ходу — прогінні коефіцієнти.

Розглянутий варіант є методом правої прогонки. Аналогічно можна отримати формули лівої прогонки. Комбінування правої і лівої прогонки дозволяє дістати метод зустрічної прогонки.

Для стійкості методу прогонки достатньо виконання умови $|\alpha_i| < 1$, $i = 1, \ldots, n-1$ або для початкової СЛАР умови діагонального переважання

$$|c_i| \geq |a_i| + |b_i|, \quad i = 2, \ldots, n-1;$$

$$|b_1| \leq 1, \quad |a_n| \leq 1,$$

причому хоча б для єдиного і повинна виконуватися нерівність.

Розглянемо на прикладі застосування методу прогонки до системи з тридіагональною матрицею

$$2x_1 - x_2 = 0;$$

$$x_1 - 3x_2 - x_3 = -8;$$

$$0{,}5x_2 + x_3 - 0{,}3x_4 = 2{,}8;$$

$$x_3 + 4x_4 = 19;$$

$$\alpha_1 = b_1 = 0{,}5; \quad \alpha_i = b_i/(c_i - a_i\alpha_{i-1}), \quad i = 2, 3;$$

$$\alpha_2 = -0{,}4; \quad \alpha_3 = 0{,}375;$$

$$\beta_1 = f_1 = 0; \quad \beta_i = \frac{f_i + a_i\beta_{i-1}}{c_i - a_i\alpha_{i-1}}, \quad i = 2, 3, 4;$$

$$\beta_2 = 3{,}2; \quad \beta_3 = 1{,}5; \quad \beta_4 = 4;$$

$$x_n = \beta_n; \quad x_i = \alpha_i x_{i+1} + \beta_i, \quad i = 3, 2, 1;$$

$$x_4 = 4; \quad x_3 = 2; \quad x_2 = 2; \quad x_1 = 1.$$

§ 2.5. СХЕМА ХАЛЕЦЬКОГО

Схема Халецького базується на факторизації матриці A початкової системи (2.1) у вигляді $A = LC$, де L і C — трикутні матриці (відповідно нижня і верхня)

$$l_{ij} = 0, \ i < j; \quad c_{ij} = 0, \ i > j; \quad c_{ii} = 1, \ i, j = 1, 2, \ldots, n,$$

тобто

$$L = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & c_{12} & \dots & c_{1n} \\ 0 & 1 & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Елементи матриць L і C знаходяться з розв'язку системи n^2 рівнянь

$$a_{ij} = \sum_{k=1}^{n} l_{ik}c_{kj}, \quad i, j = 1, \dots, n.$$

Специфічний вигляд матриць L і C дозволяє зазначений розв'язок записати за формулами

$$l_{i1} = a_{i1}, \quad i = 1, \dots, n;$$

$$l_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik}c_{kj}, \quad i \geq j > 1; \tag{2.32}$$

$$c_{1j} = a_{1j}/l_{11}, \quad j = 1, \dots, n;$$

$$c_{ij} = \left(a_{ij} - \sum_{k=1}^{i-1} l_{ik}c_{kj}\right)/l_{ii} \quad (1 < i < j). \tag{2.33}$$

Після визначення матриць L і C розв'язання системи (2.1) зводиться до розв'язання двох систем рівнянь з трикутними матрицями

$$L\bar{y} = \bar{f}, \quad C\bar{x} = \bar{y}, \tag{2.34}$$

при цьому припускається розв'язок у явній формі

$$y_1 = f_1/l_{11};$$

$$y_i = \left(f_i - \sum_{k=1}^{i-1} l_{ik}y_k\right)l_{ii}, \quad i > 1; \tag{2.35}$$

$$x_n = y_n; \quad x_i = y_i - \sum_{k=i+1}^{n} c_{ik}x_k, \quad i < n. \tag{2.36}$$

З формул (2.36) видно, що числа y_i вигідно обчислювати разом з коефіцієнтами c_{ij}.

Схема Халецького зручна для роботи на ЕОМ. У порівнянні з методом єдиного ділення зазначена схема потребує більшого числа арифметичних дій, проте вона стійкіша відносно похибок заокруглення.

Приклад.

$$3x_1 + x_2 - x_3 + 2x_4 = 6;$$

$$-5x_1 + x_2 + 3x_3 - 4x_4 = -12;$$

$$2x_1 + x_3 - x_4 = 1;$$

$$x_1 - 5x_2 + 3x_3 - 3x_4 = 3.$$

Розв'язання. Відповідно до формул (2.32), (2.33) обчислюємо елементи l_{ij}, c_{ij} і дістаємо матриці L і C у вигляді

$$L = \begin{bmatrix} 3 & & & 0 \\ -5 & 2{,}666667 & & \\ 2 & -0{,}666667 & 2 & \\ 1 & -5{,}333333 & 6 & 2{,}5 \end{bmatrix};$$

$$C = \begin{bmatrix} 1 & 0{,}333333 & -0{,}333333 & 0{,}666667 \\ & 1 & 0{,}5 & -0{,}25 \\ & & 1 & -1{,}25 \\ & & & 1 \end{bmatrix}.$$

Користуючись формулами (2.35), (2.36) визначаємо y_i і x_i ($i = 1, 2, 3, 4$); $y_1 = 2$; $y_2 = -0{,}75$; $y_3 = -1{,}75$; $y_4 = 3$; $x_1 = 1$; $x_2 = -1$; $x_3 = 2$; $x_4 = 3$.

§ 2.6. МЕТОД КВАДРАТНОГО КОРЕНЯ

Нехай є лінійна система

$$A\bar{x} = \bar{b}, \tag{2.37}$$

де $A = \|a_{ij}\|$ — симетрична матриця, тобто $A' = \|a_{ji}\| = A$; $\bar{x}$ — вектор-стовпець невідомих; $\bar{b}$ — вектор-стовпець з правих частин системи.

Метод квадратного кореня базується на факторизації матриці A системи у вигляді

$$A = LL'; \tag{2.38}$$

тут L — нижня трикутна матриця; L' — транспонована відносно до L матриця. Запишемо L і L' у вигляді

$$L = \begin{bmatrix} \alpha_{11} & 0 & \dots & 0 \\ \alpha_{21} & \alpha_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix};$$

$$L' = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \dots & \alpha_{n1} \\ 0 & \alpha_{22} & \dots & \alpha_{n2} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \alpha_{nn} \end{bmatrix}.$$

Розв'язання системи (2.37) зводиться до розв'язання двох систем вигляду $L\bar{y} = \bar{b}$, $L'\bar{x} = \bar{y}$.

Трикутна структура матриць $L = [\alpha_{ij}]\, \alpha_{ij} = 0$, $j > 1$, $L' = [\alpha_{ij}]'$ систем дозволяє записати їх розв'язання у явній формі:

$$y_1 = b_1/\alpha_{11},\ y_i = \left(b_i - \sum_{j=1}^{i-1} \alpha_{ji} y_j\right)/\alpha_{ii},\ \ i \geq 2; \qquad (2.39)$$

$$x_n = y_n/\alpha_{nn},\ x_i = \left(y_i - \sum_{j=i+1}^{n} \alpha_{ij} x_j\right)/\alpha_{ii},\ \ i < n. \qquad (2.40)$$

При цьому елементи матриці L знаходяться з розв'язку системи (2.37), до якої входять $n(n+1)/2$ рівнянь:

$$a_{k1} = \alpha_{k1}\alpha_{11};$$

$$a_{k2} = \alpha_{k1}\alpha_{21} + \alpha_{k2}\alpha_{22};$$

$$\dots\dots\dots\dots\dots$$

$$a_{kk} = \alpha_{k1}\alpha_{k1} + \alpha_{k2}\alpha_{k2} + \dots + \alpha_{kk}\alpha_{kk}$$

$$(k = 1, 2, \dots, n);$$

при $k = 1$

$$\alpha_{11} = \sqrt{a_{11}};$$

при $k = 2$

$$\alpha_{21} = a_{21}/\alpha_{11};\ \ \alpha_{22} = \sqrt{a_{22} - \alpha_{21}^2};$$

при $k = 3$

$$\alpha_{31} = a_{31}/\alpha_{11};\ \ \alpha_{32} = (a_{32} - \alpha_{31}\alpha_{21})/\alpha_{22};$$

$$\alpha_{33} = \sqrt{a_{33} - \alpha_{31}^2 - \alpha_{32}^2}$$

і т. д.

Таким чином, ми маємо можливість рядок за рядком знаходити елементи α_{ij}, якщо тільки $a_{ij} \neq 0$. При цьому елементи i-го рядка визначаються тільки елементами перших i-рядків і перших i-стовпців матриці A.

У загальному вигляді елементи α_{ij} мають такий вигляд:

$$\alpha_{11} = \sqrt{a_{11}}\ ; \quad \alpha_{1j} = a_{1j}/\alpha_{11} \quad (j > 1);$$

$$\alpha_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} \alpha_{ik}^2} \quad (1 < i \leq n); \qquad (2.41)$$

$$\alpha_{ij} = \left(a_{ij} - \sum_{k=1}^{i-1} \alpha_{ik}\alpha_{jk}\right)/\alpha_{jj} \quad (i > j);$$

$$\alpha_{ij} = 0 \text{ при } i < j.$$

При дійсних a_{ij} можуть мати місце чисто уявні α_{ij}. Метод квадратного кореня застосовується і в цьому разі.

Метод квадратного кореня потребує $n^3/6$ арифметичних дій, тобто при великих n він вдвічі швидший за метод Гаусса і потребує вдвічі менше оперативної пам'яті.

Приклад. Розв'язати систему рівнянь

$$A\bar{x} = \bar{b};$$

$$A = \begin{bmatrix} 6{,}1818 & 0{,}1818 & 0{,}3141 & 0{,}1415 & 0{,}1516 & 0{,}2141 \\ & 7{,}1818 & 0{,}2141 & 0{,}1815 & 0{,}1526 & 0{,}3114 \\ & & 8{,}2435 & 0{,}1214 & 0{,}2516 & 0{,}2618 \\ & & & 9{,}3141 & 0{,}3145 & 0{,}6843 \\ & a_{ik} & & & 5{,}3116 & 0{,}8998 \\ & & & & & 4{,}1313 \end{bmatrix};$$

$$\bar{b}' = (7{,}1818;\ 8{,}2435;\ 9{,}3141;\ 5{,}3116;\ 4{,}1313;\ 3{,}1816).$$

Розв'язання. За формулами (2.41) знаходимо коефіцієнти матриці L'

$$L' = \begin{bmatrix} 2{,}486323 & 0{,}073120 & 0{,}1126331 & 0{,}056911 & 0{,}060974 & 0{,}086111 \\ & 2{,}678991 & 0{,}076473 & 0{,}006619 & 0{,}055300 & 0{,}113892 \\ & & 2{,}867349 & 0{,}038066 & 0{,}083585 & 0{,}084472 \\ & & & 3{,}050415 & 0{,}099720 & 0{,}219198 \\ & a_{ik} & & & 2{,}299543 & 0{,}373697 \\ & & & & & 1{,}978909 \end{bmatrix}.$$

Визначивши значення коефіцієнтів α_{ij}, за формулами (2.39), (2.40) знаходимо значення невідомих:

$$\bar{y} = \{2{,}888522;\ 2{,}998364;\ 3{,}041100;\ 1{,}584361;\ 1{,}468632;\ 0{,}726854\};$$

$$\bar{x} = \{1{,}040932;\ 1{,}050668;\ 1{,}026605;\ 0{,}474071;\ 0{,}578973;\ 0{,}367300\}.$$

§ 2.7. ОБЧИСЛЮВАННЯ ВИЗНАЧНИКА ТА ОБЕРНЕНОЇ МАТРИЦІ

Використовуючи розглядувані прямі методи розв'язання системи лінійних алгебраїчних рівнянь і враховуючи факторизацію початкової системи у вигляді трикутних матриць типу (2.29), можна обчислити значення визначника матриці A за формулою

$$\det A = \det L * \det U.$$

Зазначимо, що визначник трикутних матриць дорівнює добутку діагональних елементів. Отже,

$$\det L = 1, \quad \det U = a_{11}a_{22}^{(1)} \dots a_{nn}^{(n-1)}.$$

Звідси значення визначника матриці A обчислюється за формулою

$$\det A = a_{11}a_{22}^{(1)} \dots a_{nn}^{(n-1)} \det P.$$

Оскільки матриця A одержується переставленням рядків одиничної матриці E, то

$$\det P = \begin{cases} 1, & \text{при зліченному числі } k \text{ переставлень,} \\ -1, & \text{при незліченному числі } k \text{ переставлень.} \end{cases}$$

Остаточно визначник матриці A дорівнює

$$\det A = (-1)^k a_{11}^{(1)} a_{22} \dots a_{nn}^{(n-1)},$$

тобто добуток діагональних елементів трикутної матриці з врахуванням потрібної кількості переставлень рядків дорівнює визначнику матриці A.

Якщо матриця A симетрична, то з врахуванням (2.38) маємо

$$\det A = \det L * \det L' = \det (L)^2 = (\alpha_{11}\alpha_{22} \dots \alpha_{nn})^2.$$

Якщо матриця A вироджена, то при використанні методу Гаусса з вибором головного елемента по стовпцю на деякому кроці вилучення k усі елементи k-го стовпця, які знаходяться нижче головної діагоналі і на ній, будуть дорівнювати нулю.

Отже, розглянемо скорочену систему, яку дістанемо з (2.11) на k-му кроці вилучення:

$$a_{kk}^{(k-1)}x_k + \dots + a_{kn}^{(k-1)}x_n = f_k^{(k-1)};$$

$$a_{k+1k}^{(k-1)}x_k + \dots + a_{k+1n}^{(k-1)}x_n = f_{k+1}^{(k-1)};$$

$$\dots\dots\dots\dots$$

$$a_{nk}^{(k-1)}x_k + \dots + a_{nkn}^{(k-1)}x_n = f_n^{(k-1)}.$$

При розв'язанні зазначеної системи можуть виникнути два випадки: 1) хоча б один з коефіцієнтів $a_{kk}^{(k-1)}, a_{k+1k}^{(k-1)}, \ldots, a_{nkk}^{(k-1)}$ відмінний від нуля; 2) $a_{kk}^{(k-1)} = a_{k+1k}^{(k-1)} = \ldots = a_{nkk}^{k-1} = 0$. Якщо для усіх $k = 1, 2, \ldots, n$ реалізується перший випадок, то систему (2.11) можна розв'язати методом Гаусса з вибором провідного елемента по стовпцю, і, отже, $\det A \neq 0$. Якщо ж $\det A = 0$, то при деякому k реалізується другий випадок. При цьому подальше вилучення стає неможливим і програма мусить видавати інформацію про те, що визначник матриці дорівнює нулю.

Часто розв'язання практичних задач зводиться до СЛАР вигляду (2.3) з фіксованими значеннями a_{ij}, але для різних значень компонент вектора $\bar{f}$. У цих випадках доцільно побудувати обернену матрицю A^{-1}

$$A^{-1}A\bar{x} = A^{-1}\bar{f};$$

$$\bar{x} = A^{-1}\bar{f}.$$

Знаходження матриці, оберненої до матриці A, еквівалентно розв'язанню рівняння

$$AX = E, \tag{2.42}$$

де E — одинична матриця; X — шукана квадратна матриця. Нехай $A = \{a_{ij}\}$, $X = \{x_{ij}\}$. Рівняння (2.42) можна записати у вигляді системи m^2 рівнянь

$$\sum_{k=1}^{m} a_{ik}x_{kj} = \delta_{ij}, \quad i, j = 1, 2, \ldots, m, \tag{2.43}$$

де $\delta_{ij} = 1$ при $i = j$ та $\delta_{ij} = 0$ при $i \neq j$.

Для подальшого важливо зазначити, що система (2.43) розпадається на m незалежних систем рівнянь з однією й тією ж матрицею A, але з різними правими частинами. Ці системи мають вигляд

$$A\bar{x}^{(j)} = \delta^{(j)} \quad (j = 1, 2, \ldots, m), \tag{2.44}$$

де $\bar{x}^{(j)} = \left\{x_{1j}, x_{2j}, \ldots, x_{nj}\right\}^{\mathrm{T}}$, у вектора $\delta^{(j)}$ дорівнює одиниці j-а компонента і нулю всі інші.

Для розв'язання системи (2.42) використовується метод Гаусса (звичайний або з вибором провідного елемента).

Оскільки всі системи мають одну й ту ж матрицю A, достатньо один раз здійснити прямий хід Гаусса, тобто одержати розкладання $A = LU$ та запам'ятати матриці L і U. Зворотний хід здійснюється шляхом розв'язання систем рівнянь

$$L\bar{y}^{(j)} = \delta^{(j)}, \quad \bar{y}^{(j)} = \left\{y_{1j}, y_{2j}, \ldots, y_{mj}\right\}^{\mathrm{T}};$$

$$U\bar{x}^{(j)} = \bar{y}^{(j)} \quad (j = 1, 2, \ldots, m)$$

з трикутними матрицями L і U. Наприклад, для матриці третього порядку (2.43) розпадається на три незалежні системи:

$$\begin{cases} a_{11}x_{11} + a_{12}x_{21} + a_{13}x_{31} = 1, \\ a_{21}x_{11} + a_{22}x_{21} + a_{23}x_{31} = 0, \\ a_{31}x_{11} + a_{32}x_{21} + a_{33}x_{31} = 0; \end{cases}$$

$$\begin{cases} a_{11}x_{12} + a_{12}x_{22} + a_{13}x_{32} = 0, \\ a_{21}x_{12} + a_{22}x_{22} + a_{23}x_{32} = 1, \\ a_{31}x_{12} + a_{32}x_{22} + a_{33}x_{32} = 0; \end{cases}$$

$$\begin{cases} a_{11}x_{13} + a_{12}x_{23} + a_{13}x_{33} = 0, \\ a_{21}x_{13} + a_{22}x_{23} + a_{23}x_{33} = 0, \\ a_{31}x_{13} + a_{32}x_{23} + a_{33}x_{33} = 1. \end{cases}$$

Оскільки системи будуть відрізнятися тільки правими частинами, їх слід розв'язувати паралельно

$$\left[\begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array}\right].$$

Виконавши після цього тричі зворотний хід, знаходимо обернену матрицю.

§ 2.8. ЗАГАЛЬНІ ЗАУВАЖЕННЯ ПРО ПОБУДОВУ ІТЕРАЦІЙНИХ ПРОЦЕСІВ

Ітераційні методи характеризуються тим, що розв'язання системи лінійних алгебраїчних рівнянь знаходяться у вигляді границі деякої послідовності векторів, які будуються за допомогою однакового ітераційного процесу.

Поряд з тим, що ітераційні процеси на відміну від прямих методів мають певні переваги, бо при побудові кожного наступного вже враховується попереднє наближення, кожний ітераційний процес має свою обмежену область застосування, оскільки він може виявитися розбіжним для даної системи, або збіжність процесу буде дуже повільною, що практично не дозволить досягти припустимої точності.

Область широкого застосування ітераційних методів — це системи рівнянь, до яких приводять чисельні методи розв'язання диференціальних рівнянь з частинними похідними. Матриці таких систем мають більше число нульових елементів і на відміну від прямих ітераційні методи не збільшують число ненульових елементів матриці в процесі обчислення. Ефективність ітераційних методів визначається швидкістю збіжності послідовних наближень $\bar{x}(k)$ до розв'язку.

Опишемо ітераційні процеси для розв'язання систем лінійних алгебраїчних рівнянь. Нехай система рівнянь (2.37) є неособливою матрицею A.

Ітераційний процес для системи (2.37) описується за допомогою схеми

$$\bar{x}^{(k)} = \bar{x}^{(k-1)} + H^{(k)}(\bar{b} - A\bar{x}^{(k-1)}) \quad (k = 1, 2, \ldots), \tag{2.45}$$

де задаються матриці $H^{(k)}$ і початкове наближення $\bar{x}^{(0)}$.

Отже, ітераційний процес визначається системою (2.37) і вибором матриць $H^{(k)}$.

Можна показати, що ітераційні процеси, визначені схемою (2.45), мають ту властивість, що для них точний розв'язок $\bar{x}^*$ системи (2.37) є нерухомою крапкою, тобто якщо $\bar{x}^{(k)} = \bar{x}^*$, то із (2.45) випливає, що $\bar{x}^{(k)} = \bar{x}^*$ (k — будь-яка).

Дійсне й обернене твердження. Довільний ітераційний процес

$$\bar{x}^{(k)} = C^{(k)}\bar{x}^{(k-1)} + \bar{z}^{(k)}, \tag{2.46}$$

для якого точний розв'язок $\bar{x}^*$ системи (2.37) є нерухомою крапкою, зводиться до вигляду ітераційного процесу (2.45) ($C^{(k)}$ — задана матриця, $\bar{z}^{(k)}$ — заданий вектор). Насправді, маємо

$$\bar{x}^* = C^{(k)}\bar{x}^* + \bar{z}^{(k)}. \tag{2.47}$$

З (2.46) і (2.47) знаходимо

$$\bar{x}^{(k)} = C^{(k)}\bar{x}^{(k-1)} + \bar{x}^* - C^{(k)}\bar{x}^*. \tag{2.48}$$

Цю рівність можна переписати у вигляді

$$\bar{x}^{(k)} = \bar{x}^{(k-1)} + (E - C^{(k)})A^{-1}A(\bar{x}^* - \bar{x}^{(k-1)}) \tag{2.49}$$

або, позначаючи $(E - C^{(k)})A^{-1} = H^{(k)}$, із (2.49) дістаємо

$$\bar{x}^{(k)} = \bar{x}^{(k-1)} + H^{(k)}(\bar{b} - A\bar{x}^{(k-1)}), \tag{2.50}$$

що й було потрібно.

Наведене твердження свідчить про те, що ітераційний процес у формі (2.50) є загальним для ряду ітераційних процесів.

Ітераційні процеси поділяються на стаціонарні і нестаціонарні. Стаціонарні — це такі процеси, в яких матриці $H^{(k)}$ не залежать від k, тобто $H^{(k)} = H$. У нестаціонарних процесах для кожного k дістаємо свою матрицю $H^{(k)}$. Особливе місце серед нестаціонарних процесів посідають так звані циклічні ітераційні процеси, у яких матриці $H^{(k)}$ періодично через p кроків повторюються, тобто $H^{(p+k)} = H^{(k)}$.

Можна циклічні ітераційні процеси віднести й до стаціонарних, якщо прийняти за один крок результат застосування p ітерацій вихідного процесу.

Від вибору матриць $H^{(k)}$ залежать можливості ітераційного процесу (2.45). Так, матриці $H^{(k)}$ можна обрати таким чином, щоб ітераційний процес збігався для можливо більш широкого класу систем рівнянь, і навпаки, можна звузити клас систем рівнянь з метою максимального прискорення збіжності за рахунок обліку окремих особливостей даної системи. Нижче, використовуючи загальну схему ітераційних процесів (2.45), розглянемо методи простої ітерації і Зейделя.

§ 2.9. МЕТОД ПРОСТОЇ ІТЕРАЦІЇ

Нехай є система лінійних алгебраїчних рівнянь (2.37) або у розгорнутому вигляді

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1;$$

$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n. \qquad (2.51)$$

Залишимо зліва діагональні члени і поділимо усі члени рівнянь системи на коефіцієнти при невідомих у діагональних членах. При цьому припустимо, що ці коефіцієнти не дорівнюють нулю, в противному разі можна перенумерувати члени системи так, щоб уникнути цього.

Після перетворення із системи (2.51) маємо

$$x_1 = -\frac{a_{12}}{a_{11}}x_2 - \ldots - \frac{a_{1n}}{a_{11}}x_n + \frac{b_1}{a_{11}};$$

$$x_2 = -\frac{a_{21}}{a_{22}}x_1 - \ldots - \frac{a_{2n}}{a_{22}}x_n + \frac{b_2}{a_{22}};$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n = -\frac{a_{11}}{a_{nn}}x_1 - \ldots - \frac{a_{nn-1}}{a_{nn}}x_n + \frac{b_n}{a_{nn}}, \qquad (2.52)$$

або у векторному вигляді

$$\bar{x} = \alpha\bar{x} + \beta, \qquad (2.53)$$

де

$$\alpha = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & -\frac{a_{1n-1}}{a_{11}} & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \dots & -\frac{a_{2n-1}}{a_{22}} & -\frac{a_{2n}}{a_{22}} \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & -\frac{a_{nn-1}}{a_{nn}} & 0 \end{bmatrix}; \qquad (2.54)$$

$$\beta = \left(\frac{b_1}{a_{11}}, \frac{b_2}{a_{22}}, \dots, \frac{b_n}{a_{nn}}\right)^T. \qquad (2.55)$$

Система рівнянь у вигляді (2.53) має назву *зведеної системи рівнянь*. Систему рівнянь (2.53) будемо розв'язувати методом послідовних наближень, тобто запишемо процес

$$\bar{x}^{(k)} = \alpha\bar{x}^{(k-1)} + \beta, \quad (k = 1, 2, \dots), \qquad (2.56)$$

де задається початкове наближення $\bar{x}^{(0)}$.

Якщо ітераційний процес (2.56) збігається, то можна показати, що розв'язок $\bar{x}^*$ системи рівнянь (2.53) є і розв'язком системи рівнянь (2.37). Викликає інтерес, в якому відношенні знаходиться ітераційний процес простої ітерації у формі (2.56) до загального вигляду ітераційного процесу у формі (2.45). З цією метою перетворюємо ітераційний процес (2.45) до вигляду (2.56).

$$\bar{x}^{(k)} = (E - H^{(k)}A)\bar{x}^{(k-1)} + H^{(k)}\bar{b}. \qquad (2.57)$$

Порівнюючи ітераційні процеси у формі (2.56) і (2.57), знаходимо

$$H^{(k)}\bar{b} = \beta; \quad E - H^{(k)}A = \alpha. \qquad (2.58)$$

Із першої рівності легко одержуємо

$$H^{(k)} = \begin{bmatrix} \frac{1}{a_{11}} & & 0 \\ & \frac{1}{a_{22}} & \\ & & \ddots \\ 0 & & \frac{1}{a_{nn}} \end{bmatrix}, \qquad (2.59)$$

тобто матриці H від k не залежать, і $H = D^{-1}$, де $D = \begin{bmatrix} a_{11} & & 0 \\ & a_{22} & \\ & & \ddots \\ 0 & & a_{nn} \end{bmatrix}$ — діагональна матриця до матриці A.

Легко перетворюється рівність $E - HA = \alpha$, й ітераційний процес для методу простої ітерації можна записати у вигляді

$$\bar{x}^{(k)} = \bar{x}^{(k-1)} + D^{-1}(\bar{b} - A\bar{x}^{(k-1)}), \quad \bar{x}^{(0)} \quad (k = 1, 2, \ldots). \tag{2.60}$$

Таким чином, ітераційний процес методу простої ітерації є стаціонарним і матриця H обернена діагональній. Звідси можна зробити висновок про обмежене збігання методу простої ітерації, оскільки матриця $H = D^{-1}$ вмішує лише діагональні елементи матриці A, тобто інші елементи матриці A не враховуються у цьому ітераційному процесі.

Для збіжності ітераційного процесу (2.56) достатньо, щоб яка-небудь норма матриці α була менша 1, тобто виконувалась нерівність $\|\alpha\| < 1$.

Розглянемо такі норми матриці:

$$\|\alpha\|_m = \max_i \sum_{j=1}^{n} |\alpha_{ij}|, \quad \|\alpha\|_l = \max_j \sum_{i=1}^{n} |\alpha_{ij}|,$$

$$\|\alpha\|_k = \sqrt{\sum_{ij}^{n} \alpha_{ij}^2}\ . \tag{2.61}$$

Із твердження, що для збіжності процесу простої ітерації достатньо, щоб $\|\alpha\| < 1$, можна дістати деякі критерії збіжності ітераційного процесу методу простої ітерації. Використовуючи вираз (2.61) для норми, отримаємо

$$\max_i \sum_{j=1}^{n} |\alpha_{ij}| < 1, \qquad \max_j \sum_{i=1}^{n} |\alpha_{ij}| < 1. \tag{2.62}$$

З перших двох нерівностей (2.62), приймаючи до уваги (2.54), (2.55), випливає, що для того щоб процес простої ітерації збігався, достатньо, щоб мали місце нерівності

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| < |a_{ii}| \qquad (i = 1, \ldots, n);$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| < |a_{jj}| \qquad (j = 1, \ldots, n), \tag{2.63}$$

тобто для збіжності процесу простої ітерації потрібно, щоб сума модулів усіх членів кожного рядка або кожного стовпця без діагонального була за модулем менша, ніж діагональний.

Ці критерії збіжності ітераційного процесу простої ітерації погоджуються з тим, що в ітераційній схемі процесу (2.60) використовуються тільки діагональні елементи матриці A.

За умови здійснення $|\alpha| < 1$ апріорна оцінка похибки методу простої ітерації визначається нерівністю

$$\|\bar{x}^* - \bar{x}^{(k)}\| \leq \frac{\|\alpha\|^k}{1 - \|\alpha\|} \|\beta\|, \quad k > 1, \tag{2.64}$$

а апостеріорна

$$\|\bar{x}^* - \bar{x}^{(k)}\| \le \frac{\|\alpha\|}{1 - \|\alpha\|} \|\bar{x}^{(k)} - \bar{x}^{(k-1)}\|, \quad k \le 1. \qquad (2.65)$$

Зокрема, для $\|\alpha\| < 1/2$ маємо

$$\|\bar{x}^* - \bar{x}^{(k)}\| \le \|\bar{x}^{(k)} - \bar{x}^{(k-1)}\|, \quad k \le 1. \qquad (2.66)$$

Процес ітерації, який збігається, має важливу властивість самовиправлення, тобто окрема помилка в обчисленні не позначиться на кінцевому результаті, оскільки хибні наближення можна розглядати як початковий вектор.

Твердження щодо збіжності ітераційного процесу накладає жорсткі умови на коефіцієнти вихідної системи (2.37). Проте, якщо $\det A \neq 0$, то за допомогою лінійного комбінування рівнянь останню можна завжди замінити такою еквівалентною системою, що умови збіжності будуть виконані.

Насправді, помножимо (2.37) на матрицю $D = A^{-1} - \varepsilon$, $\varepsilon = \|\varepsilon_{ij}\|$ — матриця з малими за модулем елементами. Матимемо

$$(A^{-1} - \varepsilon)A\bar{x} = D\bar{b};$$

$$\bar{x} = \bar{d} + B\bar{x},$$

де $B = \varepsilon A$; $\bar{d} = D\bar{b}$.

Множення на матрицю D еквівалентне сукупності елементарних перетворень над рівняннями системи. Практично роблять це таким чином. Із поданої системи виділяють рівняння з коефіцієнтами, модулі яких більші суми модулів інших коефіцієнтів рівняння. Кожне виділене рівняння виписують у такий рядок нової системи, щоб найбільший за модулем коефіцієнт зробився діагональним. Із виділених рівнянь, які залишилися невикористаними, системи складають незалежні між собою лінійні комбінації з таким розрахунком, щоб додержуватись указаного вище принципу комплектування нової системи так, щоб усі вільні рядки виявились заповненими. При цьому необхідно потурбуватись, щоб кожне невикористане раніше рівняння увійшло хоча б в одну лінійну комбінацію, яка є рівнянням нової системи.

Розглянемо на прикладі зведення вихідної системи рівнянь до вигляду, зручного для ітерацій:

$$\begin{aligned}
&A \quad 2x_1 + 3x_2 - 4x_3 + x_4 - 3 = 0;\\
&B \quad x_1 - 2x_2 - 5x_3 + x_4 - 2 = 0;\\
&C \quad 5x_1 - 3x_2 + x_3 - 4x_4 - 1 = 0;\\
&D \quad 10x_1 + 2x_2 - x_3 + 2x_4 + 4 = 0.
\end{aligned}$$

У рівнянні B коефіцієнт при x_3 за модулем більший інших; переставимо це рівняння на третє місце, а D — на перше

$$10x_1 + 2x_2 - x_3 + 2x_4 + 4 = 0;$$

. . . . . .

$$x_1 - 2x_2 - 5x_3 + x_4 - 2 = 0;$$

. . . . . .

Для одержання рівняння B з максимальним за модулем коефіцієнтом при x_2 необхідно скласти різницю (A) − (B)

$$\text{B} \quad x_1 + 5x_2 + x_3 - 1 = 0.$$

У нову систему увійшли рівняння A, B, D, тому у рівняння D обов'язково повинно увійти рівняння C

$$2(\text{A}) - (\text{B}) + 2(\text{C}) - \text{D};$$

$$\text{D} \quad 3x_1 + 0x_2 + 0x_3 - 9x_4 - 10 = 0.$$

Таким чином, отримаємо перетворену систему (A − D), еквівалентну початковій, яка задовольняє умову збіжності процесу ітерації. Розв'язавши цю систему відносно діагональних елементів, будемо мати систему, до якої вже можна застосувати метод ітерацій:

$$x_1 = -0{,}2x_2 + 0{,}1x_3 - 0{,}2x_4 - 0{,}4;$$

$$x_2 = 0{,}2x_1 - 0{,}2x_3 + 0{,}2;$$

$$x_3 = 0{,}2x_1 - 0{,}4x_2 + 0{,}2x_4 - 0{,}4;$$

$$x_4 = 0{,}333x_1 + 1{,}111.$$

Приклад. Розв'язати систему рівнянь з точністю $\varepsilon = 10^{-4}$

$$3x_1 + 9x_2 - 0{,}3x_4 = 3{,}5;$$

$$-10x_3 + 1{,}08x_4 - 3x_5 = -10;$$

$$0{,}3x_1 - 0{,}2x_3 - 1{,}6x_4 + 6{,}04x_5 = 14;$$

$$-9x_2 - 10x_3 + x_4 = 4;$$

$$-0{,}3x_1 + x_3 - 4x_4 + 0{,}8x_5 = -5.$$

Розв'язання. Зведемо вихідну систему рівнянь до вигляду, зручного для ітерацій

$$5{,}7x_1 + 0{,}2x_3 + 0{,}84x_4 - 0{,}4x_5 = 1;$$

$$3x_1 + 9x_2 - 0{,}3x_4 = 3{,}5;$$

$$-10x_3 + 1{,}08x_4 - 3x_5 = -10;$$

$$-0{,}3x_1 + x_3 - 4x_4 + 0{,}8x_5 = -5;$$

$$0{,}3x_1 - 0{,}2x_3 - 1{,}6x_4 + 6{,}04x_5 = 14.$$

В одержаній системі діагональні коефіцієнти значно перевищують останні коефіцієнти невідомих. Після зведення системи до нормального вигляду розв'язуємо її методом простої ітерації.

Дістанемо такі наближення коренів:

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$x_4^{(k)}$	$x_5^{(k)}$
0	0,17543860	0,33040935	1,00000000	1,48684216	2,73614502
1	0,11324748	0,40070114	0,33973539	1,87366939	2,81984186
2	0,08528202	0,42291722	0,35640371	1,89667320	2,82787657
3	0,08187097	0,42482102	0,35647768	1,89855444	2,82854652
4	0,08163814	0,42496133	0,35647988	1,89870644	2,82859850
5	0,08161932	0,42497268	0,35648072	1,89871836	2,82860279

§ 2.10. МЕТОД ЗЕЙДЕЛЯ

Ітераційний метод Зейделя розв'язання системи лінійних алгебраїчних рівнянь є деяким узагальненням методу простої ітерації, в якому при обчисленні наступного наближення для x_i-ї компоненти шуканого вектора враховуються вже обчислені раніше наближення компоненти $x_1, x_2, \ldots, x_{i-1}$, тобто в цьому процесі на кожному наближенні використовується більше інформації, ніж у методі простої ітерації.

Для розгляду ітераційного процесу методу Зейделя запишемо зведену систему рівнянь у вигляді

$$x_i = \sum_{j=1}^{n} \alpha_{ij} x_{ij} \quad (i = 1, \ldots, n). \tag{2.67}$$

Нехай задано початкове наближення $\overline{x}^{(0)}$ і вже обчислено наближення $\overline{x}^{(k-1)}$, тобто відомі компоненти $x_1^{(k-1)}, x_2^{(k-1)}, \ldots, x_n^{(k-1)}$.

Побудуємо k-те наближення для усіх компонент x_i $(i = 1, \ldots, n)$. Маємо

$$x_1^{(k)} = \sum_{j=1}^{n} \alpha_{1j} x_j^{(k-1)} + \beta_1;$$

$$x_2^{(k)} = \alpha_{21} x_1^{(k)} + \sum_{j=2}^{n} \alpha_{2j} x_j^{(k-1)} + \beta_2;$$

. . . . . .

$$x_i^{(k)} = \sum_{j=1}^{i-1} \alpha_{ij} x_j^{(k)} + \sum_{j=1}^{n} \alpha_{ij} x_j^{(k-1)} + \beta_i;$$

$$x_n^{(k)} = \sum_{j=1}^{n-1} \alpha_{nj} x_j^{(k)} + \alpha_{nn} x_n^{(k-1)} + \beta_i \quad (k = 1, \ldots, n). \tag{2.68}$$

Ітераційному процесу Зейделя можна надати дві інтерпретації. У першій його можна розглядати як нестаціонарний ітераційний процес, в якому за один крок можна здійснити перехід від вектора

$$\left\{x_1^{(k)}, x_2^{(k)}, \ldots, x_{l-1}^{(k)}, x_l^{(k-1)}, x_{l+1}^{(k-1)}, \ldots, x_n^{(k-1)}\right\}$$

до вектора

$$\left\{x_1^{(k)}, x_2^{(k)}, \ldots, x_{l-1}^{(k)}, x_l^{(k)}, x_{l+1}^{(k-1)}, \ldots, x_n^{(k-1)}\right\},$$

тобто у ньому за один крок змінюється одна компонента x_l. При цьому метод Зейделя буде циклічним процесом, де через n кроків матриці H повторюються,

$$H_{n(k-1)+l} = e_{ll},$$

де

$$e_{ll} = \begin{bmatrix} & (l) & \\ 0 & 0 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 1 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & 0 \end{bmatrix} (l). \qquad (2.69)$$

У другій інтерпретації метод Зейделя можна розглядати як стаціонарний ітераційний процес, в якому за один крок можна вважати результат застосування повного циклу, тобто здійснювати перехід від вектора $\bar{x}^{(k-1)}$ до вектора $\bar{x}^{(k)}$. При цьому систему рівнянь (2.53) запишемо у вигляді

$$\bar{x} = (\alpha_1 + \alpha_2)\bar{x} + \bar{\beta}, \qquad (2.70)$$

де

$$\alpha_1 = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ \alpha_{11} & 0 & \ldots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn-1} & 0 \end{bmatrix}, \quad \alpha_2 = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ 0 & \alpha_{22} & \ldots & \alpha_{21} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \alpha_{nn} \end{bmatrix}.$$

Якщо виходити з виразу (2.70), побудуємо ітераційний процес у вигляді

$$\bar{x}^{(k)} = \alpha_1\bar{x}^{(k)} + \alpha_2\bar{x}^{(k-1)} + \bar{\beta};$$

$$\bar{x}^{(0)} \quad (k = 1, 2, \ldots). \qquad (2.71)$$

Дістанемо

$$\bar{x}^{(k)} = (E - \alpha_1)^{-1}\alpha_2\bar{x}^{(k-1)} + (E - \alpha_1)^{-1}\bar{\beta},$$

$$\bar{x}^{(0)} \quad (k = 1, \ldots). \qquad (2.72)$$

Таким чином, один повний цикл циклічного процесу для системи (2.53) рівносильний одному кроку ітераційного процесу по відношенню до системи

$$\bar{x} = \alpha^* \bar{x} + \bar{\beta}^*, \tag{2.73}$$

де $\alpha^* = (E - \alpha_1)^{-1}\alpha_2$, $\bar{\beta}^* = (E - \alpha_1)^{-1}\bar{\beta}$.

Для ітераційного процесу Зейделя має місце те ж твердження, що й для процесу простої ітерації, тобто достатньою умовою збіжності є нерівність $\|\alpha\| < 1$, де $\|\alpha\|$ — норма матриці α для зведеної системи рівнянь.

Оцінка похибки ітераційного процесу Зейделя виражається нерівністю

$$\|\bar{x}^* - \bar{x}^{(k)}\| \le \frac{\mu}{1-\mu} \|\bar{x}^{(k)} - \bar{x}^{(k-1)}\|,$$

де

$$\mu = \max (q_i/(1 - P_i));$$

$$P_i = \sum_{j=1}^{i-1} |\alpha_{ij}|, \quad q_i = \sum_{j=i}^{n} |\alpha_{ij}| \quad (i = 1, \dots, n).$$

Звичайно метод Зейделя має кращу збіжність, ніж метод простої ітерації, хоча потребує більш громіздких обчислень. Проте принципово можливо, що метод Зейделя збігається повільніше, ніж метод простої ітерації, і навіть розбігається, коли метод простої ітерації збігається. Такі випадки, мабуть, слід віднести до «патології» в розумінні прикладної математики.

Приклад. Методом Зейделя розв'язати систему рівнянь з точністю $\varepsilon = 10^{-4}$

$$0{,}6x_2 - 0{,}3x_4 + 1{,}5x_5 = 4{,}8;$$

$$x_1 + 0{,}1x_2 - 0{,}1x_4 = -10;$$

$$1{,}51x_1 + 1{,}2x_2 - 0{,}6x_3 + 0{,}6x_5 = 20;$$

$$-0{,}6x_2 - 6x_3 + 0{,}6x_4 = 4;$$

$$-0{,}1x_1 + 0{,}6x_3 - 20x_4 - 1{,}9x_5 = -4{,}8.$$

Розв'язання. Зведемо цю систему до вигляду з діагональним переважанням

$$x_1 + 0{,}1x_2 - 0{,}1x_4 = -10;$$

$$0{,}01x_1 + 1{,}05x_2 - 0{,}06x_3 + 0{,}15x_4 + 0{,}6x_5 = 35;$$

$$-0{,}6x_2 - 6x_3 + 0{,}6x_4 = 4;$$

$$-0,1x_1 + 0,6x_2 + 20x_4 - 1,95x_5 = -4,8;$$

$$0,6x_2 - 0,3x_4 + 1,5x_5 = 4,8.$$

Одержану систему зведемо до зручного для ітерацій вигляду:

$$x_1 = -10 - 0,1x_2 + 0,1x_4;$$

$$x_2 = 33,33333 - 0,0095238x_1 + 6x_3 - 15x_4 - 60x_5;$$

$$x_3 = 0,666666 + 0,1x_2 - 0,1x_4;$$

$$x_4 = -0,24 + 0,005x_1 - 0,03x_3 + 0,0975x_5;$$

$$x_5 = 3,2 - 0,4x_2 + 0,2x_4.$$

Використовуючи процес Зейделя, послідовно отримаємо такі ітерації:

x1	x2	x3	x4	x5
0,000000	0,000000	0,000000	0,000000	0,000000
−10,000000	35,000000	4,000000	−4,800000	4,800000
−13,980000	33,266667	0,020000	−4,514000	−10,160000
−13,778067	41,584857	0,221933	−5,835700	−9,409467
−14,742056	41,354411	−0,742056	−5,769448	−13,001083
−14,712386	43,351394	−0,712386	−6,086551	−12,895654
−14,943795	43,337863	−0,943795	−6,077277	−13,757868
−14,941514	43,818212	−0,941514	−6,153403	−13,750601
−14,997161	43,825043	−0,997161	−6,152769	−13,957965
−14,997781	43,940797	−0,997781	−6,171078	−13,960571
−15,011187	43,944872	−1,011187	−6,171310	−14,010534
−15,011618	43,972817	−1,011618	−6,175721	−14,012211
−15,014854	43,974384	−1,014854	−6,175870	−14,024271
−15,015025	43,981143	−1,015025	−6,176934	−14,024928
−15,015808	43,981662	−1,015808	−6,176992	−14,027844
−15,015865	43,983300	−1,015865	−6,177250	−14,028063
−15,016055	43,983459	−1,016055	−6,177269	−14,028770
−15,016073	43,983857	−1,016073	−6,177332	−14,028838
−15,016119	43,983904	−1,016119	−6,177338	−14,029009
−15,016124	43,984000	−1,016124	−6,177353	−14,029029
−15,016135	43,984013	−1,016135	−6,177355	−14,029071
Розв'язок				
−16,016137	43,984037	−1,016137	−6,177358	−14,029076

ГЛАВА 3

РОЗВ'ЯЗАННЯ НЕЛІНІЙНИХ СИСТЕМ РІВНЯНЬ

§ 3.1. ВСТУПНІ ЗАУВАЖЕННЯ. ПОНЯТТЯ ПРО ПРИНЦИП СТИСКАЮЧИХ ВІДОБРАЖЕНЬ

Нехай є загальний вигляд нелінійної системи рівнянь

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= y_1; \\ f_2(x_1, x_2, \ldots, x_n) &= y_2; \\ &\ldots\ldots\ldots\ldots\ldots \\ f_n(x_1, x_2, \ldots, x_n) &= y_n, \end{aligned} \tag{3.1}$$

або у векторному вигляді

$$\bar{f}(\bar{x}) = \bar{y}, \tag{3.2}$$

де $\bar{f}$ — функція (або відображення) з областю визначення $D \in E^n$ та областю значень $G \in E^n$; вектор $\bar{y}$ задається. Кожне рівняння в системі (3.1) визначає деяку поверхню в E^n. Отже, розв'язками системи (3.1) є точки перетину цих поверхонь.

Для прикладу розглянемо систему двох рівнянь з двома невідомими

$$f_1 = x_1^2 - x_2 + \alpha = 0;$$

$$f_2 = -x_1 + x_2^2 + \alpha = 0,$$

де α — дійсний параметр, $1 \le \alpha \le -1$.

Якщо змінювати α у вказаному інтервалі, то мають місце такі випадки (рис. 1):

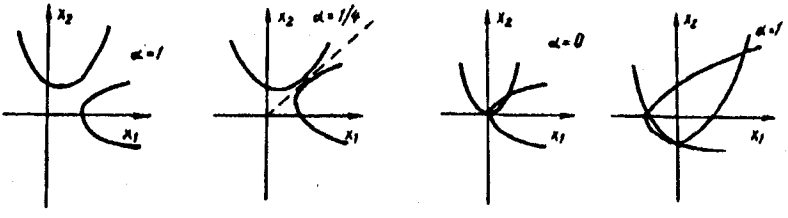


Рис. 1

$\alpha = 1$ — розв'язків немає;

$\alpha = 1/4$ — розв'язок єдиний;

$\alpha = 0$ — два розв'язки;

$\alpha = -1$ — чотири розв'язки.

Підхід до розв'язку нелінійних систем рівнянь засновано на лінеаризації вихідної системи, тобто заміну її на «близьку» лінійну систему з уточненням ітераційними методами.

Питання про існування розв'язку системи рівнянь (3.1) в околі точки (x^0, y^0) вирішує відоме в математичному аналізі твердження:

Твердження 1. Нехай функції $f_i(x_1, \ldots, x_n)$, $1 \le i \le n$, неперервно диференційовні в околі точки

$$x^{(0)} = \{x_1^{(0)}, \ldots, x_n^{(0)}\};$$

$$f_i(x_1^{(0)}, \ldots, x_n^{(0)}) = y_i^{(0)},$$

$$1 \le i \le n$$

і матриця

$$A(x^{(0)}) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(x_1^{(0)}, \ldots, x_n^{(0)}) \ldots \dfrac{\partial f_1}{\partial x_n}(x_1^{(0)}, \ldots, x_n^{(0)}) \\ \ldots\ldots\ldots\ldots \\ \dfrac{\partial f_n}{\partial x_1}(x_1^{(0)}, \ldots, x_n^{(0)}) \ldots \dfrac{\partial f_n}{\partial x_n}(x_1^{(0)}, \ldots, x_n^{(0)}) \end{pmatrix}$$

невироджена. Тоді є неперервні функції $x_i = z_i(y_1, \ldots, y_n)$ в околі точки $y^{(0)} = (y_1^{(0)}, \ldots, y_n^{(0)})$ з властивостями:

$$f_i(z_1(y_1, \ldots, y_n), \ldots, z_n(y_1, \ldots, y_n)) = y_i;$$

$$z_i(y_1^{(0)}, \ldots, y_n^{(0)}) = x_i^{(0)}, \quad 1 \le i \le n.$$

Це твердження дає підхід до розв'язання нелінійних систем рівнянь, заснований на лінеаризації системи (3.1).

Припустимо, що функції $f_i(x_1, \ldots, x_n)$ двічі неперервно диференційовні в E^n. Зобразимо ліву частину (3.1) відрізком ряду Тейлора з центром у точці $x^{(0)}$

$$f_i(\bar{x}^{(0)}) + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\bar{x}^{(0)})(x_j - x_j^{(0)}) + 0\|\bar{x} - \bar{x}^{(0)}\| = y^{(0)}.$$

Якщо відкинути в цій формулі залишковий член, дістанемо лінеаризовану в околі $\bar{x}^0$, $\bar{y}^0$ систему рівнянь (3.1), яку запишемо в матричній формі

$$f(\bar{x}^{(0)}) + A(\bar{x}^{(0)})(\bar{x} - \bar{x}^{(0)}) = \bar{y}^{(0)}. \qquad (3.3)$$

Оскільки точний розв'язок вихідної системи рівнянь невідомий, у загальному випадку

$$f(\bar{x}^{(0)}) \neq y^{(0)}.$$

Нехай $\det A(\bar{x}^{(0)}) \neq 0$. Тоді чисельно розв'язуємо систему лінійних рівнянь (3.3) й дістанемо розв'язок лінеаризованої системи

$$\bar{x} = A^{-1}(\bar{x}^{(0)})(\bar{y}^{(0)} - \bar{f}(\bar{x}^{(0)})) + \bar{x}^{(0)}. \tag{3.4}$$

Похибка чисельного розв'язку (3.3) та похибка, пов'язана з відкиданням залишкового члена, дають повну похибку (3.4), якщо прийняти розв'язок лінеаризованого рівняння за наближений розв'язок (3.1). Похибка лінеаризації визначається числом обумовленості матриці $A(\bar{x}^{(0)})$ та вибором нульового наближення.

Розглянемо лінійний n-вимірний дійсний простір E^n й визначимо в E^n якусь норму $\|\bar{x}\|$ елемента $\bar{x} = \{x_1, \ldots, x_n\}$.

Нехай задано, взагалі кажучи, нелінійне відображення $\bar{y} = \bar{\varphi}(\bar{x})$, наприклад, за допомогою n функцій

$$\begin{aligned} y_1 &= \varphi_1(x_1, \ldots, x_n); \\ y_2 &= \varphi_2(x_1, \ldots, x_n); \\ &\ldots\ldots\ldots\ldots\ldots \\ y_n &= \varphi_n(x_1, \ldots, x_n), \end{aligned} \tag{3.5}$$

які визначені на всьому просторі E^n.

Відображення $\bar{\varphi}(\bar{x})$ переводить E^n в себе, якщо для будь-якого елемента $\bar{x}$, який належить до E^n, елемент $\bar{y} = \bar{\varphi}(\bar{x})$ також належить до E^n.

Наприклад, відображення $y = \sin x$ переводить E^1 в себе, відображення $y = \sqrt{\sin x}$ цієї властивості не має.

Відображення $\bar{\varphi}(\bar{x})$, яке переводить E^n в себе, має назву стискаючого в E, якщо для довільних двох елементів $\bar{x}_1$ та $\bar{x}_2 \in E^n$ має місце нерівність

$$\|\bar{\varphi}(\bar{x}_1) - \varphi(\bar{x}_2)\| \leq q\|\bar{x}_1 - \bar{x}_2\|, \tag{3.6}$$

де коефіцієнт стискання q задовольняє нерівність

$$0 < q < 1. \tag{3.7}$$

Наприклад, $\varphi(x) = 0{,}1 \sin x$ — стискаюче відображення в E^1 з нормою $\|x\| = |x|$. Покажемо, що $\varphi(x)$ задовольняє (3.6) та (3.7). Дійсно,

$$|\varphi(x_1) - \varphi(x_2)| = |0{,}1 \sin x_1 - 0{,}1 \sin x_2| = 0{,}1\,|\sin x_1 - \sin x_2|.$$

Але із теореми про середнє маємо

$$|\sin x_1 - \sin x_2| = |\cos \xi|\,|x_1 - x_2|, \quad x_1 \leq 5 \leq x_2.$$

З двох останніх співвідношень дістанемо

$$||\varphi(x_1) - \varphi(x_2)|| \le 0{,}1|x_1 - x_2|,$$

тобто нерівність (3.6) з коефіцієнтом стискання $q = 0{,}1$. Дію стискаючого відображення ілюструє рис. 2.

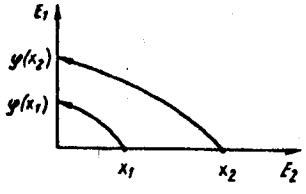


Рис. 2

Для нелінійних функцій $\bar{\varphi}(\bar{x})$ вимога визначення $\bar{\varphi}(\bar{x})$ на всьому просторі E^n виявляється занадто обмежуючою. Тому відображення $\bar{\varphi}(\bar{x})$ часто розглядається локально, тільки в кулі S, яка належить області визначення D-відображення. Куля S — це сукупність елементів $\bar{x} \in E^n$ таких, що $||\bar{x} - \bar{x}^{(0)}|| \le \bar{r}$,

$$S = \{\bar{x}: ||\bar{x} - \bar{x}^{(0)}|| \le \bar{r}\}.$$

Тут $\bar{x}^{(0)}$ — елемент; E^n — центр кулі; r — додатне число, радіус кулі.

Відображення $\bar{\varphi}(\bar{x})$ переводить кулю в себе, якщо для будь-якого $\bar{x} \in S$, елемент $\bar{\varphi}(\bar{x})$ також належить кулі S. Наприклад, відображення $y = \sqrt{\sin x}$ переводить кулю $\left|x - \frac{\pi}{2}\right| \le \frac{\pi}{2}$ з центром в точці $x^{(0)} = \pi/2$ та радіусом $\pi/2$ в себе.

Відображення $\bar{\varphi}(\bar{x})$, яке переводить кулю S в E^n, має назву *стискаючого* в S, якщо для довільних двох елементів $\bar{x}_1$ та $\bar{x}_2$ мають місце нерівності (3.6), (3.7).

§ 3.2. МЕТОД ПРОСТОЇ ІТЕРАЦІЇ. МЕТОД ЗЕЙДЕЛЯ

Стискаюче відображення є чудовим представником класу відображень, які визначають нелінійне рівняння вигляду

$$x_1 = \varphi_1(x_1, \dots, x_n);$$

$$x_2 = \varphi_2(x_1, \dots, x_n);$$

$$\dots\dots\dots\dots$$

$$x_n = \varphi_n(x_1, \dots, x_n), \qquad (3.8)$$

або у векторному вигляді

$$\bar{x} = \bar{\varphi}(\bar{x}), \qquad (3.9)$$

де інтегруючі функції $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ дійсні, визначені та неперервні в деякому околі ω ізольованого кореня $x_1^*, x_2^*, \ldots, x_n^*$ системи (3.8). Розв'язок рівняння (3.8), якщо воно існує, є нерухомою точкою перетворення (3.5).

Для рівняння вигляду (3.8) легко відповісти на питання про існування та єдиності його розв'язку, а також побудувати послідовність наближень $\{\bar{x}^{(k)}\}$, які збігаються до розв'язку.

Побудуємо послідовність $\{\bar{x}^{(k)}\}$ за формулою

$$\bar{x}^{(k+1)} = \bar{\varphi}(\bar{x}^{(k)}), \quad k = 0, 1, 2, \ldots, \tag{3.10}$$

де $\bar{x}^{(0)}$ — початкове наближення, яке має бути задано. Якщо процес ітерації збігається, то граничне значення $\bar{x}^* = \lim\limits_{k \to \infty} \bar{x}^{(k)}$ обов'язково є коренем рівняння (3.8). Таким чином, розв'язок нелінійного рівняння (3.8) можна отримати як границю послідовності ітерацій $\{\bar{x}^{(k)}\}$. Наближений розв'язок (3.8) за допомогою (3.10) має назву *методу простої ітерації.*

Твердження 2. Нехай область G замкнена та відображення (3.5) є стискаючим в G; таким чином справджується умова (3.6). Тоді, якщо для ітераційного процесу (3.10) всі послідовні наближення $\{\bar{x}^{(k)}\} \in G$ $(k = 0, 1, \ldots)$, то:

1) незалежно від вибору початкового наближення $\bar{x}^{(0)}$ процес (3.10) збігається, тобто існує $\bar{x}^* = \lim\limits_{k \to \infty} \bar{x}^{(k)}$;

2) граничний вектор $\bar{x}^*$ є єдиним розв'язком рівняння (3.8) в області G;

3) справедлива оцінка

$$\|\bar{x}^* - \bar{x}^{(k)}\| \le q^k/(1-q)\,\|\bar{x}^{(1)} - \bar{x}^{(0)}\|; \tag{3.11}$$

тут q — константа, що характеризує *коефіцієнт стискання* оператора $\bar{\varphi}(\bar{x})$.

Зауваження 1. Якщо область G збігається з усім простором E^n, то умова $\{\bar{x}^{(k)}\} \in G$ $(k = 0, 1, \ldots)$, як видно, є зайвою.

Зауваження 2. Якщо $0 \le q \le \frac{1}{2}$, то можемо показати, що $\|\bar{x}^{(k)} - \bar{x}^{(k-1)}\| \le \varepsilon$, звідки випливає нерівність $\|\bar{x}^* - \bar{x}^{(k)}\| \le \varepsilon$.

Порівняємо (3.10) з методом простої ітерації в системах лінійних рівнянь. Метод простої ітерації (3.10) є відповідним методом розв'язку лінійних рівнянь, якщо позначити $\bar{\varphi}(\bar{x}) = B\bar{x} + \bar{\alpha}$, де B — матриця; $\bar{x}$, $\bar{\alpha}$ — вектори відповідних розмірностей. Аналогом достатньої умови збігання

простих ітерацій в лінійному випадку ($\|B\| < 1$) виявляється умова стисливості ($q < 1$).

Твердження 3. Процес ітерації (3.10) збігається до єдиного розв'язку рівняння (3.8), якщо при $\bar{x} \in G$ виконується одна з умов

$$\sum_{i=1}^{n} \left| \frac{\partial \varphi_i(x)}{\partial x_j} \right| \le q_j < 1 \quad (j = 1, 2, \dots, n); \tag{3.12}$$

$$\sum_{j=1}^{n} \left| \frac{\partial \varphi_j(x)}{\partial x_i} \right| \le q_i < 1 \quad (i = 1, 2, \dots, n). \tag{3.13}$$

Отже, достатньою умовою збігання ітераційного процесу (3.10) є виконання умови

$$\|M\| \le q < 1, \tag{3.14}$$

де M — матриця з елементами $m_{ij} = \dfrac{\partial \varphi_i}{\partial x_j}$.

Якщо рівняння має вигляд (3.1), то воно попередньо повинно бути перетворено до форми (3.9), причому таким чином, щоб $\overline{\varphi(x)}$ виявилося стискаючим відображенням. Загального прийому для переходу від (3.1) до (3.9) не існує, і тут є важливим попередній аналіз задачі, наприклад, дослідження лінеаризованого рівняння. Зокрема для перетворення

$$F_1(x, y) = 0;$$

$$F_2(x, y) = 0$$

у

$$x = \varphi_1(x, y);$$

$$x = \varphi_2(x, y)$$

з виконанням умов (3.12) або (3.13) можна рекомендувати такий прийом побудови ітерованих функцій:

$$\varphi_1(x, y) = x + \alpha F_1(x, y) + \beta F_2(x, y);$$

$$\varphi_2(x, y) = y + \gamma F_1(x, y) + \delta F_2(x, y); \quad (\alpha\delta \ne \beta\gamma). \tag{3.15}$$

Коефіцієнти $\alpha, \beta, \gamma, \delta$ шукаються із умов

$$\begin{cases} \dfrac{\partial \varphi_1(x_0, y_0)}{\partial x} = 0; & \dfrac{\partial \varphi_2(x_0, y_0)}{\partial x} = 0; \\ \dfrac{\partial \varphi_1(x_0, y_0)}{\partial y} = 0; & \dfrac{\partial \varphi_2(x_0, y_0)}{\partial y} = 0, \end{cases}$$

що для (3.15) має вигляд:

$$1 + \alpha \frac{\partial F_1(x_0, y_0)}{\partial x} + \beta \frac{\partial F_2(x_0, y_0)}{\partial x} = 0;$$

$$\gamma \frac{\partial F_1(x_0, y_0)}{\partial x} + \delta \frac{\partial F_2(x_0, y_0)}{\partial x} = 0;$$

$$\alpha \frac{\partial F_1(x_0, y_0)}{\partial y} + \beta \frac{\partial F_2(x_0, y_0)}{\partial y} = 0;$$

$$1 + \gamma \frac{\partial F_1(x_0, y_0)}{\partial y} + \delta \frac{\partial F_2(x_0, y_0)}{\partial y} = 0.$$

При такому виборі параметрів умови (3.12) — (3.14) виконуються, якщо $F_1(x, y)$, $F_2(x, y)$ змінюються не швидко в околі (x_0, y_0).

Повернемось до запису системи у вигляді (3.2). Припустимо, що за допомогою обчислювального алгоритму (3.10) обчислення доведені до наближення номера k: $\bar{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$. В методі Зейделя для пошуку наступного наближення $\bar{x}^{(k+1)} = (x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_n^{(k+1)})$ насамперед слід встановити порядок обчислення його компонентів $x_i^{(k+1)}$ $(i = 1, 2, \dots, n)$. Такий порядок має бути своїм для кожної системи і для кожного кроку. Оскільки будь-яке розташування x_i^{k+1} можна провести шляхом зміни нумерації до натурального порядку $x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_n^{(k+1)}$, то правило Зейделя достатньо записати для цього порядку таким чином:

$$x_1^{(k+1)} = \varphi_1(x_1^k, x_2^k, \dots, x_n^k);$$

$$x_2^{(k+1)} = \varphi_2(x_1^{k+1}, x_2^k, \dots, x_n^k);$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$x_n^{(k+1)} = \varphi_n(x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_{n-1}^{(k+1)}, x_n^k). \qquad (3.16)$$

Після обчислення $x_i^{(k+1)}$ $(i = 1, 2, \dots, n)$ відбувається пошук наступного наближення $x^{(k+2)}$: обирають послідовність обчислення його компонентів $x_i^{(k+2)}$ та виконують його розрахунок за допомогою рівнянь, аналогічних (3.16), і т. д.

§ 3.3. МЕТОД НЬЮТОНА РОЗВ'ЯЗАННЯ СИСТЕМ НЕЛІНІЙНИХ РІВНЯНЬ

Розглянемо, взагалі кажучи, нелінійну систему рівнянь

$$f_1(x_1, x_2, \dots, x_n) = 0;$$

$$f_2(x_1, x_2, \dots, x_n) = 0;$$

. . . . . .

$$f_n(x_1, x_2, \dots, x_n) = 0 \tag{3.17}$$

з дійсними лівими частинами, яка має у векторному запису вигляд

$$\overline{f}(\overline{x}) = 0. \tag{3.18}$$

Припустимо, що знайдено k-те наближення

$$\overline{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$$

одного з ізольованих коренів $\overline{x} = (x_1, x_2, \dots, x_n)$ рівняння (3.18). Тоді точний корінь рівняння (3.18) можна подати у вигляді

$$\overline{x} = \overline{x}^{(k)} + \overline{\varepsilon}^{(k)}, \tag{3.19}$$

де $\overline{\varepsilon}^{(k)} = (\varepsilon_1^{(k)}, \varepsilon_2^{(k)}, \dots, \varepsilon_n^{(k)})$ — похибка кореня.

Підставляючи вираз (3.19) у рівняння (3.18) матимемо

$$\overline{f}(\overline{x}^{(k)} + \overline{\varepsilon}^{(k)}) = 0. \tag{3.20}$$

Припускаючи, що функція $\overline{f}(\overline{x})$ неперервно диференційовна в деякій опуклій області G, яка містить $\overline{x}$ та $\overline{x}^{(k)}$, розкладемо ліву частину рівняння (3.20) за степенями малого параметра $\overline{\varepsilon}^{(k)}$, обмежуючись лінійними членами

$$\overline{f}(\overline{x}^{(k)} + \overline{\varepsilon}^{(k)}) = \overline{f}(\overline{x}^{(k)}) + \overline{f}'(\overline{x}^{(k)})\overline{\varepsilon}^{(k)} = 0 \tag{3.21}$$

або в розгорнутому вигляді

$$f_1(x_1^{(k)} + \varepsilon_1^{(k)}, x_2^{(k)} + \varepsilon_2^{(k)}, \dots, x_n^{(k)} + \varepsilon_n^{(k)}) =$$

$$= f_1(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) + f'_{1x_1}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_1^{(k)} +$$

$$+ f'_{1x_2}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_2^{(k)} + \dots + f'_{1x_n}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_n^{(k)} = 0;$$

$$f_2(x_1^{(k)} + \varepsilon_1^{(k)}, x_2^{(k)} + \varepsilon_2^{(k)}, \dots, x_n^{(k)} + \varepsilon_n^{(k)}) =$$

$$= f_2(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) + f'_{2x_1}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_1^{(k)} +$$

$$+ f'_{2x_2}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_2^{(k)} + \dots + f'_{2x_n}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_n^{(k)} = 0;$$

. . . . . .

$$f_n(x_1^{(k)} + \varepsilon_1^{(k)}, x_2^{(k)} + \varepsilon_2^{(k)}, \dots, x_n^{(k)} + \varepsilon_n^{(k)}) =$$

$$= f_n(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) + f'_{nx_1}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_1^{(k)} +$$

$$+ f'_{nx_2}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_2^{(k)} + \dots + f'_{nx_n}(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})\varepsilon_n^{(k)} = 0. \tag{3.21'}$$

З формул (3.21) та (3.21′) випливає, що під похідною $\overline{f}'(\overline{x})$ треба розуміти матрицю Якобі системи функцій $f_1, f_2, \ldots, f_n$ відносно змінних $x_1, x_2, \ldots \ldots, x_n$, тобто

$$\overline{f}'(\overline{x}) = F(\overline{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Система (3.21′) є лінійною системою відносно поправок $\varepsilon_i^{(k)}$ $(i = 1,2, \ldots, n)$ з матрицею $F(\overline{x})$; тому формула (3.21) може бути записана у вигляді

$$\overline{f}(\overline{x}^{(k)}) + F(\overline{x}^{(k)})\overline{\varepsilon}^{(k)} = 0.$$

Звідси, припускаючи, що матриця $F(\overline{x}^{(k)})$ невироджена, отримаємо

$$\overline{\varepsilon}^{(k)} = - F^{-1}(\overline{x}^{(k)})\overline{f}(\overline{x}^{(k)}).$$

Отже,

$$\overline{x}^{(k+1)} = \overline{x}^{(k)} - F^{-1}(\overline{x}^{(k)})\overline{f}(\overline{x}^{(k)}) \quad (k = 0, 1, 2, \ldots). \tag{3.22}$$

За початкове наближення слід обирати вектор розв'язку $\overline{x}^{(0)}$, достатньо близький до шуканого розв'язку $\overline{x}$ системи.

Оскільки при безпосередньому використанні формули (3.22) необхідно обчислювати на кожному кроці обернену матрицю до матриці Якобі, що ускладнює обчислювальний процес, зручно замість вказаної формули обчислення вести за такою схемою:

$$F(\overline{x}^{(k)})\overline{\varepsilon}^{(k)} = - \overline{f}(\overline{x}^{(k)}); \tag{3.23}$$

$$\overline{x}^{(k+1)} = \overline{x}^{(k)} + \overline{\varepsilon}^{(k)}. \tag{3.24}$$

При реалізації цієї обчислювальної схеми для кожного наближення розв'язується система лінійних рівнянь з матрицею $F(\overline{x}^{(k)})$, а потім за знайденим приростом $\overline{\varepsilon}^{(k)}$ відшукується наступне наближення $\overline{x}^{(k+1)}$.

Оскільки часто за рядом причин важко отримати в явному вигляді матрицю Якобі $F(\overline{x}^{(k)})$, при чисельній реалізації методу використовується її скінченнорізницева апроксимація, наприклад

$$(F(\overline{x}^{(k)}))_{ij} \approx \frac{f_i(\overline{x}^{(k)} + h\overline{e}_j) - f_i(\overline{x}^{(k)})}{h},$$

де h — крок диференціювання; $\overline{e}_j$ — j-й одиничний орт.

При використанні методу Ньютона припускається, що в деякій області G, яка містить розв'язок $\bar{x}^*$ системи (3.17), функції $f_i(\bar{x})$ мають неперервні похідні першого порядку і в деякому околі точки $\bar{x}^*$ матриця Якобі невироджена. Якщо $F(\bar{x})$ в деякому околі кореня $\bar{x}^*$ задовольняє умову

$$\|F(\bar{x}) - F(\bar{x}^*)\| \le L\|\bar{x} - \bar{x}^*\|,$$

то маємо квадратичну збіжність.

Недоліком методу Ньютона є те, що якщо початкове наближення $\bar{x}^{(0)}$ знаходиться далеко від розв'язку $\bar{x}^*$, то метод Ньютона найчастіше розбігається.

Існують модифікації методу Ньютона вигляду

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + t^{(k)}\bar{P}^{(k)},$$

де

$$\bar{P}^{(k)} = -F^{-1}(\bar{x}^{(k)})\bar{f}(\bar{x}^{(k)})$$

(так званий ньютонівський напрям); $t^{(k)} > 0$ — деяка довжина кроку вздовж $\bar{P}^{(k)}$. Якщо функції $f_i(\bar{x})$ неперервні та виконується умова $\|F^{-1}(\bar{x}^{(k)})\| \le C$ для будь-яких $\bar{x}$ із розглядуваної області G, то гарантується збіжність до розв'язку $\bar{x}^*$ з будь-якої початкової точки $\bar{x}_0 \in G$. При цьому поблизу від $\bar{x}^*$ модифікації повністю збігаються з традиційним методом Ньютона, що забезпечує високу швидкість збігання. Одна з перших модифікацій використовує відому властивість ньютонівського напряму $\bar{P}^{(k)}$, яка полягає в тому, що при всіх достатньо малих $t > 0$ (незалежно від вигляду норми $\|-\|$) виконується нерівність

$$\|\bar{f}(\bar{x}^{(k)} + t\bar{P}^{(k)})\| \le \|\bar{f}(\bar{x}^{(k)})\|,$$

тобто $\bar{P}^{(k)}$ — напрям спадання норми відхилу $\|\bar{f}(\bar{x})\|$.

Критерієм закінчення ітерацій для методу Ньютона та його модифікацій може бути нерівність

$$\|\bar{f}(\bar{x}^{(k)})\| \le \varepsilon, \tag{3.25}$$

де ε — спочатку дане додатне число.

Якщо отримання матриці Якобі потребує порівняно невеликих обчислювальних витрат, то метод Ньютона дуже ефективний. Коли обчислювальні витрати, необхідні для знаходження матриці Якобі, є значними, замість методу Ньютона використовуються його різні модифікації.

Приклад. Методом Ньютона наближено знайти додатний розв'язок системи

$$\begin{cases} x^2 + y^2 + z^2 = 1, \\ 2x^2 + y^2 - 4z = 0, \\ 3x^2 - 4y + z^2 = 0. \end{cases}$$

Розв'язання. На підставі початкового наближення $x_0 = y_0 = z_0 = 0{,}5$ маємо

$$\bar{f}(\bar{x}) = \begin{pmatrix} x^2 + y^2 + z^2 - 1, \\ 2x^2 + y^2 - 4z, \\ 3x^2 - 4y + z^2. \end{pmatrix}$$

Звідси

$$\bar{f}(\bar{x}^{(0)}) = \begin{pmatrix} 0{,}25 + 0{,}25 + 0{,}25 - 1 \\ 0{,}50 + 0{,}25 - 2{,}00 \\ 0{,}75 - 2{,}00 + 0{,}25 \end{pmatrix} = \begin{pmatrix} -0{,}25 \\ -1{,}25 \\ -1{,}00 \end{pmatrix}.$$

Побудуємо матрицю Якобі

$$F(\bar{x}) = \begin{pmatrix} 2x & 2y & 2z \\ 4x & 2y & -4 \\ 6x & -4 & 2z \end{pmatrix}.$$

Дістанемо

$$F(\bar{x}^{(0)}) = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -4 \\ 3 & -4 & 1 \end{pmatrix}.$$

Відповідно з (3.23) маємо систему рівнянь

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -4 \\ 3 & -4 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_x^{(0)} \\ \varepsilon_y^{(0)} \\ \varepsilon_z^{(0)} \end{pmatrix} = \begin{pmatrix} 0{,}25 \\ 1{,}25 \\ 1{,}00 \end{pmatrix};$$

розв'язуючи її методом виключення Гаусса, отримуємо:

$$\varepsilon_x^{(0)} = 0{,}375; \quad \varepsilon_y^{(0)} = 0; \quad \varepsilon_z^{(0)} = -0{,}125.$$

За знайденим приростом дістанемо перше наближення (3.24):

$$\bar{x}^{(1)} = \bar{x}^{(0)} + \bar{\varepsilon}^{(0)};$$

$$\bar{x}^{(1)} = \begin{pmatrix} 0{,}875 \\ 1{,}5 \\ 1{,}375 \end{pmatrix}.$$

Далі обчислюємо друге наближення $\bar{x}^{(2)}$. Маємо

$$\bar{f}(\bar{x}^{(1)}) = \begin{pmatrix} 0{,}15625 \\ 0{,}28125 \\ 0{,}43750 \end{pmatrix}; \quad F(\bar{x}^{(1)}) = \begin{pmatrix} 1{,}750 & 1 & 0{,}750 \\ 3{,}500 & 1 & -4 \\ 5{,}250 & -4 & 0{,}750 \end{pmatrix}.$$

Розв'язуючи систему рівнянь

$$F(\bar{x}^{(1)})\bar{\varepsilon}^{(1)} = -f(x^{(1)}),$$

отримуємо

$$\varepsilon_x^{(1)} = -0{,}085183; \quad \varepsilon_y^{(1)} = -0{,}0033784; \quad \varepsilon_z^{(1)} = -0{,}0050675.$$

Використовуючи знайдений приріст, будуємо друге наближення:

$$\bar{x}^{(2)} = \bar{x}^{(1)} + \bar{\varepsilon}^{(1)};$$

$$\bar{x}^{(2)} = \begin{pmatrix} 0{,}78982 \\ 0{,}49662 \\ 0{,}36993 \end{pmatrix}.$$

Аналогічно відшукуються подальші наближення:

$$\bar{x}^{(3)} = \begin{pmatrix} 0{,}78521 \\ 0{,}49662 \\ 0{,}36992 \end{pmatrix}; \quad \bar{f}(\bar{x}^{(3)}) = \begin{pmatrix} 0{,}00001 \\ 0{,}00004 \\ 0{,}00005 \end{pmatrix}$$

і т. д.

Обмежуючись третім наближенням, дістаємо:

$$x = 0{,}7852; \; y = 0{,}4966; \; z = 0{,}3699.$$

§ 3.4. МЕТОДИ КВАЗІНЬЮТОНІВСЬКОГО ТИПУ

Одним з недоліків Ньютона є необхідність обчислювати матрицю Якобі та розв'язувати систему лінійних алгебраїчних рівнянь. Це потребує значних витрат машинних дій, обсяг яких різко зростає із зростанням розмірності системи (3.17). Тому були розроблені модифікації методу Ньютона, в яких протягом ітераційного процесу замість побудови самої матриці Якобі або її оберненої будується їх апроксимація. Це дозволяє суттєво скоротити число арифметичних дій на ітерації. Такі методи розв'язку систем нелінійних рівнянь отримали назву *квазіньютонівських*. Більшість відомих квазіньютонівських методів збігається локально з надлінійною швидкістю збіжності при тих самих припущеннях про властивості функцій $\bar{f}_i(\bar{x})$, які були зроблені при використанні методу Ньютона, що має квадратичну швидкість збіжності. Квазіньютонівські методи можна поділити на два тісно пов'язаних між собою класи методів у залежності від того, що апроксимується — матриця Якобі або їй обернені.

Розглянемо перший із класів, де матриця B_k з розмірами $n \times n$ апроксимує матрицю $F(\bar{x}^{(k)})$. Перед початком ітерацій задають початкову точку $\bar{x}^{(0)}$, а матрицю B_0 звичайно отримують, або припускаючи, що вона є одиничною, або апроксимуючи $F(\bar{x}^{(k)})$ за скінченнорізницевими формулами. Потім для $k = 0, 1, \ldots$ обчислюють

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} - B_k^{-1}\bar{f}(\bar{x}^{(k)}); \tag{3.26}$$

$$\Delta\bar{x}^{(k)} = \bar{x}^{(k+1)} - \bar{x}^{(k)};$$

$$\Delta\bar{f}^{(k)} = \bar{f}(\bar{x}^{(k+1)}) - \bar{f}(\bar{x}^{(k)});$$

$$B_{k+1} = B_k + \frac{(\Delta\bar{f}^{(k)} - B_k\Delta x^{(k)})\bar{C}_k^T}{((\Delta x^{(k)})^T\bar{C}_k)}, \tag{3.27}$$

де $\bar{C}_k$ — n-вимірний вектор, що є параметром розглядуваного класу методів. Якщо $\bar{C}_k$ взяти таким, що дорівнює $\Delta x^{(k)}$, то матимемо перший метод Бройдена. Вибір $\bar{C}_k = \Delta\bar{f}^{(k)}$ відповідає методу Пірсона, а $\bar{C}_k = \Delta\bar{f}^{(k)} - B_k\Delta x^{(k)}$ — симетричному методу першого рангу.

У другому з розглядуваних тут класів квазіньютонівських методів матриця H_k з розмірами $n \times n$ апроксимує матрицю $F^{-1}(\bar{x}^{(k)})$. Перед початком ітерацій задають початкову точку $\bar{x}^{(0)}$ і матрицю H_0, яка звичайно або дорівнює одиничній, або є оберненою до скінченнорізницевої апроксимації матриці $F(\bar{x}^{(k)})$. Потім обчислюють

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + H_kF(\bar{x}^{(k)}); \tag{3.28}$$

$$H_{k+1} = H_k + \frac{(\Delta\bar{x}^{(k)} - H_k\Delta\bar{f}^{(k)})\bar{\alpha}_k^T}{((\Delta\bar{f}^{(k)T}\bar{\alpha}_k)}, \tag{3.29}$$

де $\bar{\alpha}_k$ — n-вимірний вектор, який є параметром розглядуваного класу методів. Конкретний вигляд вектора $\bar{\alpha}_k$ відповідає відповідному методу: наприклад, $\bar{\alpha}_k = \Delta\bar{f}^{(k)}$ — другому методу Бройдена, $\bar{\alpha}_k = \Delta\bar{x}^{(k)}$ — методу Мак-Корміка.

Зауважимо, що якщо задати B_0^{-1}, то можна вести перерахунок не B_k, а матриць B_k^{-1} за формулою

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(\Delta\bar{x}^{(k)} - B_k^{-1}\Delta\bar{f}^{(k)})\bar{C}_k^TB_k^{-1}}{(\bar{C}_k^TB_k^{-1}\Delta\bar{f}^{(k)})}, \tag{3.30}$$

еквівалентною (3.27). Це потребує порядку $0(n^2)$ арифметичних дій замість $0(n^3)$, необхідних для розв'язання системи лінійних рівнянь $B_k\Delta\bar{x}^{(k)} = -\bar{f}(\bar{x}^{(k)})$.

Як видно з (3.30), між формулами (3.27) та (3.29) має місце певний зв'язок. Так, якщо $H_k = B_k^{-1}$, то $H_{k+1} = B_{k+1}^{-1}$ при $\overline{C}_k = B_k^T \overline{\alpha}_k$. Таким чином, один і той самий метод може реалізуватися двома різними формулами (3.27) і (3.29), які еквівалентні теоретично, але їх чисельна реалізація може відрізнятися за ефективністю.

Розглянемо, наприклад, перший метод Бройдена. Його можна реалізувати за формулою (3.27) так, що це потребує загалом $0(n^2)$ арифметичних дій. Це виявляється можливим, якщо подати матрицю B_k у вигляді добутку $Q_k R_k$, де Q_k — ортогональна, а R_k — верхня трикутна матриця. Дійсно, у цьому разі розв'язання системи потребує тільки $0(n^3)$ арифметичних дій. Маючи ж $B_k = Q_k R_k$, на подання матриці B_{k+1}, яка задовольняє (3.27) у вигляді $Q_{k+1} R_{k+1}$, необхідно $0(n^2)$ арифметичних дій. Важлива перевага формули (3.27) перед (3.39) полягає в тому, що в (3.27) немає необхідності множення матриці на вектор, оскільки

$$\Delta \overline{f}^{(k)} - B_k \Delta \overline{x}^{(k)} = \overline{f}(\overline{x}^{(k+1)}).$$

Існують квазіньютонівські методи, які враховують симетричність матриці Якобі і виробляють послідовність симетричних матриць B_k (або H_k). Ці методи також можна поділити на два класи. У першому з них матриця B_k апроксимує $F(x)$. На відміну від описаного вище класу, що задається формулами (3.26) і (3.27), тут потрібна симетричність матриці B_0, і замість (3.27) використовується формула

$$J_{k+1} = J_k + \frac{(\Delta \overline{f}^{(k)} - B_k \Delta \overline{x}^{(k)}) \overline{C}_k^T + \overline{C}_k (\Delta \overline{f}^{(k)} - B_k \Delta \overline{x}^{(k)})^T}{((\Delta x^{(k)})^T \overline{C}_k} -$$

$$- \frac{((\Delta \overline{f}^{(k)} - B_k \Delta \overline{x}^{(k)})^T \Delta \overline{x}^{(k)} \overline{C}_k \overline{C}_k^T}{((\Delta \overline{x}^{(k)})^T \overline{C}_k)^2}, \qquad (3.31)$$

де значення параметра $\overline{C}_k = \Delta \overline{x}^{(k)}$ відповідає симетричному варіанту Пауелла методу Бройдена, а $\overline{C}_k = \Delta \overline{f}^{(k)}$ — методу Давідона—Флечера—Пауелла.

У другому із розглядуваних класів квазіньютонівських методів матриця H_k апроксимує матрицю $F^{-1}(\overline{x}^{(k)})$. Тут матриця H_0 повинна бути симетричною, а замість (3.29) використовується формула

$$H_{k+1} = H_k + \frac{(\Delta \overline{x}^{(k)} - H_k \Delta \overline{f}^{(k)}) \overline{\alpha}_k^T + \overline{\alpha}_k (\Delta \overline{x}^{(k)} - H_k \Delta \overline{f}^{(k)})^T}{((\Delta \overline{f}^{(k)})^T \overline{\alpha}_k} -$$

$$- \frac{((\Delta \overline{x}^{(k)} - H_k \Delta \overline{f}^{(k)})^T \Delta \overline{f}^{(k)} \overline{\alpha}_k \overline{\alpha}_k^T}{((\Delta \overline{f}^{(k)})^T \overline{\alpha}_k)^2}, \qquad (3.32)$$

де $\alpha_k = \Delta\bar{x}_k$ відповідає методу Бройдена—Флечера—Гольдфарба—Шенно, що є одним з найкращих (з обчислювальної точки зору), який враховує симетричність матриці Якобі.

Зазначимо, що формула (3.31) еквівалентна формулі

$$B_{k+1}^{-1} = B_k^{-1} + [(\overline{C}_k^T B_k^{-1}\Delta\bar{f}^{(k)})(B_k^{-1}\overline{C}_k(\Delta\bar{x}^{(k)} - B_k^{-1}\Delta\bar{f}^{(k)})^T + (\Delta\bar{x}^{(k)} -$$

$$- B_k^{-1}\Delta\bar{f}^{(k)})\overline{C}_k^T B_k^{-1}) - ((\Delta\bar{f}^{(k)})^T(\Delta\bar{x}^{(k)} - B_k^{-1}\Delta\bar{f}^{(k)}))B_k^{-1}\overline{C}_k\overline{C}_k^T B_k^{-1} +$$

$$+ (\overline{C}_k^T B_k^{-1}\overline{C}_k)(\Delta\bar{x}^{(k)} - B_k^{-1}\Delta\bar{f}^{(k)})(\Delta\bar{x}^{(k)} - B_k^{-1}\Delta\bar{f}^{(k)})^T] /$$

$$/ [(\overline{C}_k^T B_k^{-1}\Delta\bar{f}^{(k)})^2 + (\overline{C}_k^T B_k^{-1}\Delta\bar{f}^{(k)})((\Delta\bar{f}^{(k)})^T(\Delta\bar{x}^{(k)} - B_k^{-1}\Delta\bar{f}^{(k)}))],$$

яка виключає необхідність розв'язку системи лінійних рівнянь. Що ж до зв'язку між формулами (3.31) і (3.32), то вони не еквівалентні.

Описані вище квазіньютонівські методи збігаються лише при достатньо доброму початковому наближенні $\bar{x}^{(0)}$. Для розширення області їх збіжності можна використати прийом, який має назву *одномірного пошуку.*

Нехай маємо квазіньютонівський напрям $\bar{P}_k = -B_k^{-1}\bar{f}(\bar{x}^{(k)})$ (або $\bar{P}_k = -H_k\bar{f}(\bar{x}^{(k)})$). Використаємо довжину кроку $t_k = 1$ та перевіримо нерівність

$$\|\bar{f}(\bar{x}^{(k)} + t_k\bar{P}_k\| \le \|\bar{f}(\bar{x}^{(k)})\|, \tag{3.34}$$

де $\|-\|$ — евклідова норма. Якщо воно виконується, то закінчуємо одновимірний пошук та вважаємо

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + t_k\bar{P}_k, \tag{3.35}$$

тобто зменшуємо довжину кроку (встановлюючи, наприклад, $t_k = t_k/2$), доки не виконається (3.34). На цьому закінчуємо одновимірний пошук і переходимо до формули (3.35).

Як видно, одновимірний пошук (в разі успіху) забезпечує монотонне зменшення норми відхилу $\|\bar{f}(\bar{x}^{(k)})\|$ із зростанням k. Якщо квазіньютонівський напрям $\bar{P}_k$ сильно відрізняється від ньютонівського, то одновимірний пошук може виявитися невдалим, і тоді необхідно поновити матрицю B_k (або H_k), прирівнявши її, наприклад, скінченнорізницевій апроксимації матриці Якобі $F(\bar{x}^{(k)})$ (або $F^{-1}(\bar{x}^{(k)})$). Критерієм закінчення ітерацій для квазіньютонівських методів є нерівність (3.25).

Чисельний розв'язок нелінійних алгебраїчних рівнянь є складною і не до кінця розв'язаною задачею обчислювальної математики. Для розв'язання систем нелінійних рівнянь можна використати метод Ньютона, квазіньютонівські методи та ін. Проста ітерація має лінійну збіжність, метод Ньютона — квадратичну, а квазіньютонівські — надлінійну швидкість збігання. Незважаючи на те, що квазіньютонівські методи мають гіршу порівняно з методом Ньютона теоретичну збіжність, вони потребу-

ють при своїй реалізації меншої кількості машинних дій і в багатьох випадках класичного методу Ньютона. Проте всі ці методи мають локальну збіжність, тобто збіжність при доброму початковому наближенні. Для отримання цього початкового наближення під час розв'язання систем нелінійних рівнянь використовують ті чи інші методи спуску і комбінують їх з методами, які мають більшу швидкість збігання.

Зауважимо, що розв'язання системи нелінійних рівнянь може бути зведене до задачі мінімізації функції. Задача пошуку мінімуму функції n змінних є окремим випадком розв'язання екстремальних задач.

§ 3.5. МЕТОД ПОКООРДИНАТНОГО СПУСКУ

Розглянемо нелінійну систему рівнянь вигляду (3.17). Припустимо, що функції f_i дійсні та неперервно диференційовні в їх загальній області визначення. Розглянемо функцію

$$U(\bar{x}) = \sum_{i=1}^{n} |f_i(\bar{x})|^2 = (\bar{f}(\bar{x})\bar{f}(\bar{x})). \qquad (3.36)$$

Очевидно, що кожний розв'язок системи (3.17) перетворює на нуль функцію $U(\bar{x})$; навпаки, числа $x_1, x_2, \ldots, x_n$, для яких функція $U(\bar{x})$ дорівнює нулю, є коренями системи (3.17).

Ідея всіх методів спуску полягає в тому, що згідно з початковим наближенням — точки $\bar{x}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\} \in D$ — перейти в точку $\bar{x}^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}\} \in D$ так, щоб значення $U(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$ зменшилось $U(\bar{x}^{(1)}) < U(\bar{x}^{(0)})$. Вважатимемо, що система (3.17) має лише ізольований розв'язок, який є точкою строгого мінімуму функції $U(\bar{x})$. Таким чином, задача зводиться до відшукування мінімуму функції $U(\bar{x})$ в n-вимірному просторі $E^n = \{x_1, x_2, \ldots, x_n\}$.

Нехай $\bar{x}$ — вектор-корінь системи (3.17) і $\bar{x}^{(0)}$ — його нульове наближення. Через точку $\bar{x}^{(0)}$ проведемо поверхню рівня функції $U(\bar{x})$. Якщо точка $\bar{x}^{(0)}$ достатньо близька до кореня $\bar{x}$, то при наших припущеннях поверхня рівня $U(\bar{x}) = U(\bar{x}^{(0)})$ буде схожа на еліпсоїд.

З точки $\bar{x}^{(0)}$ рухаємось по нормалі до поверхні $U(\bar{x}) = U(\bar{x}^{(0)})$ доти, поки ця нормаль не торкнеться в деякій точці $\bar{x}^{(1)}$ якоїсь іншої поверхні рівня $U(\bar{x}) = U(\bar{x}^{(1)})$. Відштовхуючись від точки $\bar{x}^{(1)}$ знову зсуваємося по нормалі до поверхні рівня $U(\bar{x}) = U(\bar{x}^{(1)})$ доти, поки ця нормаль не торкнеться в деякій точці $\bar{x}^{(2)}$ нової іншої поверхні $U(\bar{x}) = U(\bar{x}^{(2)})$ і т.д. (рис. 3).

Оскільки $U(\bar{x}^{(0)}) > U(\bar{x}^{(1)}) > U(\bar{x}^{(2)}) > \ldots$, то рухаючись таким шляхом, швидко можна наблизитись до точки з найменшим значенням U, яка відповідає шуканому кореню $\bar{x}$ системи (3.17).

Нагадаємо, що градієнт функції $U(\bar{x})$ визначається формулою

$$\text{grad } U(\bar{x}) = \left(\frac{\partial U}{\partial x_1}, \frac{\partial U}{\partial x_2}, \ldots, \frac{\partial U}{\partial x_n}\right).$$

Вектор grad $U(\bar{x})$ ортогональний лініям рівня $U(\bar{x}) = C = \text{const}$, його напрям збігається з напрямом максимального зростання $U(\bar{x})$ у заданій точці. У точці мінімуму функції grad $U(\bar{x}) = 0$.

Визначимо ітераційний процес

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} - h_k \text{ grad } U(\bar{x}^{(k)}), \qquad (3.37)$$

де h_k — крок спуску на k-й ітерації; $\bar{x}^{(0)}$ — задане початкове наближення до точки мінімуму. При $h_k = h = \text{const}$ формула (3.37) являє собою метод градієнтного спуску з постійним кроком. Ітераційний процес (3.37) продовжується до виконання якої-небудь умови закінчення алгоритму, наприклад

$$|U(\bar{x}^{(k+1)}) - U(\bar{x}^{(k)})| < \varepsilon$$

або

$$\|\text{grad } U(\bar{x}^{(k+1)})\| < \varepsilon,$$

де ε — задана точність.

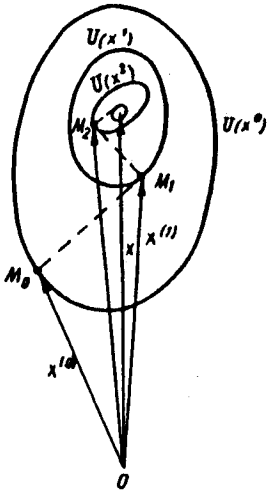


Рис. 3

Для випадку змінного кроку залишається визначити множник h_k. Для цього розглянемо скалярну функцію $\Phi(h) = U[\bar{x}^{(k)} - h \text{ grad } U(\bar{x}^{(k)})]$.

Функція $\Phi(h)$ дає зміну рівня функції U вздовж відповідної нормалі до поверхні рівня в точці $\bar{x}^{(k)}$. Множник $h = h_k$ треба обирати таким чином, щоб $\Phi(h)$ мала мінімум. Беручи похідну по h і прирівнюючи її до нуля, дістанемо рівняння

$$\Phi'(h) = \frac{d}{dh} U[\bar{x}^{(k)} - h \text{ grad } U(\bar{x}^{(k)})] = 0. \qquad (3.38)$$

Найменший додатний корінь рівняння (3.38) і дає нам значення h_k. Рівняння (3.38) розв'язується чисельно, тому вкажемо метод чисельного відшукування чисел h_k. Вважатимемо, що h — мала величина, квадратами і вищими степенями якої можна знехтувати. Маємо

$$\Phi(h) = \sum_{i=1}^{n} \{f_i[\bar{x}^{(k)} - h \text{ grad } U(\bar{x}^{(k)})]^2\}.$$

Розкладаючи функції f_i за степенями h з точністю до лінійних членів, дістанемо

$$\Phi(h) = \sum_{i=1}^{n} \left[f_i(\bar{x}^{(\kappa)}) - h \frac{\partial f_i(\bar{x}^{(k)})}{\partial \bar{x}} \operatorname{grad} U(\bar{x}^{(k)}) \right]^2,$$

де

$$\frac{\partial f_i}{\partial \bar{x}} = \left\{ \frac{\partial f_i}{\partial x_1}, \frac{\partial f_i}{\partial x_2}, \ldots, \frac{\partial f_i}{\partial x_n} \right\}.$$

Звідси

$$\Phi'(h) = -2 \sum_{i=1}^{n} \left[f_i(\bar{x}^{(k)}) - h \frac{\partial \bar{f}_i(\bar{x}^{(k)})}{\partial \bar{x}} \operatorname{grad} U(\bar{x}^{(k)}) \right] \times$$

$$\times \frac{\partial \bar{f}_i(\bar{x}^{(k)})}{\partial \bar{x}} \operatorname{grad} U(\bar{x}^{(k)}) = 0.$$

Отже,

$$h_k = \frac{\sum_{i=1}^{n} f_i(\bar{x}^{(k)}) \frac{\partial \bar{f}_i(\bar{x}^{(k)})}{\partial \bar{x}} \operatorname{grad} U(\bar{x}^{(k)})}{\sum_{i=1}^{n} \left[\frac{\partial \bar{f}_i(\bar{x}^{(k)})}{\partial \bar{x}} \operatorname{grad} U(\bar{x}^{(k)}) \right]^2} =$$

$$= \frac{(\bar{f}(\bar{x}^{(k)}) \cdot F(\bar{x}^{(k)}) \operatorname{grad} U(\bar{x}^{(k)}))}{(F(\bar{x}^{(k)}) \operatorname{grad} U(\bar{x}^{(k)}) \cdot F(\bar{x}^{(k)}) \operatorname{grad} U(\bar{x}^{(k)}))},$$

де

$$F(x) = \frac{\partial \bar{f}}{\partial \bar{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

— матриця Якобі вектор-функції $\bar{f}$.

Далі маємо

$$\frac{\partial U}{\partial x_j} = \frac{\partial}{\partial x_j} \left\{ \sum_{i=1}^{n} |f_i(\bar{x})|^2 \right\} = 2 \sum_{i=1}^{n} f_i(\bar{x}) \frac{\partial f_i(\bar{x})}{\partial x_j}.$$

Звідси

$$\operatorname{grad} U(\bar{x}) = 2\begin{bmatrix} \sum\limits_{i=1}^{n} f_i(\bar{x})\dfrac{\partial f_i(\bar{x})}{\partial x_1} \\ \cdots\cdots\cdots \\ \sum\limits_{i=1}^{n} f_i(\bar{x})\dfrac{\partial f_i(\bar{x})}{\partial x_n} \end{bmatrix} = 2F^T(\bar{x})\bar{f}(\bar{x});$$

тут $F^T(\bar{x})$ — транспонована матриця Якобі. Тому, як висновок,

$$2h_k = \frac{(\bar{f}^{(k)} \cdot F_k F_k^T \bar{f}^{(k)})}{(F_k F_k^T \bar{f}^{(k)} \cdot F_k F_k^T \bar{f}^{(k)})}, \qquad (3.39)$$

де для того, щоб формула була короткою, прийнято $\bar{f}^{(k)} = \bar{f}(\bar{x}^{(k)}) \cdot F_k = F(\bar{x}^{(k)})$. Причому

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} - 2h_k F_k^T \bar{f}^{(k)} \quad (k = 0, 1, 2, \dots). \qquad (3.40)$$

Приклад. Методом градієнтного спуску приблизно обчислити корені системи

$$x + x^2 - 2yz = 0{,}1;$$

$$y - y^2 + 3xz = -0{,}2;$$

$$z + z^2 + 2xy = 0{,}3,$$

розташовані в околі початку координат. Тут

$$\bar{f} = \begin{bmatrix} x + x^2 - 2yz - 0{,}1 \\ y - y^2 + 3xz + 0{,}2 \\ z + z^2 + 2xy - 0{,}3 \end{bmatrix};$$

$$F = \begin{bmatrix} 1 + 2x & -2z & -2y \\ 3z & 1 - 2y & 3x \\ 2y & 2x & 1 + 2z \end{bmatrix}.$$

Розв'язання. Підставляючи нульове наближення, матимемо

$$\bar{f}^{(0)} = \begin{bmatrix} -0{,}1 \\ 0{,}2 \\ -0{,}3 \end{bmatrix}; \quad F_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E.$$

За формулами (3.39), (3.40) отримуємо перше наближення:

$$2h_0 = \frac{(\bar{f}^{(0)} \cdot \bar{f}^{(0)})}{(\bar{f}^{(0)} \cdot \bar{f}^{(0)})} = 1;$$

$$\bar{x}^{(1)} = \bar{x}^{(0)} - 1 * E\bar{f}^{(0)} = \begin{bmatrix} 0,1 \\ -0,2 \\ 0,3 \end{bmatrix}.$$

Аналогічно шукаємо друге наближення $\bar{x}^{(1)}$. Маємо:

$$\bar{f}^{(1)} = \begin{bmatrix} 0,13 \\ 0,05 \\ 0,05 \end{bmatrix}; \quad F_1 = \begin{bmatrix} 1,2 & -0,6 & 0,4 \\ 0,9 & 1,4 & 0,3 \\ -0,4 & 0,2 & 1,6 \end{bmatrix}.$$

Звідси

$$F_1^T \bar{f}^{(1)} = \begin{bmatrix} 0,181 \\ 0,002 \\ 0,147 \end{bmatrix}; \quad F_1 F_1^T \bar{f}^{(1)} = \begin{bmatrix} 0,2748 \\ 0,2098 \\ 0,1632 \end{bmatrix}.$$

Отже,

$$2h_1 = \frac{0,13 \cdot 0,2748 + 0,05 \cdot 0,2098 + 0,05 \cdot 0,1632}{0,2748^2 + 0,2098^2 + 0,1632^2} = 0,3719;$$

$$\bar{x}^{(2)} = \begin{bmatrix} 0,0327 \\ -0,2007 \\ 0,2453 \end{bmatrix}$$

і т. д.

§ 3.6. МЕТОД ПРОДОВЖЕННЯ РОЗВ'ЯЗАННЯ ПО ПАРАМЕТРУ

Розглянутий в § 3.3 ітераційний процес Ньютона розв'язання системи нелінійних алгебраїчних або трансцендентних рівнянь суттєво залежить від вибору початкового наближення. Якщо початковий вектор обраний недостатньо близько до шуканого розв'язку системи, то ітераційний процес або може збігатися дуже повільно, або взагалі розбігатися. Але оскільки розв'язок системи невідомий і в багатьох випадках неможливо вказати обмежену область, якій належить шуканий розв'язок, то проблема вибору початкового наближення, яке забезпечує збіжність ітераційного процесу, у вказаних випадках утруднює використання методу Ньютона. Для подолання вищезгаданих труднощів використовують метод продовження розв'язання по параметру.

Розглянемо цей метод. Замість вихідної системи нелінійних рівнянь $\bar{f}(\bar{x}) = 0$ взято систему рівнянь з параметром λ у явному вигляді

$$\bar{\psi}(\bar{x}, \lambda) = 0. \tag{3.41}$$

При цьому потрібно, щоб при $\lambda = 0$ розв'язок системи $\bar{\psi}(\bar{x}, \lambda) = 0$ шукався просто, а при $\lambda = 1$ система $\bar{\psi}(\bar{x}, 1) = 0$ була еквівалентною вихідній системі $\bar{f}(\bar{x}) = 0$.

Припустимо, що вектор-функція $\bar{\psi}(\bar{x}, \lambda)$ неперервна й диференційовна по λ потрібну кількість разів, а рівняння $\bar{\psi}(\bar{x}, \lambda) = 0$ має розв'язок при всіх значеннях λ в інтервалі $[0, 1]$.

Нехай розв'язок системи $\bar{\psi}(\bar{x}, 0) = 0$ дорівнює $\bar{x}^0$, а розв'язок системи $\bar{\psi}(\bar{x}, 1) = 0$ — $\bar{x}^*$, тобто $\bar{f}(\bar{x}^*) = 0$. Тоді система (3.41) з вказаними вимогами визначає вектор-функцію $\bar{x} = \bar{x}(\lambda)$, що залежить від аргументу λ, яка задовольняє умови $\bar{x}(0) = \bar{x}^{(0)}$, $\bar{x}(1) = \bar{x}^*$.

Таким чином, для вектор-функції $\bar{x}(\lambda)$ може бути сформульована задача Коші з початковою умовою $\bar{x}(0) = \bar{x}^{(0)}$, а значення цієї функції при $\lambda = 1$ $\bar{x}(1) = \bar{x}^*$ є шуканим розв'язком вихідної нелінійної системи рівнянь $\bar{f}(\bar{x}) = 0$.

Для отримання диференціального рівняння, що описує цю задачу Коші, продиференціюємо (3.41) по параметру λ. Дістанемо

$$F(\bar{x}, \lambda) \frac{d\bar{x}}{d\lambda} = -\frac{\partial \bar{\psi}}{\partial \lambda}, \tag{3.42}$$

де $F(\bar{x}, \lambda)$ — матриця Якобі вектор-функції $\bar{\psi}(\bar{x}, \lambda)$. Задача Коші для вектор-функції $\bar{x} = \bar{x}(\lambda)$ визначається диференціальним рівнянням (3.42) та початковою умовою

$$\bar{x}(0) = \bar{x}^0. \tag{3.43}$$

У загальному випадку система диференціальних рівнянь (3.42) нелінійна.

Для введення параметра λ, тобто побудови системи (3.41), є декілька способів. Розглянемо один з найбільш простих.

Задамо довільно вектор $\bar{x}^0$ та обчислимо значення вектор-функції $\bar{f}(\bar{x}^0) = \bar{f}^0$. Вектор-функцію $\bar{\psi}(\bar{x}, \lambda)$ будуємо у вигляді

$$\bar{\psi}(\bar{x}, \lambda) = \bar{f}(\bar{x}) - (1 - \lambda)\bar{f}^0. \tag{3.44}$$

Легко перевірити виконання необхідних вимог. При $\lambda = 0$: $\bar{\psi}(\bar{x}, 0) = \bar{f}(\bar{x}) - \bar{f}^0$; $\bar{f}(\bar{x}) = \bar{f}^0$; $\bar{x} = \bar{x}^0$.

При $\lambda = 1$: $\bar{\psi}(\bar{x}, 1) = \bar{f}(\bar{x})$; $\bar{f}(\bar{x}) = 0$; $\bar{x} = \bar{x}^*$.

Задача Коші для системи диференціальних рівнянь (3.42) для вектор-функції (3.44) матиме вигляд

$$F(\bar{x}) \frac{d\bar{x}}{d\lambda} = -\bar{f}^0, \quad \bar{x}(0) = \bar{x}^{(0)}, \tag{3.45}$$

де $F(\bar{x})$ — матриця Якобі вектор-функції $\bar{f}(\bar{x})$.

Розглянемо ще один спосіб побудови системи (3.41), яка містить параметр λ. Зобразимо шукану вектор-функцію:

$$\bar{\psi}(\bar{x}, \lambda) = (1 - \lambda)(D\bar{x} - \bar{p}) + \lambda\bar{f}(\bar{x}), \tag{3.46}$$

де D і $\bar{p}$ — задані квадратна матриця та вектор вимірності n. Перевіримо виконання необхідних вимог. При $\lambda = 1$: $\bar{\psi}(\bar{x}, 1) = \bar{f}(\bar{x})$. Розв'язок лінійної системи $D\bar{x} = \bar{p}$ не складний; тому вважаємо, що воно знайдено й дорівнює $\bar{x} = \bar{x}^0$.

Матрицю D і вектор $\bar{p}$ можна задати на основі різних міркувань: зокрема, вектор-функцію $D\bar{x} - \bar{p}$ апроксимувати вектор-функцією $\bar{f}(\bar{x})$ в околі розв'язку системи $\bar{f}(\bar{x}) = 0$. Для цього можемо використати метод найменших квадратів. Обираємо m довільних значень $\bar{x}^{(i)}$ $(i = 1, 2, \ldots, m, m \geq n + 1)$ в околі розв'язку системи $\bar{f}(\bar{x}) = 0$. Позначимо через $\bar{\omega}_i$ $(i = 1, 2, \ldots, n)$ вектори-стовпці матриці D, тобто $D = (\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_n)$. Елементи векторів $\bar{\omega}_j$ позначимо через ω_{ij} $(i, j = 1, 2, \ldots, n)$. Тоді величини ω_{ij}, p_i можна знайти на підставі умов

$$\sum_{k=1}^{m}\left[\sum_{i=1}^{n}\omega_{ij}x_j^{(k)} - p_i - f_i(\bar{x}^{(k)})\right]^2 = \min \quad (i = 1, 2, \ldots, n). \tag{3.47}$$

При $m = n + 1$ поліноми першого степеня $\sum_{i=1}^{n}\omega_{ij}x_i - p_i$ будуть інтерполяційними поліномами. Із умови (3.47) дістанемо n лінійних систем з $n + 1$ невідомими, кожне для визначення $n(n + 1)$ коефіцієнтів ω_{ij}, p_i.

Розглянемо окремий випадок вибору точок $\bar{x}^{(i)}$, а саме, коли $\bar{x}^{(i)} = \bar{x}_0 + h\bar{e}_i$ $(i = 1, 2, \ldots, n)$, де $\bar{x}_0$ — задана точка, h — додатне число, $\bar{e}_i$ — i-й координатний орт.

Вектори $\bar{\omega}_i$, $\bar{p}_i$ шукаємо із системи

$$D(\bar{x}_0 + h\bar{e}_i) - \bar{p} = \bar{f}(\bar{x}_0 + h\bar{e}_i); \quad (i = 1, 2, \ldots, n);$$

$$D\bar{x}_0 - \bar{p} = \bar{f}(\bar{x}_0).$$

Віднімаючи із перших систем останню, маємо

$$\bar{\omega}_i = D\bar{e}_i = \frac{\bar{f}(\bar{x}_0 + h\bar{e}_i) - \bar{f}(\bar{x}_0)}{h} \quad (i = 1, 2, \ldots, n). \tag{3.48}$$

У цьому разі матриця D збігається з дискретним аналогом матриці Якобі $F_h(\bar{x})$, тільки в матриці D у виразі (3.48) h не обов'язково повинна бути малою.

Вектор $\bar{p}$ визначається із виразу

$$\bar{p} = D\bar{x}_0 - \bar{f}(\bar{x}_0). \tag{3.49}$$

Крім вказаного вибору точок, можна також використати такий: $\bar{x}_0$, $\bar{x}_0 + h\bar{e}_i$, $\bar{x}_0 - h\bar{e}_i$ $(i = 1, 2, \ldots, n)$. Дістанемо системи:

$$D(\bar{x}_0 + h\bar{e}_i) - \bar{p} = \bar{f}(\bar{x}_0 + h\bar{e}_i);$$

$$D(\bar{x}_0 - h\bar{e}_i) - \bar{p} = \bar{f}(\bar{x}_0 - h\bar{e}_i) \quad (i = 1, 2, \ldots, n).$$

Звідки

$$\bar{\omega}_i = D\bar{e}_i = \frac{\bar{f}(\bar{x}_0 + h\bar{e}_i) - \bar{f}(\bar{x}_0 - h\bar{e}_i)}{2h} \quad (i = 1, 2, \ldots, n). \quad (3.50)$$

Вектор $\bar{p}$ знаходимо за формулою (3.49). Необхідно зауважити, що вектор-функція $D\bar{x} - \bar{p}$ з матрицею (3.50) у багатьох випадках більш точно апроксимує вектор-функцію $\bar{f}(\bar{x})$, ніж з матрицею (3.48). Але при цьому необхідно майже в два рази більше обчислень, ніж для побудови матриці (3.48). В разі наближення величини h до нуля в (3.48) і (3.50) матриця D наближається до матриці Якобі $F(\bar{x}_0)$.

Для вектор-функції $\bar{\psi}(\bar{x}, \lambda)$, яка визначається за виразом (3.46), задача Коші формулюється для рівняння

$$[(1 - \lambda)D + F(\bar{x})]\frac{d\bar{x}}{d\lambda} = D\bar{x} - \bar{p} - \bar{f}(\bar{x}). \quad (3.51)$$

Початкова умова визначається розв'язком системи $D\bar{x} = \bar{p}$, тобто $\bar{x}(0) = \bar{x}^0$.

Для отримання вектор-функції $\bar{x} = \bar{x}(\lambda)$ необхідно знайти розв'язок задачі Коші для рівняння (3.45) або (3.51). За винятком випадків, коли вдається знайти аналітичний розв'язок рівнянь (3.45) або (3.51), для їх розв'язання можна використовувати чисельний метод розв'язання задачі Коші.

Але є можливість використати й інший підхід, не використовуючи чисельного інтегрування диференціальних рівнянь. При цьому підході весь інтервал $0 \le \lambda \le 1$ поділяють точками λ_j $(j = 0, 1, \ldots, m)$, $\lambda_0 = 0$, $\lambda_m = 1$ на декілька інтервалів. У точці $\lambda_0 = 0$ розв'язок знаходиться із рівняння $\bar{\psi}(\bar{x}, 0) = 0$; вважається, що він є відомим і дорівнює $\bar{x}^0$. Потім шукається розв'язок рівняння $\bar{\psi}(\bar{x}, \lambda_1) = 0$ з початковим наближенням $\bar{x}(\lambda_1)$ і т. д. Останнім відшукується розв'язок рівняння $\bar{\psi}(\bar{x}, \lambda_m) = 0$ з початковим наближенням $\bar{x}(\lambda_{m-1})$. В результаті отримуємо шуканий розв'язок $\bar{x}(1) = \bar{x}^*$. Для кожного фіксованого λ_j система нелінійних рівнянь розв'язується методом Ньютона. Метод комбінування ітераційного процесу з продовженням по параметру завжди стійкий. При цьому в багатьох випадках рахунок можна вести з достатньо великим кроком по λ. Природно, що цей метод потребує збільшення обчислень, але це компенсується його безвідмовністю.

ГЛАВА 4

ОБЧИСЛЕННЯ ВЛАСНИХ ЗНАЧЕНЬ ТА ВЛАСНИХ ВЕКТОРІВ МАТРИЦЬ

§ 4.1. ПОСТАНОВКА ЗАДАЧІ. ЗАГАЛЬНА ХАРАКТЕРИСТИКА МЕТОДІВ

У цій главі розглядаються питання чисельного розв'язання задач на власні значення. Велика кількість задач з механіки, фізики та техніки вимагає знайдення власних значень і власних векторів матриць, тобто таких значень λ, для яких існує нетривіальний розв'язок однорідної системи лінійних алгебраїчних рівнянь

$$A\bar{x} = \lambda\bar{x} \tag{4.1}$$

і визначення цих нетривіальних розв'язків. Тут A — дійсна квадратна матриця порядку n з елементами a_{jk}, а $\bar{x}$ — вектор із компонентами x_1, x_2, ... , x_n. Кожному власному значенню λ_i відповідає хоча б один нетривіальний розв'язок. Якщо навіть матриця A дійсна, її власні значення (всі або деякі), отже, і власні вектори, можуть бути недійсними.

Власні значення λ матриці A є коренями рівняння

$$D(\lambda) = |A - \lambda E| = 0$$

або у розгорнутому вигляді

$$D(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ a_{n1} & a_{n2} & \dots & a_{11} - \lambda \end{vmatrix} =$$

$$= (-1)^n [\lambda^n - p_1\lambda^{n-1} - \dots - p_{n-1}\lambda - p_n], \tag{4.2}$$

де E — одинична матриця порядку n. Рівняння (4.2) називають *характеристичним рівнянням матриці A*.

Власним вектором $\bar{x}_i(x_{1i}, x_{2i}, \dots, x_{ni})$, що відповідає власному значенню λ_i, називають *ненульовий розв'язок* однорідної системи рівнянь

$$(A - \lambda_i E)\bar{x}_i = 0. \tag{4.3}$$

Оскільки $A\bar{x}_i = \lambda_i\bar{x}_i$, то власний вектор $\bar{x}_i$ після помноження на матрицю A буде колінеарним до вектора $\bar{x}_i$. Отже, задача знаходження власних значень і власних векторів зводиться до знаходження коефіцієнтів характеристичного рівняння (4.2), визначення його коренів і знаходження нетривіального розв'язку системи (4.3). Якщо для даного власного значення

λ_i система (4.3) має кілька лінійно незалежних розв'язків, то цьому власному значенню відповідають кілька власних векторів. Кожному простому (не кратному) власному значенню відповідає один (із точністю до напряму) власний вектор, а сукупність усіх простих власних векторів, що відповідає сукупності простих власних значень, лінійно незалежна. Таким чином, якщо всі власні значення матриці прості, то вона має n лінійно незалежних власних векторів, які утворюють базис простору. Кратному власному значенню кратності p може відповідати від 1 до p лінійно незалежних власних векторів.

Усі чисельні методи знаходження власних значень і власних векторів можна розділити на дві групи. До першої відносяться методи (назвемо їх прямими), у яких спочатку шукається характеристичне рівняння (його коефіцієнти $p_1, p_2, \ldots, p_n$). При його розв'язанні знаходять власні значення матриці і потім власні вектори, що їм відповідають. Іншу групу становлять ітераційні методи, які з розвитком обчислювальної техніки знаходять у порівнянні з прямими методами все ширше використання.

В ітераційних методах власні значення знаходяться як границі деяких числових послідовностей без попереднього визначення коефіцієнтів характеристичного рівняння. При цьому, як правило, водночас обчислюються і власні вектори.

При розв'язанні деяких задач виникає необхідність знаходження всіх власних значень і власних векторів, що їм відповідають. У цьому випадку задача називається *повною проблемою власних значень*. Але існують задачі, у яких немає необхідності у повних відомостях, і можна обмежитись меншим обсягом знань: наприклад, достатньо вказати межі, у яких знаходяться усі власні значення, як це інколи буває при вивченні стійкості або нестійкості процесів, або знайти власне значення, близьке до відомого числа, коли розглядаються явища резонансу, тощо. Усі задачі такого роду називаються *частковими проблемами власних значень* і для кожної із них створюються свої методи розв'язання.

Використання прямих методів дозволяє вирішувати повну проблему власних значень, а ітераційними методами можна також розв'язувати й часткову проблему власних значень.

Задача на власні значення легко розв'язується для деяких простих форм матриць: діагональної, тридіагональної, трикутної або майже трикутної. Наприклад, визначник трикутної (зокрема, діагональної) матриці дорівнює добутку діагональних елементів. У цьому випадку $A - \lambda E$ також трикутна або діагональна матриця. Тому власні значення трикутної (діагональної) матриці дорівнюють діагональним елементам. Легко перевірити, що діагональна матриця має n власних ортонормованих векторів $e_i = \{0, \ldots, 0, 1, 0, \ldots, 0\}^T$, що відповідають власним значенням $\lambda_i = a_{ii}$; навпаки, матриця із такими власними векторами діагональна.

Багато чисельних методів розв'язання задач на власні значення побудовано на зведенні матриці до однієї з наведених вище простих форм або до специфічних форм за допомогою перетворень подібності.

Матриця B

$$B = S^{-1} \cdot A \cdot S \tag{4.4}$$

називається *подібною* до матриці A. Нехай λ, $\bar{y}$ є власне значення і власний вектор матриці B, тоді $\lambda\bar{y} = B\bar{y} = S^{-1} \cdot A \cdot S \cdot \bar{y}$, що після множення зліва на матрицю S дає $\lambda(S\bar{y}) = A(S\bar{y})$. Звідси видно, що λ та $S\bar{y}$ є власне значення і власний вектор матриці A. Отже, перетворення подібності не змінює власних значень матриці і за деяким законом перетворює її власні вектори.

§ 4.2. МЕТОД ДАНИЛЕВСЬКОГО

Простий та витончений метод знаходження характеристичного многочлена та вирішення повної проблеми власних значень запропонував А.М. Данилевський. Тут наводиться видозміна методу, зручного для реалізації на персональних обчислювальних машинах (ПОМ).

Розглянемо ідею методу. Вихідна матриця

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

для якої знаходиться характеристичний многочлен, за допомогою подібних перетворень перетворюється на матрицю

$$P = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & P_1 \\ 1 & 0 & 0 & \dots & 0 & 0 & P_2 \\ 0 & 1 & 0 & \dots & 0 & 0 & P_3 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 & P_n \end{pmatrix}, \tag{4.5}$$

яка має нормальну форму Фробеніуса, тобто матриця має в явному вигляді в останньому стовпці шукані коефіцієнти характеристичного рівняння (4.2). Оскільки подібні матриці мають один і той же характеристичний многочлен, а

$$|P - \lambda E| = (-1)^n \left[\lambda^n - \sum_{i=1}^{n} P_i \lambda^{i-1}\right], \tag{4.6}$$

то і

$$|A - \lambda E| = (-1)^n \left[\lambda^n - \sum_{i=1}^{n} P_i \lambda^{i-1}\right]. \tag{4.6'}$$

Тому для обґрунтування методу досить показати, яким чином з матриці A будується матриця P.

Подібні перетворення матриці A до матриці P здійснюються послідовно. На першому кроці матриця A перетворюється на подібну до неї матрицю $A^{(1)}$, в якій передостанній стовпець має необхідний вигляд. На другому кроці матриця $A^{(1)}$ перетворюється на подібну до неї матрицю $A^{(2)}$, в якій вже два передостанніх стовпці мають необхідний вигляд, і т.д.

На першому кроці матриця A помножується справа на матрицю

$$C_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & a_{1n} \\ 1 & 0 & 0 & \dots & 0 & 0 & a_{2n} \\ 0 & 1 & 0 & \dots & 0 & 0 & a_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 & a_{n-1n} \\ 0 & 0 & 0 & \dots & 0 & 1 & a_{nn} \end{pmatrix} = \begin{pmatrix} 0 & t \\ E_{n-1} & \tau \end{pmatrix} \tag{4.7}$$

і зліва на матрицю, їй обернену

$$C_1^{-1} = \begin{pmatrix} -a_{2n}/a_{1n} & 1 & 0 & \dots & 0 & 0 \\ -a_{3n}/a_{1n} & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -a_{nn}/a_{1n} & 0 & 0 & \dots & 0 & 1 \\ 1/a_{1n} & 0 & 0 & \dots & 0 & 0 \end{pmatrix} = \begin{pmatrix} -\tau/t & E_{n-1} \\ 1/t & 0 \end{pmatrix}. \tag{4.8}$$

У рівностях (4.7) та (4.8) одночасно записані матриці $C_1 \cdot C^{-1}$ в клітинній формі.

Легко перевіряється, що

$$C_1^{-1}A = \begin{pmatrix} a'_{11} & a'_{12} & \dots & a'_{1,n-1} & 0 \\ a'_{21} & a'_{22} & \dots & a'_{2,n-1} & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a'_{n-1,1} & a'_{n-1,2} & \dots & a'_{n-1,n-1} & 0 \\ a'_{n1} & a'_{n2} & \dots & a'_{n,n-1} & 1 \end{pmatrix}, \tag{4.9}$$

де

$$a_{ik}' = a_{i+1,k} - a_{1k}\frac{a_{i+1,n}}{a_{1n}} \quad (i = k = 1, \dots, n-1);$$

$$a_{nk}' = a_{1k}/a_{1n} \quad (k = 1, \dots, n). \tag{4.10}$$

Перший крок дає

$$A^{(1)} = C_1^{-1}AC_1 = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1,n-2}^{(1)} & 0 & a_{1n}^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} & \dots & a_{2,n-2}^{(1)} & 0 & a_{2n}^{(1)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n-1,1}^{(1)} & a_{n-1,2}^{(1)} & \dots & a_{n-1,n-2}^{(1)} & 0 & a_{n-1,n}^{(1)} \\ a_{n1}^{(1)} & a_{n2}^{(1)} & \dots & a_{n,n-2}^{(1)} & 0 & a_{nn}^{(1)} \end{pmatrix}, \tag{4.11}$$

де

$$a_{ik}^{(1)} = a_{i,k+1}' \quad (i = 1, \dots, n; \quad k = 1, \dots, n-1);$$

$$a_{in}^{(1)} = \sum_{k=1}^{n} a_{ik}' a_{kn} \quad (i = 1, \dots, n). \tag{4.12}$$

Таким чином, елементи матриці $A^{(1)}$ можуть бути отримані з елементів матриці A за формулами

$$a_{ik}^{(1)} = a_{i+1,k+1} - a_{1,k+1} \frac{a_{i+1,n}}{a_{1n}} \quad (i = 1, \dots, n-1; \ k = 1, \dots, n-1);$$

$$a_{nk}^{(1)} = a_{1,k+1} / a_{1n} \quad (k = 1, \dots, n-1);$$

$$a_{in}^{(1)} = \sum_{k=1}^{n} \left(a_{i+1,k} - a_{1,k} \frac{a_{i+1,n}}{a_{1n}} \right) a_{k,n} \quad (i = 1, \dots, n-1);$$

$$a_{nn}^{(1)} = 1 / a_{1n} \sum_{k=1}^{n} a_{1,k} a_{k,n}. \tag{4.13}$$

На другому кроці матриця $A^{(1)}$ помножується справа на матрицю

$$C_2 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & a_{1n}^{(1)} \\ 1 & 0 & 0 & \dots & 0 & 0 & a_{2n}^{(1)} \\ 0 & 1 & 0 & \dots & 0 & 0 & a_{3n}^{(1)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 & a_{n-1,n}^{(1)} \\ 0 & 0 & 0 & \dots & 0 & 1 & a_{nn}^{(1)} \end{pmatrix} = \begin{pmatrix} 0 & f^{(1)} \\ E_{n-1} & \tau^{(1)} \end{pmatrix} \tag{4.14}$$

і зліва на зворотну до неї матрицю

$$C_2^{-1} = \begin{pmatrix} -a_{2n}^{(1)}/a_{1n}^{(1)} & 1 & 0 & \dots & 0 & 0 \\ -a_{3n}^{(1)}/a_{1n}^{(1)} & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -a_{nn}^{(1)}/a_{1n}^{(1)} & 0 & 0 & \dots & 0 & 1 \\ 1/a_{1n}^{(1)} & 0 & 0 & \dots & 0 & 0 \end{pmatrix} = \begin{pmatrix} -\tau^{(1)}/f^{(1)} & E_{n-1} \\ 1/f^{(1)} & 0 \end{pmatrix}. \tag{4.15}$$

Очевидно, що елементи матриці

$$A^{(2)} = C_2^{-1} A^{(1)} C_2 = \left\{ a_{ik}^2 \right\} \tag{4.16}$$

обчислюються за формулами

$$a_{ik}^{(2)} = a_{i+1,k+1}^{(1)} - a_{1,k+1}^{(1)} \frac{a_{i+1,n}^{(1)}}{a_{1n}^{(1)}} \quad (i = 1, \dots, n-1; \ k = 1, \dots, n-1);$$

$$a_{nk}^{(2)} = a_{1,k+1}^{(1)} / a_{1n}^{(1)} \quad (k = 1, \ldots, n-1);$$

$$a_{in}^{(2)} = \sum_{k=1}^{n} \left(a_{i+1,k}^{(1)} - a_{1,k}^{(1)} \frac{a_{i+1,n}^{(1)}}{a_{1n}^{(1)}} \right) a_{k,n}^{(1)}, \quad (i = 1, \ldots, n-1);$$

$$a_{nn}^{(2)} = (1/a_{1n}^{(1)}) \sum_{k=1}^{n} a_{1k}^{(1)} a_{k,n}^{(1)}. \tag{4.17}$$

З цих формул (враховуючи, що $a_{i,n-1}^{(1)} = 0$ при $i \neq n$, $a_{n,n-1}^{(1)} = 1$) випливає

$$a_{i,n-2}^{(2)} = \begin{cases} 0, & i \neq n-1; \\ 1, & i = n-1; \end{cases} \qquad a_{i,n-1}^{(2)} = \begin{cases} 0, & i \neq n; \\ 1, & i = n. \end{cases} \tag{4.18}$$

Це означає, що два передостанніх стовпці матриці $A^{(2)}$ мають потрібний вигляд. Продовжуючи цей процес, після $n-1$ кроків прийдемо до матриці

$$A^{(n-1)} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & a_{1n}^{(n-1)} \\ 1 & 0 & 0 & \ldots & 0 & 0 & a_{2n}^{(n-1)} \\ 0 & 1 & 0 & \ldots & 0 & 0 & a_{3n}^{(n-1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 0 & 1 & a_{nn}^{(n-1)} \end{pmatrix}, \tag{4.19}$$

що має нормальну форму Фробеніуса та подібна до вихідної матриці A. При цьому на кожному кроці елементи матриці $A^{(j)}$ знаходяться по елементах матриці $A^{(j-1)}$ так само, як ми знаходили за формулами (4.17) елементи $A^{(2)}$ по елементах $A^{(1)}$. При цьому припускається, що всі елементи $a_{1n}^{(j)}$ відмінні від нуля. Якщо на j-му кроці виявиться, що $a_{1n}^{(j)} = 0$, то продовжувати процес у такому вигляді не буде можливим. При цьому можуть виникнути два випадки.

1. Серед елементів $a_{2n}^{(j)}, a_{3n}^{(j)}, \ldots, a_{n-j,n}^{(j)}$ є хоча б один, відмінний від нуля, наприклад $a_{i_0 n}^{(j)}$. Для продовження процесу поміняємо в $A^{(j)}$ місцями перший та i_0-й рядки та одночасно 1-й та i_0-й стовпці. Таке перетворення матриці $A^{(j)}$ буде подібним. Після отримання матриці $\overline{A}^{(j)}$ процес можна продовжувати, оскільки стовпці матриці $A^{(j)}$, приведені до необхідного вигляду, не будуть зіпсовані.

2. Всі елементи $a_{2n}^{(j)}, a_{3n}^{(j)}, \ldots, a_{n-j,n}^{(j)}$ дорівнюють нулю. Тоді матриця $A^{(j)}$ має вигляд

$$A^{(j)} = \begin{pmatrix} B & 0 \\ B_1 & F \end{pmatrix},$$

де F — квадратична матриця порядку $j+1$, яка має нормальний вигляд Фробеніуса; де B — квадратична матриця порядку $n-j-1$, але

$$|A^{(j)} - \lambda E| = |B - \lambda E_{n-j-1}| \, |F - \lambda E_{j+1}|, \qquad (4.20)$$

тобто характеристичний многочлен матриці F є дільником характеристичного многочлена матриці A. Як видно з (4.20), для знаходження характеристичного многочлена матриці A необхідно ще знайти характеристичний многочлен матриці B, для чого використати можна той же метод.

Підраховано, що кількість операцій множення та ділення, необхідних для одержання характеристичного многочлена матриці порядку n, становить $n(n-1)(2n+3)/2$.

Можна досягти більшої точності. Вибираючи як елемент $t^{(j)}$, на який проводиться ділення, найбільший за абсолютною величиною елемент останнього стовпця матриці $A^{(j)}$, має вигляд

$$C_j = \begin{pmatrix} E_1 & 0 & \tau_1^{(j)} \\ 0 & 0 & t^{(j)} \\ 0 & E_2 & \tau_2^{(j)} \end{pmatrix}; \quad C_j^{-1} = \begin{pmatrix} E_1 & -\tau_1^{(j)}/t^{(j)} & 0 \\ 0 & -\tau_2^{(j)}/t^{(j)} & E_2 \\ 0 & 1/t^{(j)} & 0 \end{pmatrix},$$

де останній стовпець в C_j збігається з останнім стовпцем матриці $A^{(j-1)}$; $t^{(j)}$ — найбільший за абсолютною величиною елемент цього стовпця; E_1; E_2 — одиничні матриці. Порядок E_2 повинен бути більше $j-1$. При такому способі випадок 1 не може виникнути.

Розглянемо питання щодо пошуку власних векторів матриці A. Якщо матриця має нормальну форму Фробеніуса, її власні вектори знаходяться просто. Нехай $\bar{y} = (y_1, \ldots, y_n)$ — власний вектор матриці P відповідний власному значенню λ. Тоді система $P\bar{y} = \lambda\bar{y}$ може бути розписана таким чином:

$$\begin{cases} P_1 y_n = \lambda y_1; \\ y_1 + P_2 y_n = \lambda y_2; \\ y_2 + P_3 y_n = \lambda y_3; \\ \ldots\ldots\ldots \\ y_{n-2} + P_{n-1} y_n = \lambda y_{n-1}; \\ y_{n-1} + P_n y_n = \lambda y_n. \end{cases} \qquad (4.21)$$

Компонента $y_n \neq 0$, оскільки в іншому випадку всі $y_i = 0$ і $\bar{y} = 0$, що неможливо (власний вектор — не нульовий вектор).

Поклавши $y_n = 1$ із (4.21), дістанемо

$$y_n = 1; \quad y_{n-1} = \lambda - P_n;$$

$$y_{n-2} = \lambda^2 - P_n\lambda - P_{n-1}, \ldots, y_1 = \lambda^{n-1} - P_n\lambda^{n-2} - \ldots - P_3\lambda - P_2,$$

тобто знайдені всі компоненти власного вектора $\bar{y}$. Позначимо через S матрицю

$$S = C_1 C_2 \ldots C_{n-1}. \qquad (4.22)$$

Тоді, якщо $\bar{y}$ — власний вектор матриці P, що відповідає власному значенню λ, то з

$$S^{-1}AS = P = A^{(n-1)}$$

випливає, що

$$S^{-1}AS\bar{y} = P\bar{y} = \lambda\bar{y};$$

$$AS\bar{y} = \lambda S\bar{y}.$$

А це означає, що вектор $\bar{x} = S\bar{y}$ є власним вектором матриці A, відповідним λ.

Таким чином, визначивши всі власні вектори матриці P та помноживши їх на матрицю S, дістанемо всі власні вектори матриці A.

Приклад. Знайти власні значення та власні вектори матриці

$$A = \begin{pmatrix} 5 & 2 & -1 & -1 \\ 3 & 3 & 0 & 0 \\ 1 & -2 & 4 & 1 \\ 3 & 0 & 0 & 3 \end{pmatrix}.$$

Розв'язання. Згідно з методом Данилевського маємо

$$A^{(1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 3 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 5 & 2 & -1 & -1 \\ 3 & 3 & 0 & 0 \\ 1 & -2 & 4 & 1 \\ 3 & 0 & 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix} =$$

$$= \begin{pmatrix} 3 & 0 & 0 & -3 \\ 0 & 3 & 0 & -3 \\ 6 & -3 & 0 & -21 \\ -2 & 1 & 1 & 9 \end{pmatrix};$$

$$A^{(2)} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -7 & 0 & 1 & 0 \\ 3 & 0 & 0 & 1 \\ -1/3 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 0 & 0 & -3 \\ 0 & 3 & 0 & -3 \\ 6 & -3 & 0 & -21 \\ -2 & 1 & 1 & 9 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -21 \\ 0 & 0 & 1 & 9 \end{pmatrix} =$$

$$= \begin{pmatrix} 3 & 0 & 0 & 0 \\ -3 & 0 & 0 & 54 \\ 1 & 1 & 0 & -45 \\ 0 & 0 & 1 & 12 \end{pmatrix}.$$

Далі процес продовжувати не можна й не потрібно, оскільки маємо виключний випадок 2. Матимемо

$$|A - \lambda E| = |A^{(2)} - \lambda E| = (3 - \lambda)(\lambda^3 - 12\lambda^2 + 45\lambda - 54) =$$

$$= -(\lambda - 3)^3(\lambda - 6);$$

$$\lambda_1 = \lambda_2 = \lambda_3 = 3; \quad \lambda_4 = 6.$$

Власні вектори матриці $A^{(2)}$:

$$\bar{y}_1 = (0, 18, -9, 1); \quad \bar{y}_2 = (1, -1, 0, 0), \quad \bar{y}_3 = (0, 9, -6, 1).$$

Матриця

$$S = C_1C_2 = \begin{pmatrix} 0 & 0 & -1 & -9 \\ 0 & 0 & 0 & -3 \\ 1 & 0 & 1 & 6 \\ 0 & 1 & 3 & 6 \end{pmatrix}.$$

Власні вектори матриці A

$$\bar{x}_1 = S\bar{y}_1 = (0, 1, -5, 7);$$

$$\bar{x}_2 = S\bar{y}_2 = (0, 0, 1, -1);$$

$$\bar{x}_3 = S\bar{y}_3 = (1, 1, 0, 1).$$

§ 4.3. МЕТОД ЛЕВЕР'Є

Відомо багато інших способів одержання характеристичного многочлена.

Розглянемо метод Левер'є, що дозволяє вирішити проблему власних значень, в основу якого покладено обчислювання слідів степенів матриці A. Вказаний метод потребує більшої кількості операцій, ніж розглянутий вище, але зовсім не чутливий до частинних особливостей матриці, зокрема «провалів» проміжних визначників.

Нехай характеристичний поліном матриці A записано у вигляді (4.2), де $\lambda_1, \lambda_2, \ldots, \lambda_n$ — його корені, серед яких деякі можуть бути рівні. Позначимо

$$\sum_{i=1}^{n} \lambda_i^k = S_k. \tag{4.23}$$

Суми S_k, $k = \overline{1, n}$ — степенів коренів многочлена зв'язані з коефіцієнтами рівняння (4.2) формулами Ньютона

$$kp_k = S_k - \sum_{i=1}^{k-1} p_i S_{k-i}, \quad k = 1, \ldots, n. \tag{4.24}$$

Якщо обчислити сліди $S_1, S_2, \ldots, S_n$ матриць $A, A^2, \ldots, A^n$, то з (4.24) можна послідовно обчислювати коефіцієнти p_k.

Покажемо, як визначаються числа S_k:

$$S_1 = S_p A = \sum_{i=1}^{n} a_{ii};$$

$$S_2 = S_p A^2 = \sum_{i=1}^{n} a_{ii}^{(2)},$$

. . . . . .

Оскільки матриця A^k має своїми власними значеннями числа $\lambda_1^k, \lambda_2^k, \ldots$ $\ldots, \lambda_n^k$, то $S_p A^k = S_k = \sum_{i=1}^{n} \lambda_i^k$.

Таким чином, процес обчислення зводиться до послідовного обчислення степенів матриці A, обчислення їх слідів (суми діагональних елементів) і, нарешті, до розв'язання рекурентної системи (4.24). Обчислення n степенів матриці A (в останньої матриці A^n треба знайти тільки діагональні елементи) потребує великої кількості одноманітних операцій, які легко реалізуються за допомогою ПВМ. Кількість необхідних за методом Лавер'є множень дорівнює $\frac{1}{2}(n-1)(2n^3 - 2n^2 + n + 2)$.

Зазначимо, що при обчисленні степенів матриці корисно здійснювати контроль за допомогою стовпця, що складається із сум елементів кожного рядка матриці A.

Результат множення матриці A на цей стовпець повинен збігатися з аналогічним стовпцем матриці A^2. Дійсно, нехай Σ_1 — стовпець сум матриці A; Σ_2 — стовпець сум матриці A^2. Нехай $U = (1, 2, \ldots, 1)'$. Тоді

$$\Sigma_1 = AU; \quad \Sigma_2 = A^2 U \to \Sigma_2 = A\,\Sigma_1.$$

Очевидно, сказане вірне й для інших степенів.

Визначивши з допомогою вказаного методу коефіцієнти характеристичного полінома вигляду (4.2), знаходимо його корені, які є шуканими власними значеннями. Для відшукання власних векторів $\bar{x}_i$, відповідних власним значенням λ_i, необхідно знайти нетривіальний розв'язок системи рівнянь (4.1), для чого можна скористатися способом, описаним у попередньому параграфі, заснованим на методі виключення Гаусса.

Приклад. Побудувати характеристичний поліном матриці

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 4 & 6 & -1 \\ 4 & 4 & 1 \end{pmatrix}.$$

Розв'язання. У відповідності з методом Левер'є будуємо степені A^k $(k = 2, 3)$

$$A^2 = \begin{pmatrix} 1 & -1 & 1 \\ 4 & 6 & -1 \\ 4 & 4 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 1 \\ 4 & 6 & -1 \\ 4 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -3 & 3 \\ 24 & 28 & -3 \\ 24 & 24 & 1 \end{pmatrix};$$

$$A^3 = \begin{pmatrix} 1 & -1 & 1 \\ 4 & 6 & -1 \\ 4 & 4 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -3 & 3 \\ 24 & 28 & -3 \\ 24 & 24 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -7 & 7 \\ 124 & 132 & -7 \\ 124 & 124 & 1 \end{pmatrix}.$$

Звідси

$$S_1 = S_p A = 1 + 6 + 1 = 8;$$

$$S_2 = S_p A^2 = 1 + 28 + 1 = 30;$$

$$S_3 = S_p A^3 = 1 + 132 + 1 = 134.$$

Відповідно до формул (4.24) маємо

$$p_1 = S_1 = 8;$$

$$p_2 = 1/2(S_2 - p_1 S_1) = 1/2(30 - 8 * 8) = -17;$$

$$p_3 = 1/3(S_3 - p_1 S_2 - p_2 S_1) = 1/3(134 - 8 * 30 + 17 * 8) = 10.$$

Таким чином, враховуючи вигляд (4.2), дістанемо

$$\lambda^3 - 8\lambda^2 + 17\lambda - 10 = 0.$$

Власні значення матриці A: $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 5$.

§ 4.4. ЗВЕДЕННЯ СИМЕТРИЧНОЇ МАТРИЦІ ДО ТРИДІАГОНАЛЬНОГО ВИГЛЯДУ

Метод, що розглядається тут, відомий як прямий метод Якобі. Базується він на тій самій ідеї, що й метод Данилевського: подібні матриці мають однаковий характеристичний многочлен. Існуючі методи, побудовані на перетворенні подібності (4.4), відрізняються вибором матриці S в залежності від матриці A. При цьому зручно як матрицю S використовувати ортогональні матриці, оскільки для них $S^{-1} = S'$ і немає необхідності знаходити обернені матриці.

Розглянемо симетричні матриці. Проблема обчислення власних значень дійсних симетричних матриць суттєво спрощується, відносно розв'язання тієї ж задачі для матриць загального вигляду, тому що власні значення перших завжди добре зумовлені.

Наведемо деякі важливі особливості симетричної матриці.

1. Усі власні значення дійсні.

2. Власні вектори, що відповідають власним значенням, взаємно ортогональні.

3. Кожному власному значенню відповідає стільки ж лінійно незалежних власних векторів, як кратність власного значення.

4. За допомогою перетворення подібності кожну симетричну матрицю можна привести до діагонального вигляду.

У прямому методі Якобі симетрична матриця A за допомогою ланцюжка перетворень зводиться зі збереженням подібності до тридіагональної матриці. *Обертанням* будемо називати перетворення координат з ортогональною елементарною матрицею обертання

$$T_{ij} = \begin{bmatrix} & & i & & j & \\ & 1 & & & & \\ & & c & & -s & \cdots\ i \\ & & & 1 & & \\ & & s & & c & \cdots\ j \\ & & & & & 1 \end{bmatrix} \tag{4.25}$$

при $c^2 + s^2 = 1$.

Геометрично обертання може бути інтерпретовано як поворот базисних векторів e_i, e_j на деякий кут, здійснюваний в площині, що натягнена на вектори e_i, e_j.

Покажемо, що будь-яку симетричну матрицю можна привести до тридіагональної форми за допомогою ланцюжка перетворень подібності з матрицями вигляду T_{ij}.

Проведемо необхідні перетворення: $B = AT_{ij}$, $C = T_{ij}'B = T_{ij}'AT_{ij}$. Тут усі стовпці матриці B збігаються зі стовпцями матриці A, за винятком i-го та j-го стовпців, які утворюються з відповідних стовпців матриці A за формулами

$$B_i = cA_i + sA_j; \quad B_j = -sA_i + cA_j. \tag{4.26}$$

У свою чергу рядки матриці C збігаються з рядками матриці B, за винятком i-го та j-го, які одержуються з відповідних рядків матриці B за такими самими формулами:

$$C^i = cB^i + sB^j; \quad C^j = -sB^i + cB^j. \tag{4.27}$$

При цьому для побудови рядків c^i та c^j необхідно обчислити тільки чотири елементи c_{ji}, c_{ij}, c_{ii}, c_{jj}, причому c_{ji} тільки для контролю, оскільки $c_{ij} = c_{ji}$, але вони одержуються через різні обчислення. Інші елементи рядків c^i та c^j не тільки теоретично дорівнюють відповідним елементам стовпців B_i та B_j, але при їх обчисленні виконуються однакові дії.

Покажемо, що c та s можна вибирати так, щоб $c_{i-1\,j} = 0$. Нехай $1 < i < j$. Дійсно, $c_{i-1\,j} = b_{i-1\,j} = -sa_{i-1\,i} + ca_{i-1\,j}$, тому достатньо взяти $\frac{s}{c} = \frac{a_{i-1\,j}}{a_{i-1\,i}}$ та відповідно

$$s = \frac{a_{i-1\,j}}{\pm\sqrt{a_{i-1\,i}^2 + a_{i-1\,j}^2}}, \qquad c = \frac{a_{i-1\,i}}{\pm\sqrt{a_{i-1\,i}^2 + a_{i-1\,j}^2}}.$$

Вибір знака знаменника не має значення.

Розглянемо зведення симетричної матриці до тридіагонального вигляду. За рахунок перетворень за допомогою $T_{23}, \ldots, T_{2n}$ анулюються по черзі елементи першого рядка, починаючи з третього; за рахунок $T_{34}, \ldots, T_{3n}$ — елементи другого рядка, починаючи з четвертого. При цьому елементи першого рядка більше змінюватись не будуть. Дійсно, перші два елементи першого рядка не будуть змінюватись при перетвореннях за допомогою $T_{34}, \ldots, T_{3n}$. Елементи, які дорівнюють нулю, підлягатимуть лінійним однорідним перетворенням і тому залишаться нульовими. Далі за рахунок перетворень за допомогою $T_{45}, \ldots, T_{4n}$ анулюються елементи третього рядка, починаючи з п'ятого, тощо.

Із цього ясно, що кожне наступне перетворення не буде змінювати раніше анульовані елементи. Таким чином, максимум через $\frac{(n-1)(n-2)}{2}$ перетворень ми перейдемо від симетричної матриці A до тридіагональної матриці S

$$S = \begin{pmatrix} s_{11} & s_{12} & 0 & 0 & \ldots & 0 \\ s_{21} & s_{22} & s_{23} & 0 & \ldots & 0 \\ 0 & s_{32} & s_{33} & s_{34} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & 0 & 0 & \ldots & s_{nn} \end{pmatrix}. \tag{4.27}$$

Далі будується характеристичний поліном

$$\varphi_{k+1}(\lambda) = (s_{k+1\,k+1} - \lambda)\varphi_k(\lambda) - s_{k\,k+1}^2\varphi_{k-1}, \quad k = 1, \ldots, n-1$$

та обчислюються його корені.

Обчислення власних векторів для матриці S може бути здійснено через розв'язання методом прогонки, наведеним у гл. 2, відповідної трикутної системи

$$\begin{aligned} &(s_{11} - \lambda_i)v_1 + s_{12}v_2 = 0; \\ &s_{12}v_1 + (s_{22} - \lambda_i)v_2 + s_{23}v_3 = 0; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &s_{n-1\,n}v_{n-1} + (s_{nn} - \lambda_i)v_n = 0 \end{aligned} \tag{4.28}$$

для компонент $\bar{v}_1, \ldots, \bar{v}_n$ власного вектора $\bar{v}_l$, що належить λ_l. Тут зручно задатись першою компонентою, а потім послідовно обчислювати другу, третю і т. д.

Для переходу від власних векторів матриці S до власних векторів матриці A треба використати співвідношення

$$S = [T_{23} \ldots T_{n-1\,n}]' \, A T_{23} \ldots T_{n-1\,n},$$

із якого виходить, що кожен власний вектор $\bar{x}$ матриці A висловлюється через власний вектор $\bar{v}$ матриці S за формулою

$$\bar{x} = T_{23} T_{24} \ldots T_{n-1\,n} \bar{v}, \tag{4.29}$$

тобто $\bar{x}$ знаходять із $\bar{v}$ за допомогою кількох множень на матриці поворотів T_{ij}. При кожному окремому множенні будуть змінюватись лише дві компоненти попереднього вектора — i-та та j-та за формулами

$$v_i' = cv_i - sv_j; \qquad v_j' = sv_i - cv_j, \tag{4.30}$$

де v_i та v_j — компоненти попереднього вектора; v_i' та v_j' — наступного.

Хоча кількість множень у цьому процесі надто значна, помилки округлення накопичуються повільно, через те, що вони помножуються на коефіцієнти c і s, які за модулем менші за одиницю.

Якщо виявиться, що один або кілька елементів $s_{k-1\,k}$ дорівнюють нулю, то матриця S розділиться на два або кілька ящиків Якобі і задача обчислення власних значень тільки полегшиться. Це явище мабуть буде мати місце, якщо вихідна матриця має кратні власні значення.

Важливою перевагою методу Якобі є ортогональні перетворення, що використовуються у методі і не збільшують помилок. При цьому хоча для матриці A будується характеристичний поліном, відшукання його коренів полегшується тим, що відоме їхнє приблизне розташування.

Вказаний процес можна застосовувати і до несиметричної матриці, тільки при цьому замість тридіагональної матриці дістанемо майже трикутну

$$\begin{pmatrix} s_{11} & s_{12} & 0 & \ldots & 0 & 0 \\ s_{21} & s_{22} & s_{23} & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ s_{n-1\,1} & s_{n-1\,2} & s_{n-1\,3} & \ldots & s_{n-1\,n-1} & s_{n-1\,n} \\ s_{n1} & s_{n2} & s_{n3} & \ldots & s_{n\,n-1} & s_{nn} \end{pmatrix}.$$

Приклад. Розв'язати повну проблему власних значень для симетричної матриці

$$A = \begin{pmatrix} 1{,}00 & 2{,}00 & 3{,}00 \\ 2{,}00 & 3{,}00 & -5{,}00 \\ 3{,}00 & -5{,}00 & 2{,}00 \end{pmatrix}.$$

Розв'язання. Застосовуючи прямий метод Якобі, отримаємо тридіагональну матрицю з характеристичним поліномом $\lambda^3 - 6\lambda^2 - 27\lambda + 114 = 0$.

Власні числа матриці A: $\lambda_1 = 3{,}16919$; $\lambda_2 = -4{,}74696$; $\lambda_3 = 7{,}57777$.

Власні вектори, відповідні власним значенням: $\bar{x}_1 = (0{,}8430;\ 0{,}4304;\ 0{,}3226)$; $\bar{x}_2 = (0{,}5287;\ -0{,}5524;\ -0{,}6445)$; $\bar{x}_3 = (-0{,}9912;\ 0{,}7139;\ -0{,}6932)$.

§ 4.5. ІНТЕРАЦІЙНИЙ МЕТОД ЯКОБІ ДЛЯ СИМЕТРИЧНИХ МАТРИЦЬ

Розглянемо метод, заснований на підборі такої нескінченної послідовності елементарних обертань, яка в границі перетворює дану симетричну матрицю A в діагональну, подібну вихідній:

$$\Omega = \dots T_{ij}^{(k)\prime} \dots T_{ij}^{(2)\prime} T_{ij}^{(1)\prime} A T_{ij}^{(1)} T_{ij}^{(2)} \dots T_{ij}^{(k)}.$$

При цьому на k-му кроці має місце перетворення

$$A_k = T_{ij}^{(k)\prime} A_{k-1} T_{ij}^{(k)}, \quad k = 1, 2, \dots \ (A_0 = A),$$

де T_{ij}^{k} — елементарна матриця обертання. На практиці процес закінчується тоді, коли позадіагональні елементи стають незначними в порівнянні із заданою точністю.

Для дійсних симетричних матриць застосовуються плоскі обертання. При цьому, як і в прямому методі Якобі, тут використовуються такі ж перетворення T_{ij} (4.25), але послідовність обертів (i, j) та їхні кути (c, s) добираються іншим способом.

Розглянемо дію обертання на сферичну норму матриці (точніше на її квадрат)

$$S_A = \|A\|_E^2 = \sum_{i,j=1}^{n} |a_{ij}|^2. \tag{4.31}$$

Як і в прямому методі Якобі, зупинимось на матрицях $B = AT_{ij}$ та $C = T_{ij}'AT_{ij}$. Характеристичні рівняння матриць A, B, C збігаються; отже, подібні матриці мають однакові власні числа.

Із формул (4.26) виходить, що

$$b_{ki}^2 + b_{kj}^2 = a_{ki}^2 + a_{kj}^2. \tag{4.32}$$

Покажемо, що $b_{ki} = ca_{ki} + sa_{kj}$; $b_{kj} = -sa_{ki} + ca_{kj}$; $b_{ki}^2 + b_{kj}^2 = a_{ki}^2(c^2 + s^2) + a_{kj}^2(c^2 + s^2) = a_{ki}^2 + a_{kj}^2$, тобто при переході до матриці B елементи i-го та j-го стовпців змінюються так, що попарні суми квадратів зберігаються. Оскільки елементи інших стовпців узагалі не змінюються, то

$S_B = S_A$ (сферична норма матриці не змінюється при множенні матриці справа на матрицю обертання).

Аналогічно можна показати, що $S_C = S_B$

$$c_{ik}^2 + c_{jk}^2 = b_{ik}^2 + b_{jk}^2. \tag{4.33}$$

Отже, $S_A = S_B = S$. Ітераційний метод Якобі якраз і заснований на збереженні сферичної норми при обертанні.

Розіб'ємо суму, що входить в сферичну норму (4.31), на діагональну та недіагональну частини

$$S_1 = \sum_{i=1}^{n} |a_{ii}|^2, \quad S_2 = \sum_{i,j=1}^{n} |a_{ij}|^2. \tag{4.34}$$

При елементарному перетворенні обертання $T_{ij}'AT_{ij}$ недіагональні елементи a_{ki}, a_{kj}, a_{ik}, a_{jk} при $k \neq j$ змінюються так, що попарно суми їхніх квадратів зберігаються. Крім цих елементів, є ще один позадіагональний елемент a_{ij}, що змінюється. При елементарному обертанні S_2 змінюється настільки, наскільки зміниться $2a_{ij}^2$. Для максимального зменшення S_2 за один оберт c і s добираються таким чином, щоб анулювати елемент c_{ij}. На підставі того, що

$$c_{ij} = cb_{ij} + sb_{jj} = cs(a_{jj} - a_{ii}) + (c^2 - s^2)a_{ij},$$

для визначення s та c дістаємо

$$(c^2 - s^2)a_{ij} = cs(a_{ii} - a_{jj});$$

$$c^2 + s^2 = 1. \tag{4.35}$$

Піднесемо перше рівняння (4.35) в квадрат, виключимо з нього s за допомогою другого рівняння, при цьому отримаємо

$$c^4 - c^2 + a_{ij}^2\,[4a_{ij}^2 + (a_{ii} - a_{jj})^2]^{-1} = 0;$$

$$c = \sqrt{\frac{1}{2}\left(1 + \frac{1}{\sqrt{1+\mu^2}}\right)};$$

$$s = (\operatorname{sign}\mu)\sqrt{\frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\mu^2}}\right)};$$

$$\mu = \frac{2a_{ij}^{(k-1)}}{a_{ii}^{(k-1)} - a_{jj}^{(k-1)}},$$

якщо $a_{ii}^{(k-1)} \neq a_{jj}^{(k-1)}$.

При такому визначенні обертань сума S_2 з кожним обертом зменшується, а сума S_1 — збільшується і можна підібрати таку послідовність обертань, щоб $S_2 = 0$. Після достатньої кількості обертів всі позадіаго-

нальні елементи стануть зневажливо малі, а матриця A перетвориться на діагональну. Діагональні елементи одержаної матриці і будуть шуканими власними значеннями.

Оскільки при черговому оберті раніше знищений елемент знову може стати ненульовим, то анулювати всі позадіагональні елементи за кінцеву кількість обертів не можна.

Найвигідніше підібрати обертання таким чином, щоб при черговому оберті перетворювати в нуль максимальний по модулю позадіагональний елемент. Але знаходити такий елемент на ЕОМ невигідно. Найбільш ефективним виявилось обнулювання так званого оптимального елемента — найбільшого по модулю елемента в найбільшій із сум

$$r_i = \sum_{\substack{j=1 \\ i \neq j}} a_{ij}^2, \quad i = 1, 2, \ldots, n, \tag{4.36}$$

де r_i — сума квадратів позадіагональних елементів i-ї ітерації.

За міру близькості матриці A_k до діагональної приймається величина

$$t(A_k) = \sum_{i \neq j}^{n} |a_{ij}^{(k)}|^2. \tag{4.37}$$

Остаточним вважається N крок ітераційного процесу, при якому $t(A_N) < \varepsilon$.

Збіжність ітераційного методу Якобі випливає з таких міркувань. Оптимальний елемент становить не менш $1/n - 1$ частини суми (4.36) свого рядка, яка повинна бути не менше $1/n$ частини S_2. Отже, за один оберт позадіагональна частина сферичної норми зменшиться не менше $2/n(n - 1)$ частини своєї величини (перетворюються в нуль два симетричних елемента). Отже, після N обертів величина S_2 складає $(1 - 2/n(n-1))^N$ частину своєї первісної величини, тобто $S_2 \to 0$ при $N \to \infty$. Метод Якобі з вибором оптимального елемента завжди збігається, причому чим далі від розв'язання збіжність не гірша від лінійної, а поблизу — квадратична.

При реалізації матриць A_k на ЕОМ доцільно не вводити проміжну матрицю $C = A_{k-1}T_k$, а користуватись підсумковими формулами для $a_{ij}^{(k)}$:

$$a_{lm}^{(k)} = a_{lm}^{(k-1)}, \quad l, m = 1, \ldots, n; \quad l, m \neq i, j; \quad a_{ij}^{(k)} = 0;$$

$$a_{ii}^{(k)} = c^2 a_{ii}^{(k-1)} + 2sca_{ij}^{(k-1)} + s^2 a_{jj}^{(k-1)};$$

$$a_{jj}^{(k)} = s^2 a_{ii}^{(k-1)} - 2sca_{ij}^{(k-1)} + c^2 a_{jj}^{(k-1)};$$

$$a_{im}^{(k)} = ca_{im}^{(k-1)} + sa_{jm}^{(k-1)}, \quad m = i+1, \ldots, n; \quad m \neq j;$$

$$a_{jm}^{(k)} = -sa_{im}^{(k-1)} + ca_{jm}^{(k-1)}, \quad m = j+1, \ldots, n;$$

$$a_{li}^{(k)} = ca_{li}^{(k-1)} + sa_{lj}^{(k-1)}, \quad l = 1, \ldots, i-1;$$

$$a_{lj}^{(k)} = -sa_{li}^{(k-1)} + ca_{lj}^{(k-1)}, \quad l = 1, \ldots, j-1, \quad l \neq i.$$

Власні вектори матриці A визначаються на підставі того, що оскільки вектори

$$\bar{e}_1 = \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}, \quad \bar{e}_2 = \begin{pmatrix} 0 \\ 1 \\ . \\ . \\ 0 \end{pmatrix}, \ldots, \quad \bar{e}_n = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 1 \end{pmatrix}$$

є власними векторами діагональної матриці, власними векторами матриці A будуть стовпці матриці обертання

$$T = \prod_{i,j=1}^{n} T_{ij}. \tag{4.39}$$

Приклад. Розв'язати з залученням ітераційного методу Якобі повну проблему власних значень для заданої симетричної матриці з точністю $\varepsilon = 10^{-4}$

$$A = \begin{pmatrix} 1{,}00 & 0{,}42 & 0{,}54 & 0{,}66 \\ 0{,}42 & 1{,}00 & 0{,}32 & 0{,}44 \\ 0{,}54 & 0{,}32 & 1{,}00 & 0{,}22 \\ 0{,}66 & 0{,}44 & 0{,}22 & 1{,}00 \end{pmatrix}. \tag{4.40}$$

Розв'язання. Наведемо для кожної ітерації вигляд матриці обертання $T^{(k)}$ та матриць A_k. При цьому на кожному кроці перевіряється ступінь мализни величини ε, контролюється виконання збереження сферичної норми матриць A_k.

Ітерація 1

$$T_1^{(1)} = \begin{pmatrix} 7{,}07107*10^{-1} & 0{,}00000 & 0{,}00000 & -7{,}07107*10^{-1} \\ 0{,}00000 & 1{,}00000 & 0{,}00000 & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 1{,}00000 & 0{,}00000 \\ 7{,}07107*10^{-1} & 0{,}00000 & 0{,}00000 & 7{,}07107*10^{-1} \end{pmatrix};$$

$$A_1 = \begin{pmatrix} 1{,}66000 & 6{,}08112*10^{-1} & 5{,}37401*10^{-1} & 0{,}00000 \\ 6{,}08112*10^{-1} & 1{,}00000 & 3{,}20000*10^{-1} & 1{,}41422*10^{-1} \\ 5{,}37401*10^{-1} & 3{,}20000*10^{-1} & 1{,}00000 & -2{,}26274*10^{-1} \\ 0{,}00000*10^{-1} & 1{,}41422*10^{-1} & -2{,}26274*10^{-1} & 3{,}40000*10^{-1} \end{pmatrix}.$$

Точність ε: 1,6248000. Сферична норма: 6,4959984.

Ітерація 2

$$T^{(2)} = \begin{pmatrix} 8{,}59349*10^{-1} & -5{,}11390*10^{-1} & 0{,}00000 & 0{,}00000 \\ 5{,}11390*10^{-1} & 8{,}59349*10^{-1} & 0{,}00000 & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 1{,}00000 & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 0{,}00000 & 1{,}00000 \end{pmatrix};$$

$$A_2 = \begin{pmatrix} 2{,}02188 & 0{,}00000 & 6{,}25460*10^{-1} & 7{,}23216*10^{-3} \\ 0{,}00000 & 6{,}38119*10^{-1} & 1{,}69963*10^{-4} & 1{,}21530*10^{-2} \\ 6{,}25460*10^{-1} & 1{,}69963*10^{-4} & 1{,}00000 & -2{,}26274*10^{-1} \\ 7{,}23216*10^{-3} & 1{,}21530*10^{-2} & -2{,}26274*10^{-1} & 3{,}40000*10^{-1} \end{pmatrix}.$$

Точність ε: $8{,}8520002*10^{-1}$. Сферична норма: 6,4959989.

Ітерація 3

$$T^{(3)} = \begin{pmatrix} 9{,}03506*10^{-1} & 0{,}00000 & -4{,}28576*10^{-1} & 0{,}00000 \\ 0{,}00000 & 1{,}00000 & 0{,}00000 & 0{,}00000 \\ 4{,}28576*10^{-1} & 0{,}00000 & 9{,}03506*10^{-1} & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 0{,}00000 & 1{,}00000 \end{pmatrix};$$

$$A_3 = \begin{pmatrix} 2{,}31857 & 7{,}28420*10^{-5} & 0{,}00000 & -9{,}04415*10^{-2} \\ 7{,}28420*10^{-5} & 6{,}38119*10^{-1} & 1{,}53562*10^{-4} & 1{,}21530*10^{-2} \\ 0{,}00000 & 1{,}53562*10^{-4} & 7{,}03314*10^{-1} & -2{,}07540*10^{-1} \\ -9{,}04415*10^{-2} & 1{,}21530*10^{-2} & -2{,}07540*10^{-1} & 3{,}40000*10^{-1} \end{pmatrix}.$$

Точність ε: $1{,}0280008*10^{-1}$. Сферична норма: 6,4959984.

Ітерація 4

$$T^{(4)} = \begin{pmatrix} 1{,}00000 & 0{,}00000 & 0{,}00000 & 0{,}00000 \\ 0{,}00000 & 1{,}00000 & 0{,}00000 & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 9{,}10667*10^{-1} & 4{,}13142*10^{-1} \\ 0{,}00000 & 0{,}00000 & -4{,}13142*10^{-1} & 9{,}10667*10^{-1} \end{pmatrix};$$

$$A_4 = \begin{pmatrix} 2{,}31857 & 7{,}28420*10^{-5} & 3{,}73651*10^{-2} & -8{,}23620*10^{-2} \\ 7{,}28420*10^{-5} & 6{,}38119*10^{-1} & -4{,}88108*10^{-3} & 1{,}11308*10^{-2} \\ 3{,}73651*10^{-2} & -4{,}88108*10^{-3} & 7{,}97469*10^{-1} & -1{,}49012*10^{-8} \\ -8{,}23620*10^{-2} & 1{,}11308*10^{-2} & -1{,}49012*10^{-8} & 2{,}45846*10^{-1} \end{pmatrix}.$$

Точність ε: $1{,}6654763*10^{-2}$. Сферична норма: 6,4959984.

Ітерація 5

$$T^{(5)} = \begin{pmatrix} 9{,}99214*10^{-1} & 0{,}00000 & 0{,}00000 & 3{,}96424*10^{-2} \\ 0{,}00000 & 1{,}00000 & 0{,}00000 & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 1{,}00000 & 0{,}00000 \\ -3{,}96424*10^{-2} & 0{,}00000 & 0{,}00000 & 9{,}99214*10^{-1} \end{pmatrix};$$

$$A_5 = \begin{pmatrix} 2{,}32183 & -3{,}68468*10^{-4} & 3{,}73358*10^{-2} & 2{,}79397*10^{-9} \\ -3{,}68468*10^{-4} & 6{,}38119*10^{-1} & -4{,}88108*10^{-3} & 1{,}11250*10^{-2} \\ 3{,}73358*10^{-2} & -4{,}88108*10^{-3} & 7{,}97469*10^{-1} & 1{,}48123*10^{-3} \\ 2{,}79379*10^{-9} & 1{,}11250*10^{-2} & 1{,}48123*10^{-3} & 2{,}42578*10^{-1} \end{pmatrix}.$$

Точність ε: $3{,}0877562*10^{-3}$. Сферична норма: 6,4959993.

Ітерація 6

$$T^{(6)} = \begin{pmatrix} 9{,}99701*10^{-1} & 0{,}00000 & -2{,}44706*10^{-2} & 0{,}00000 \\ 0{,}00000 & 1{,}00000 & 0{,}00000 & 0{,}00000 \\ 2{,}44706*10^{-2} & 0{,}00000 & 9{,}99701*10^{-1} & 0{,}00000 \\ 0{,}00000 & 0{,}00000 & 0{,}00000 & 1{,}00000 \end{pmatrix};$$

$$A_6 = \begin{pmatrix} 2{,}32275 & -4{,}87801*10^{-4} & 1{,}86265*10^{-9} & 3{,}62495*10^{-5} \\ -4{,}87801*10^{-4} & 6{,}38119*10^{-1} & -4{,}87061*10^{-3} & 1{,}12502*10^{-2} \\ 1{,}86265*10^{-9} & -4{,}87061*10^{-3} & 7{,}96555*10^{-1} & 1{,}48079*10^{-3} \\ 3{,}62495*10^{-5} & 1{,}12502*10^{-2} & 1{,}48079*10^{-3} & 2{,}42578*10^{-1} \end{pmatrix}.$$

Точність ε: $2{,}9983872*10^{-4}$. Сферична норма: 6,4960003.

Ітерація 7

$$T^{(7)} = \begin{pmatrix} 1{,}00000 & 0{,}00000 & 0{,}00000 & 0{,}00000 \\ 0{,}00000 & 9{,}99605*10^{-1} & 0{,}00000 & -2{,}80926*10^{-2} \\ 0{,}00000 & 0{,}00000 & 1{,}00000 & 0{,}00000 \\ 0{,}00000 & 2{,}80926*10^{-2} & 0{,}00000 & 9{,}99605*10^{-1} \end{pmatrix};$$

$$A_7 = \begin{pmatrix} 2{,}32275 & -4{,}86590*10^{-4} & 1{,}86265*10^{-9} & 4{,}99388*10^{-5} \\ -4{,}86590*10^{-4} & 6{,}38283*10^{-1} & -4{,}82708*10^{-3} & 9{,}31323*10^{-10} \\ 1{,}86265*10^{-9} & -4{,}82708*10^{-3} & 7{,}96706*10^{-1} & 1{,}61703*10^{-3} \\ 4{,}99388*10^{-5} & 9{,}31323*10^{-10} & 1{,}61703*10^{-3} & 2{,}42261*10^{-1} \end{pmatrix}.$$

Точність ε: $5{,}2309584*10^{-5}$. Сферична норма: 6,4959989. Власні числа матриці A: $\lambda_1 = 2{,}32275$; $\lambda_2 = 0{,}638283$; $\lambda_3 = 0{,}796706$; $\lambda_4 = 0{,}242261$. Власні вектори $\bar{x}_i$ відповідають λ_i:

$$\bar{x}_1 = \begin{pmatrix} 0{,}579643 \\ 0{,}459997 \\ 0{,}433459 \\ 0{,}514326 \end{pmatrix}; \quad \bar{x}_2 = \begin{pmatrix} 0{,}380449 \\ -0{,}850275 \\ -0{,}358896 * 10^{-1} \\ 0{,}361941 \end{pmatrix}; \quad \bar{x}_3 = \begin{pmatrix} -0{,}503284 * 10^{-1} \\ 0{,}237226 \\ -0{,}128462 \\ 0{,}5249596 \end{pmatrix};$$

$$\bar{x}_4 = \begin{pmatrix} -0{,}718846 \\ -0{,}956990 * 10^{-1} \\ 0{,}387435 \\ 0{,}569206 \end{pmatrix}.$$

§ 4.6. *QR*-АЛГОРИТМ

Перейдемо до обчислення власних значень несиметричних матриць за допомогою ітераційних методів. Це набагато складніша задача, ніж у випадку симетричних матриць, і раціональні методи її розв'язання виявляються значно складнішими за методи, що викладені у попередніх параграфах. Більш суттєвим є те, що власні значення несиметричних матриць можуть бути дуже погано обумовлені, тобто незначна зміна елементів матриці веде до зміни власних значень. Ця можливість поганої обумовленості власних значень не дозволяє розраховувати на те, що знайдеться такий метод, який буде точно обчислювати власні значення для всіх несиметричних матриць. Якщо задана матриця A, то у найкращому випадку можна сподіватись, що метод буде давати достатньо точні власні значення для матриць типу $A + E$, де елементи матриці E у якомусь смислі малі. Якщо ж знайдені власні значення будуть суттєво відрізнятись від точних, то це пояснюється поганою обумовленістю цих власних значень.

Останнім часом для розв'язання проблеми власних значень несиметричної матриці успішно використовують QR-алгоритм Френсиса—Кублановської. Цей метод базується на зведенні вихідної матриці A до клітинно-діагональної форми, для якої легко знаходяться власні значення. Даний метод вигідний для верхніх майже трикутних матриць (зменшується число арифметичних дій на кожній ітерації). Тому на першому етапі вихідна матриця зводиться до форми Хессенберга, яка дозволяє провести розкладання за менший час. Матриці, у яких нижче головної діагоналі є тільки одна ненульова діагональ, що безпосередньо прилягає до головної, називаються *матрицями Хессенберга* ($a_{ij} = 0$, $i \geq j + 2$) і мають вигляд

$$\begin{bmatrix} X & X & X & \dots & X & X \\ X & X & X & \dots & X & X \\ 0 & X & X & \dots & X & X \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & X & X \end{bmatrix}. \qquad (4.41)$$

Вихідна матриця A може бути ефективно зведена до вигляду (4.41) за допомогою методу Гівенса, суть якого полягає в тому, що вихідна матриця множиться на елементарні матриці плоских обертань (4.25), підібраних так, щоб при обертанні анулювались відповідні елементи. Зведення за методом Гівенса має $(n-2)$ основних етапів, на r-м із яких з'являються нулі у r-му рядку і у r-му стовпчику, при цьому нулі, одержані на попередніх $(r-1)$ етапах, не зникають. На початку r-го основного етапу перші $(r-1)$ рядків і стовпців утворюють тридіагональну матрицю. Проілюструємо це для випадку $n=6$, $r=3$.

$$\begin{bmatrix} X & X & 0 & \dots & 0 & 0 & 0 \\ X & X & X & \dots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & X & X & \dots & X & \underline{X} & \underline{X} \\ 0 & 0 & X & \dots & X & \underline{X} & \underline{X} \\ 0 & 0 & \underline{X} & \dots & X & X & X \\ 0 & 0 & \underline{X} & \dots & X & X & X \end{bmatrix} \quad (4.42)$$

Основний r-й етап складається із $(n-r-1)$ допоміжних етапів, на яких нулі з'являються послідовно у позиціях $r+2, r+3, \dots, n$ r-го рядка і r-го стовпця. У (4.42) підкреслені елементи, які виключаються на r-му основному етапі, і відмічені стрілками ті рядки і стовпці, які використовуються. Одразу ж видно, що перші $(r-1)$ рядки і стовпці цілком не змінюються, хоча деякі ненульові елементи беруть участь у перетворенні, вони замінюються лінійною комбінацією нулів. Тому у дійсності на r-му основному етапі ми маємо справу лише з матрицею порядку $(n-r+1)$ у нижньому правому куті поточної матриці. Нуль в i-й позиції з'являється за допомогою обертання у площині $(r+1, i)$, тобто при множенні матриці зліва на матрицю обертання (4.25), де

$$c = \cos\theta = \frac{a_{r+1,r}}{\sqrt{a_{r+1,r}^2 + a_{i,r}^2}};$$

$$s = \sin\theta = \frac{a_{i,r}}{\sqrt{a_{r+1,r}^2 + a_{i,r}^2}}. \quad (4.43)$$

Якщо $x = a_{r+1,r}^2 + a_{i,r}^2 = 0$, то беремо $\cos\theta = 1$, $\sin\theta = 0$ і справа домножуємо на обернену матрицю T^{-1} $(T^{-1} = T')$ T — ортогональна матриця). При множенні справа на обернену матрицю обертання форма Хессенберга зберігається.

Дослідження показали, що цей метод чисельно стійкий і дає добру точність.

Припустимо, що матриця A зведена до форми Хессенберга, й перейдемо до другого етапу — розгляд питання про застосування до неї QR-алгоритму. Провідна ідея QR-алгоритму полягає у поданні матриці Хессенберга у вигляді добутку $A = QR$, де Q — ортогональна, R — верхньотрикутна матриці. Переписуючи співвідношення $A = QR$ у вигляді $R = Q^{-1}A$, зводимо задачу до знаходження такої ортогональної матриці Q, для якої

$Q^{-1}A$ було б верхньотрикутною матрицею. Знову використовуємо метод Гівенса. Домножуючи матрицю Хессенберга вигляду (4.41) зліва на матрицю обертання (4.25), у якої синуси та косинуси розташовані у лівому верхньому куті, дістанемо

$$\begin{bmatrix} \sin\theta & \cos\theta & & \\ -\cos\theta & \sin\theta & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11}\sin\theta + a_{21}\cos\theta & \dots & a_{1n}\sin\theta + a_{2n}\cos\theta \\ -a_{11}\cos\theta + a_{21}\sin\theta & \dots & -a_{1n}\cos\theta + a_{2n}\sin\theta \\ a_{32} & \dots & a_{3n} \\ \dots & \dots & \dots \\ & & a_{nn} \end{bmatrix}.$$

Якщо вибрати θ із умови $-a_{11}\cos\theta + a_{21}\sin\theta = 0$ або $\theta = \operatorname{arctg}(a_{11}/a_{21})$, то елемент a_{21} при цьому обернеться на нуль. Таким чином, у результаті множення на послідовність матриць плоских обертань T_i із відповідними значеннями кутів, всі піддіагональні елементи матриці Хессенберга послідовно обернуться на 0. Дістанемо $R = T'_{n-1}T'_{n-2}\dots T'_1A$. Звідки $A = T_1T_2\dots T_{n-1}R$ або $A = QR$, де $Q = T_1T_2\dots T_{n-1}$ — ортогональна матриця. Далі будуємо нову матрицю $A_1 = RQ$. Оскільки $A = QR = Q(RQ)Q^{-1} = QA_1Q^{-1}$, то, як видно, матриці A і A_1 подібні, тобто мають однакові власні значення. QR-факторизація проводиться із матрицею A_1. Здійснюючи факторизацію та перестановку співмножників, отримаємо послідовність матриць

$$A_k = Q_kR_k; \quad A_{k+1} = R_kQ_k, \quad k = 0, 1, \dots .$$

Одержана послідовність подібних матриць A_k при $k \to \infty$ прямує до клітинно-діагонального вигляду.

Зведення матриці A до форми Хессенберга не мало б сенсу, якби цю процедуру потрібно було виконувати після кожного кроку алгоритму. Дійсно, відмінні від 0 позадіагональні елементи матриць T_i, розташовані лише у позиціях $(i+1, i)$ та $(i, i+1)$. Звідси випливає, що матриця Q сама є матрицею Хессенберга. Оскільки R — верхня трикутна матриця, добуток RQ також є матрицею Хессенберга. Таким чином, якщо вихідна матриця зведена до форми Хессенберга, то і всі матриці A_k, що генеруються QR-алгоритмом, будуть автоматично зберігати цю форму.

Потрібно зазначити, що навіть при початковому зведенні до форми Хессенберга і випливаючої звідси економії при побудові матриць A_k, алгоритм може виявитись неефективним через повільну збіжність до нуля елементів, що містяться під діагоналлю. Швидкість збіжності описується виразом

$$a_{ij}^{(k)} = o\left(\frac{|\lambda_i|^k}{|\lambda_j|^k}\right), \quad k \to \infty, \quad i > j, \tag{4.44}$$

із якого видно, що, коли два власних значення (припустимо λ_i та λ_{i+1}) близькі, то позадіагональний елемент із індексами $(i+1, i)$ буде прямувати до нуля дуже повільно.

Пропонуємо послабити цю проблему збіжності таким чином. Припустимо, що $\overline{\lambda}_n$ є непоганим наближенням до λ_n, і розглянемо матрицю $\overline{A} = A - \overline{\lambda}_n E$, власні значення якої дорівнюють $\lambda_1 - \overline{\lambda}_n, \ldots, \lambda_n - \overline{\lambda}_n$. Якщо застосувати QR-алгоритм до $\overline{A}$, то позадіагональні елементи останнього рядка матриці A_k будуть прямувати до нуля як степені відношень $(\lambda_n - \overline{\lambda}_n)/(\lambda_i - \overline{\lambda}_n)$, а не як степені λ_n/λ_i $(i = 1, \ldots, n-1)$. Найповільніше із них збігається елемент у позиції $(n, n-1)$, причому для матриці $\overline{A}$ швидкість збіжності описується відношенням $(\lambda_n - \overline{\lambda}_n)/(\lambda_{n-1} - \overline{\lambda}_n)$, а для вихідної матриці A — відношенням λ_n/λ_{n-1}. Якщо, наприклад $\lambda_n = 0{,}99$, $\lambda_{n-1} = 1{,}1$ і $\overline{\lambda}_n = 2{,}0$, то $\lambda_n/\lambda_{n-1} = 0{,}9$, а $|\lambda_n - \overline{\lambda}_n| / |\lambda_{n-1} - \overline{\lambda}_n| = 0{,}1$, тобто елемент $(n, n-1)$ матриці із зсунутими власними значеннями прямує до нуля у 20 разів швидше.

Добре наближення $\overline{\lambda}_n$, яке слід було б використовувати як параметр зсуву, нам, як правило, невідоме. Але під час QR-алгоритму елементи $a_{nn}^{(k)}$ матриць A_k будуть прямувати до λ_n, тобто ми можемо їх використовувати як параметри зсувів, тобто, зробивши k кроків, на кроці $k+1$ можемо виконати QR-розклад матриці $\overline{A}_k = A_k - a_{nn}^{(k)} E$. У той же час можна виконати перетворення зсуву на кожному кроці QR-алгоритму, використовуючи як параметр елемент (n, n) поточної матриці. Оскільки при кожному зсувові власні значення вихідної матриці змінюються на величину зсуву, то необхідно стежити за нагромадженням цих величин. Фактично, саме сума зсуву збігається до власного значення λ_n. Критерієм збіжності виступає достатня мализна елементів останнього рядка. Коли цього буде досягнуто, можна відкинути останні рядок і стовпець матриці і перейти до відшукання власного значення λ_{n-1} на основі отриманої матриці розміру $(n-1) \times (n-1)$.

Зазначимо, що власні значення цієї підматриці, а отже,й вихідної матриці, були змінені на сумарну величину усіх зсувів (яка служила наближенням до λ_n); отже, після обчислення власних значень матриці до них треба додати цю величину. Можна вчинити інакше, повертаючи зсув назад після кожного кроку QR-алгоритму, тоді матриці A_k будуть мати ті ж самі власні значення.

Приклад. За допомогою методу QR-алгоритму знайти власні значення матриці A.

Обчислення виконати із точністю $E = 10^{-5}$.

$$A = \begin{vmatrix} 0{,}40463 & 0{,}59641 & 1{,}00000 & 1{,}00000 \\ 0{,}89066 & 1{,}00000 & 0{,}74534 & 0{,}36879 \\ 0{,}67339 & 0{,}79956 & 0{,}46333 & 0{,}52268 \\ 2{,}00000 & 0{,}25761 & 0{,}61755 & 0{,}12392 \end{vmatrix}.$$

Матриця у формі Хессенберга

$$H = \begin{vmatrix} 0{,}404630 & 1{,}399033 & 0{,}521915 & 0{,}354987 \\ 2{,}290574 & 0{,}967666 & 0{,}932533 & 0{,}102094 \\ 0{,}000000 & 0{,}946979 & 0{,}624823 & -0{,}351149 \\ 0{,}000000 & 0{,}000000 & -0{,}234233 & -0{,}005239 \end{vmatrix}.$$

Трикутна форма матриці

$$K = \begin{vmatrix} 2{,}86362 & 0{,}74984 & -0{,}18953 & -0{,}08824 \\ 0{,}00001 & -1{,}14335 & 0{,}62901 & -0{,}40016 \\ 0{,}00000 & 0{,}00000 & 0{,}51392 & 0{,}23627 \\ 0{,}00000 & 0{,}00000 & 0{,}00000 & -0{,}24231 \end{vmatrix}.$$

У результаті виконання 15 ітерацій одержана трикутна матриця, у якої на діагоналі розташовані власні значення вихідної матриці $\lambda_1 = -2{,}86362$; $\lambda_2 = -1{,}14335$; $\lambda_3 = 0{,}51392$; $\lambda_4 = -0{,}24231$.

§ 4.7. МЕТОД ОБЕРНЕНИХ ІТЕРАЦІЙ

Для відшукання власних векторів матриці використовується метод обернених ітерацій. Розглянемо його. Вибираємо довільний вектор $\bar{b}$ і лінійну неоднорідну систему

$$(A - \bar{\lambda}_k E)\bar{x} = \bar{b}, \tag{4.45}$$

де $\bar{\lambda}_k$ — наближене значення для власного числа λ_k. Оскільки $\det(A - \bar{\lambda}_k E) \neq 0$, то система має єдиний розв'язок. Покажемо, що знайдений із неї вектор $\bar{x}$ майже дорівнює власному вектору $\bar{x}_k$, що відповідає власному значенню $\bar{\lambda}_k$.

Для простоти обмежимось випадком, коли матриця n-го порядку має n лінійно незалежних власних векторів $\bar{x}_j$, $j = 1, \ldots, n$. Тоді власні вектори утворюють базис, по якому можна розкласти вектори $\bar{x}$ і $\bar{b}$:

$$\bar{x} = \sum_{j=1}^{n} a_j \bar{x}_j, \quad \bar{b} = \sum_{j=1}^{n} c_j \bar{x}_j.$$

Підставляючи ці розклади у систему (4.45) і враховуючи, що $A\bar{x}_j = \lambda_j \bar{x}_j$, дістанемо

$$\sum_{j=1}^{n} a_j(\lambda_j - \bar{\lambda}_k)\bar{x}_j = \sum_{j=1}^{n} c_j \bar{x}_j.$$

Звідси в силу лінійної незалежності $\bar{x}_j$ випливає, що при довільному j

$$a_j(\lambda_j - \bar{\lambda}_k) = c_j, \text{ або } a_j = c_j/(\lambda_j - \bar{\lambda}_k).$$

Як видно з формули, якщо $j = k$, то $\lambda_k \approx \tilde{\lambda}_k$, і коефіцієнт a_k буде дуже великим, у противному разі він невеликий. Іншими словами, при оберненій ітерації, тобто при переході від $\bar{b}$ до $\bar{x}$, компонента a_k сильно збільшується у порівнянні з іншими компонентами і вектор $\bar{x}$ виявляється близьким до $\bar{x}_k$.

У випадку, коли вектор $\bar{b}$ вибраний невдало, знайдений вектор $\bar{x}$ може значно відрізнятись від $\bar{x}_k$, тоді ітерації слід повторити за формулами

$$(A - \tilde{\lambda}_k E)\bar{x}^{(s)} = \bar{x}^{(s-1)}, \quad \bar{x}^{(0)} = \bar{b}. \tag{4.46}$$

Звичайно двох-трьох ітерацій достатньо, при цьому на кожній з них обов'язково потрібно нормувати знайдені вектори, щоб не отримувати у розрахунках занадто великих чисел, які викликають на ЕОМ переповнення.

Метод обернених ітерацій застосовується для знаходження власних векторів як у випадку простих λ_k, так і у випадку кратних власних значень. Щоб знайти усі власні вектори для кратного власного значення, кількість лінійно незалежних векторів $\bar{b}_r$ повинна відповідати кратності кореня. Оберненими ітераціями для кожного $\bar{b}$ буде побудовано вектор $\bar{x}_r$. Серед цих векторів буде стільки лінійно незалежних, скільки власних векторів відповідають власному значенню λ_k.

Знаходження власного вектора потребує (на одну ітерацію) не більше $2/3 \cdot n^3$ арифметичних дій, тобто для знайдення всіх їх потрібно близько n^4 арифметичних дій. Таким чином, при великих порядках матриці метод буде не економічним, але при $n \leq 10$ досить задовільним. Особливу зручність дає його простота, універсальність і стабільність стійкості алгоритму.

У деяких окремих випадках розрахунки суттєво спрощуються і прискорюються. Найбільш важливий випадок тридіагональної матриці. При цьому лінійна система рівнянь (4.45) для визначення компонент власних векторів також буде тридіагональною, і її розв'язують економічним методом прогонки.

Виділимо одне суттєве зауваження. Оскільки $\det(A - \tilde{\lambda}_k E) \approx 0$, то при відшуканні власних векторів у формулах прямого ходу методу виключень (прогонки) на головній діагоналі з'явиться хоча б один дуже малий елемент. Для формального ведення розрахунків діагональні елементи не повинні обертатись на нуль; для цього потрібно, щоб похибка власного значення була не дуже малою, тобто становила б 10—15 останніх двійкових розрядів числа на ЕОМ. Щоб уникнути ділення на 0, треба вносити у λ_i невеликі похибки.

Приклад. Знайти власні значення матриці A за допомогою методу QR-алгоритму та власні вектори за методом ітерацій:

$$A = \begin{vmatrix} 1{,}022551 & 0{,}115069 & -0{,}287028 & -0{,}429969 \\ 0{,}228401 & 0{,}742521 & -0{,}176368 & -0{,}283720 \\ 0{,}326141 & 0{,}097221 & 0{,}197209 & -0{,}216487 \\ 0{,}433864 & 0{,}148965 & -0{,}193686 & 0{,}0064772 \end{vmatrix}.$$

Матриця у формі Хессенберга

$$H = \begin{vmatrix} 1{,}022551 & -0{,}430736 & 0{,}020506 & -0{,}307858 \\ 0{,}588874 & -0{,}047172 & 0{,}039884 & -0{,}511079 \\ 0{,}000000 & 0{,}016830 & 0{,}325898 & 0{,}029246 \\ 0{,}000000 & 0{,}000000 & 0{,}000024 & 0{,}667481 \end{vmatrix}.$$

Трикутна форма матриці

$$R = \begin{vmatrix} 0{,}667480 & 0{,}000000 & 0{,}418900 & -1{,}086180 \\ 0{,}000000 & 0{,}667480 & -0{,}051130 & 0{,}195630 \\ 0{,}000000 & 0{,}000000 & 0{,}346180 & -0{,}035130 \\ 0{,}000000 & 0{,}000000 & 0{,}000040 & 0{,}287620 \end{vmatrix}.$$

У результаті виконання 40 ітерацій одержані власні числа матриці A: $\lambda_1 = 0{,}66748$; $\lambda_2 = 0{,}66748$; $\lambda_3 = 0{,}34618$; $\lambda_4 = 0{,}28762$.

Власні вектори

$$\bar{x}_1 = (0{,}555003;\ -1{,}02163;\ 0{,}129450;\ 0{,}0961189);$$

$$\bar{x}_2 = (0{,}976771;\ 0{,}636737;\ 0{,}517676;\ 0{,}632935);$$

$$\bar{x}_3 = (-0{,}611585;\ -0{,}366955;\ -1{,}22309;\ -0{,}244691);$$

$$\bar{x}_4 = (-0{,}598402;\ -0{,}398937;\ -0{,}199437;\ -0{,}997345).$$

§ 4.8. РОЗВ'ЯЗАННЯ ЧАСТКОВОЇ ПРОБЛЕМИ ВЛАСНИХ ЗНАЧЕНЬ

У випадку дуже великих розріджень матриць, для яких треба визначити тільки кілька власних значень, розглянуті методи не особливо корисні. Такі проблеми часто виникають як в рівняннях з частковими похідними, так і в інших областях. Наведемо характерний приклад такої задачі: є матриця розміром 5000×5000, в кожному рядку якої міститься порядку десяти відмінних від нуля елементів, і потрібно знайти тільки кілька (може, три або чотири) власних значень. Для такої задачі попередні методи взагалі не підходять. Мають також місце задачі, коли треба визначити максимальне власне значення матриці A та відповідний власний вектор, тобто розв'язати часткову проблему власних значень.

Класичним методом, який іноді корисний для великих розріджених систем у загальному випадку, є степеневий. Його використовують для розв'язання часткової проблеми власних значень.

Припустимо, що власні значення $\lambda_1, \lambda_2, \ldots, \lambda_n$ матриці A дійсні і задовольняють умову

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|. \tag{4.47}$$

При цьому позначимо відповідні власні вектори через $\bar{x}^{(1)}, \ldots, \bar{x}^{(n)}$. Візьмемо довільний вектор $\bar{y}^{(0)} = (y_1^{(0)}, \ldots, y_n^{(0)})$ і побудуємо рекурентну послідовність векторів

$$\bar{y}^{(0)}, \bar{y}^{(1)} = A\bar{y}^{(0)}, \bar{y}^{(2)} = A\bar{y}^{(1)} = A^2\bar{y}^{(0)}, \ldots, \bar{y}^{(k)} = A\bar{y}^{(k-1)} = \ldots = A^k\bar{y}^{(0)}.$$

Щоб з'ясувати, яким буде $y^{(k)}$ при великих k, розкладаємо $y^{(0)}$ за власними векторами $\bar{x}^{(i)}$:

$$\bar{y}^{(0)} = a_1\bar{x}^{(1)} + a_2\bar{x}^{(2)} + \ldots + a_n\bar{x}^{(n)}. \tag{4.48}$$

Прийнявши до уваги, що $A^k\bar{x}^{(i)} = \lambda_i^k\bar{x}^{(i)}$, дістанемо

$$\bar{y}^{(k)} = A^k\bar{y}^{(0)} = a_1\lambda_1^k\bar{x}^{(1)} + a_2\lambda_2^k\bar{x}^{(2)} + \ldots + a_n\lambda_n^k\bar{x}^{(n)}. \tag{4.49}$$

Припустимо, що $|\lambda_1| > |\lambda_2|$ і $a_1 \neq 0$. Тоді при великих значеннях k в правій частині (4.49) перший доданок буде, очевидно, головним. Для знаходження λ_1 векторну рівність (4.49) зручніше записати в складових. Введемо таке означення:

$$\bar{y}^{(k)} = (y_1^{(k)}, y_2^{(k)}, \ldots, y_n^{(k)}), \quad \bar{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}).$$

Рівність (4.49) рівносильна n чисельним різностям:

$$y_s^{(k)} = \beta_{1s}\lambda_1^k + \beta_{2s}\lambda_2^k + \ldots + \beta_{ns}\lambda_n^k, \quad \beta_{is} = a_i x_s^{(i)}, \quad s = 1, \ldots, n. \tag{4.50}$$

Відношення складових $y_s^{(k+1)}$ та $y_s^{(k)}$ буде дорівнювати

$$\frac{y_s^{(k+1)}}{y_s^{(k)}} = \frac{\beta_{1s}\lambda_1^{k+1} + \beta_{2s}\lambda_2^{k+1} + \ldots + \beta_{ns}\lambda_n^{k+1}}{\beta_{1s}\lambda_1^k + \beta_{2s}\lambda_2^k + \ldots + \beta_{ns}\lambda_n^k} =$$

$$= \lambda_1 \frac{1 + \gamma_{2s}\mu_2^{k+1} + \ldots + \gamma_{ns}\mu_n^{k+1}}{1 + \gamma_{2s}\mu_2^k + \ldots + \gamma_{ns}\mu_n^k}, \tag{4.51}$$

де $\gamma_{is} = \dfrac{\beta_{is}}{\beta_{1s}}$, $\quad \mu_i = \dfrac{\lambda_i}{\lambda_1}$.

Оскільки $|\mu_i| < 1$, $(i > 1)$, при наближеному зростанні k правильне співвідношення

$$\frac{y_S^{(k+1)}}{y_S^{(k)}} = \lambda_1 + 0\,(|\mu_2|^k), \tag{4.52}$$

і для достатньо великих k з прийнятою точністю

$$\lambda_1 \approx y_S^{(k+1)}/y_S^{(k)}. \tag{4.53}$$

Права частина залежить від номера S взятої складової і, якщо ця частина буде мати однакове значення при будь-яких S у межах прийнятої точності, то це є деякою, щоправда, не достовірною гарантією того, що взято достатньо велике значення k.

При достатньо великих k в зображенні (4.49) вектора $\overline{y}^{(k)}$ всі доданки справа, починаючи з другого, будуть мати значення менше прийнятої похибки обчислень, і залишиться лише перший доданок. Звідси отримуємо правило для наближеного знаходження власного вектора $\overline{x}^{(1)}$

$$\overline{x}^{(1)} \approx \overline{y}^{(k)}/a_1\lambda_1^k.$$

Оскільки $\overline{x}^{(1)}$ визначено лише з точністю до чисельного множника, то постійний множник $(a_1\lambda_1^k)^{-1}$ можна замінити будь-яким числом і вважати $\overline{x}^{(1)} = c_k\overline{y}^{(k)}$.

Головною перевагою степеневого методу є те, що вектори $\overline{y}^{(k)}$ одержуються тільки за допомогою множення матриці на вектор, ніяких перетворень самої матриці A при цьому робити не потрібно. Основний недолік цього методу полягає в тому, що він може збігатись дуже повільно. Як видно з (4.51), швидкість збіжності в першу чергу визначається відношенням λ_2/λ_1. Якщо це відношення за модулем близьке до 1, що характерно для багатьох практичних задач, то швидкість буде повільною. Можна спробувати обминути цю складність за допомогою зсувів, як це було зроблено в QR-алгоритмі. Якщо застосувати степеневий метод до матриці $A - \sigma E$ з власними значеннями $\lambda_1 - \sigma_1, \ldots, \lambda_n - \sigma_n$, то швидкість збіжності буде визначатися відношенням $|\lambda_2 - \sigma|/|\lambda_1 - \sigma|$. За умови, звичайно, що $\lambda_1 - \sigma$ залишиться максимальним по модулю власним значенням. У припущенні дійсності всіх λ_i можна показати, що мінімум цього відношення досягається при $\sigma = (\lambda_2 + \lambda_1)/2$. Але навіть при цьому оптимальному виборі збіжність може все ж таки виявитись дуже повільною.

Приклад 1. За допомогою степеневого методу знайти найбільше власне число матриці (4.40) та відповідний власний вектор.

Розв'язання. Послідовні ітерації наближених значень власного числа з відповідною точністю подані в табл. 6 (*а*)

Власний вектор, що відповідає $\lambda_{max} = 2{,}32269$:

$$\overline{x} = (0{,}579653;\ 0{,}459957;\ 0{,}433569;\ 0{,}514257).$$

Таблиця 6

Номер ітерації	Наближене значення λ_{max}	Точність
	а	
1	2,42224	3,15826
2	2,32802	$0{,}942228 * 10^{-1}$
3	2,32116	$0{,}685978 * 10^{-2}$
4	2,32175	$0{,}588894 * 10^{-3}$
5	0,232234	$0{,}590801 * 10^{-3}$
6	0,232260	$0{,}258923 * 10^{-3}$
7	2,32269	$0{,}9799 * 10^{-4}$
	б	
1	2,28994	0,225514
2	2,31894	$0{,}289927 \cdot 10^{-1}$
3	2,32231	$0{,}337172 \cdot 10^{-2}$
4	2,32270	$0{,}390053 \cdot 10^{-3}$
5	2, 32274	$0{,}455379 \cdot 10^{-4}$

Степеневим методом можна знайти будь-яке власне число матриці. Нехай потрібно знайти λ_2 при умові $|\lambda_2| < |\lambda_1|$. Тоді у послідовності $\overline{y}^{(k)}$ ($k = 0, 1, \ldots$) слід виключити із $\overline{y}^{(k)}$ частину, що містить λ_1. Головною частиною буде доданок, що містить λ_2, який можна знаходити вищевказаним шляхом. Але при великих k доданок із λ_1 виступає головною частиною, і виключення її пов'язане із великою втратою точності обчислень. Тому обчислення λ_2 буває пов'язане з необхідністю значного збільшення точності обчислень, зокрема з необхідністю знаходити λ_1 із більшою кількістю вірних значущих цифр і з більшою втратою точних знаків при визначенні λ_2. Ще більше втрачаються вірні знаки при знаходженні власних значень, менших за модулем за λ_2. Але це трапляється нечасто.

Розглянемо питання про відшукання λ_2 у випадку, коли дійсні нерівності

$$|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \ldots \tag{4.54}$$

Нехай дано порівнянні частини номера s у $y^{(k)}$ та $y^{(k+1)}$

$$y_s^{(k)} = \beta_{1s}\lambda_1^k + \beta_{2s}\lambda_2^k + \beta_{3s}\lambda_3^k + \ldots;$$

$$y_s^{(k+1)} = \beta_{1s}\lambda_1^{k+1} + \beta_{2s}\lambda_2^{k+1} + \beta_{3s}\lambda_3^{k+1} + \ldots \tag{4.55}$$

Вилучимо член із λ_1, для чого побудуємо комбінацію

$$y_s^{(k+1)} - \lambda_1 y_s^{(k)} = \beta_{2s}\lambda_2^k(\lambda_2 - \lambda_1) + \beta_{3s}\lambda_3^k(\lambda_3 - \lambda_1) + \ldots \tag{4.56}$$

Заміною k на $k-1$ отримаємо ще одну рівність

$$y_s^{(k)} - \lambda_1 y_s^{(k-1)} = \beta_{2s}\lambda_2^{k-1}(\lambda_2 - \lambda_1) + \beta_{3s}\lambda_3^{k-1}(\lambda_3 - \lambda_1) + \ldots \qquad (4.57)$$

При $\beta_{2s} \neq 0$ із двох останніх рівностей дістанемо

$$\frac{y_s^{(k+1)} - \lambda_1 y_s^{(k)}}{y_s^k - \lambda_1 y_s^{k-1}} = \lambda_2 \frac{1 + \gamma_{3s}^* \mu_3^k + \gamma_{4s}^* \mu_4^k + \ldots}{1 + \gamma_{3s}^* \mu_3^{k-1} + \gamma_{4s}^* \mu_4^{k-1} + \ldots} = \lambda_2 \cdot [1 + o(\mu_3^k)];$$

$$\gamma_{is}^* = \frac{\beta_{is}(\lambda_i - \lambda_1)}{\beta_{2s}(\lambda_2 - \lambda_1)}; \quad \mu_i = \lambda_i/\lambda_2 \quad (i = 3, 4, \ldots).$$

Із $|\mu_2| < 1$ випливає правило наближеного обчислення λ_2 вірне з тим більшою точністю, чим більше значення має k.

$$\lambda_2 \approx \frac{y_s^{(k+1)} - \lambda_1 y_s^{(k)}}{y_s^{(k)} - \lambda_1 y_s^{(k-1)}}. \qquad (4.58)$$

При використанні формули (4.58) слід мати на увазі, що при знаходженні λ_1 число k вибирається настільки великим, щоб за прийнятої кількості вірних знаків у правих частинах рівностей вигляду (4.55) усі члени, починаючи із других, знаходились поза прийнятою точністю і їх можна було б відкинути. Тоді у правій частині зберігаються тільки перші члени, і λ_1 знайдеться як відношення $y_s^{(k+1)}$ до $y_s^{(k)}$. Нехай λ_1 знайдено. При знаходженні λ_2 слід зменшити значення k настільки, щоб у правих частинах рівності (4.55) у прийнятій точності ще зберігалися другі стовпці із λ_2, і поза цією точністю залишились члени із λ_3, λ_4... Аналогічне повинно відбуватись і при заміні k на $k - 1$. Але тоді у правих частинах рівності (4.55) перші члени будуть перевищувати другі, а другі — перевищуватимуть треті у менше число разів, ніж при знаходженні λ_1. Відповідно, величина λ_2, яку ми визначаємо із (4.58), міститиме менше достовірних знаків у порівнянні з λ_1.

У формулі (4.58) права частина залежить від номера s обраної порівнянної, і число цифр, що збігаються, у значеннях λ_2, отриманих для різних s, дає деяку неповну можливість судити про дійсне число одержаних вірних знаків.

Для симетричної матриці A можна вказати інший обчислювальний процес, який прямує до найбільшого за модулем власного значення. Симетрична матриця A має повну систему власних векторів $\bar{x}^{(1)}, \ldots, \bar{x}^{(n)}$, і їх завжди можна вважати ортонормованими.

Формули (4.50) дозволяють обчислити скалярний добуток

$$(\bar{y}^{(k)}, \bar{y}^{(k)}) = a_1^2\lambda_1^{2k} + a_2^2\lambda_2^{2k} + \ldots + a_n^2\lambda_n^{2k};$$

$$(\bar{y}^{(k+1)}, \bar{y}^{(k)}) = a_1^2\lambda_1^{2k+1} + a_2^2\lambda_2^{2k+1} + \ldots + a_n^2\lambda_n^{2k+1}.$$

Тому при збільшенні k

$$\frac{(\bar{y}^{(k+1)}, \bar{y}^{(k)})}{(\bar{y}^{(k)}, \bar{y}^{(k)})} = \lambda_1 + o(|\mu_2|^{2k}); \tag{4.59}$$

і для достатньо великих k має місце рівність

$$\lambda_1 \approx \frac{(\bar{y}^{(k+1)}, \bar{y}^{(k)})}{(\bar{y}^{(k)}, \bar{y}^{(k)})}. \tag{4.60}$$

Із порівняння (4.52) із (4.59) видно, що перше з розглянутих правил обчислення λ_1 має швидкість збіжності не повільніше геометричної прогресії зі знаменником $|\mu_2|$, тоді як у другому способі аналогічним знаменником є $|\mu_2|^2$.

Приклад 2. За допомогою методу скалярних добутків знайти найбільше власне число матриці (4.40) і відповідний власний вектор.

Розв'язання. Послідовні ітерації значень власного числа з відповідаючою даному номеру точністю запишемо у табл. 6 (*б*).

Власний вектор, що відповідає $\lambda_{max} = 2{,}32274$:

$$\bar{x} = (0{,}579653;\ 0{,}459957;\ 0{,}433569;\ 0{,}514257).$$

ГЛАВА 5

МЕТОДИ РОЗВ'ЯЗАННЯ ЗАДАЧ КОШІ ДЛЯ ЗВИЧАЙНИХ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ

§ 5.1. ВСТУПНІ ЗАУВАЖЕННЯ. ПОСТАНОВКА ЗАДАЧІ

Більшість реальних процесів у науці і техніці описуються системою диференціальних рівнянь у частинних похідних. Але безпосередніх методів розв'язування їх не існує і розв'язуються вони за допомогою зведення тим чи іншим способом до звичайних диференціальних рівнянь або до систем лінійних алгебраїчних рівнянь. Конкретна прикладна задача може бути зведена до диференціальних рівнянь будь-якого порядку, або до системи рівнянь будь-якого порядку. Нехай y — функція однієї змінної x. Звичайним диференціальним рівнянням для функції y називають співвідношення у вигляді

$$F(x, y(x), y'(x), \ldots, y^{(n)}(x)) = 0, \tag{5.1}$$

де F — задана функція від $n+2$ змінних, а незалежна змінна x змінюється в деякому скінченному або нескінченному інтервалі. Рівняння (5.1) являє собою найбільш загальний вигляд звичайного диференціального рівняння порядку n, де порядок визначається як порядок старшої похідної від невідомої функції y, що входить у це рівняння. Загалом припускається, що рівняння може бути розв'язане відносно старшої похідної, і, таким чином, його можна записати у вигляді

$$y^{(n)}(x) = f(x, y(x), y'(x), \ldots, y^{(n-1)}(x)). \tag{5.2}$$

Якщо функція f залежить від функції y і її похідних лінійно, то рівняння називається лінійним

$$y^{(n)}(x) = a_0(x) + a_1(x)y(x) + \ldots + a_{n-1}(x)y^{(n-1)}(x), \tag{5.3}$$

де $a_0(x), a_1(x), \ldots, a_{n-1}(x)$ — задані функції.

Рівняння (5.2) можна розглядати і у випадку, коли y і f є вектор-функціями; при цьому одержуємо систему рівнянь n-го порядку. Найпростішим випадком є система рівнянь n-го порядку

$$\bar{y}'(x) = \bar{f}(x, \bar{y}(x)), \tag{5.4}$$

де $\bar{y}$ і $\bar{f}$ — n-вимірні вектори з координатами $y_1, \ldots, y_n$ і $f_1, \ldots, f_n$.

Взагалі достатньо розглянути саме систему рівнянь першого порядку, оскільки одне рівняння n-го порядку зводиться до системи з n рівнянь першого порядку (і, отже, систему m рівнянь n-го порядку можемо звести до системи nm рівнянь першого порядку). Здійснити таке зведення можна, наприклад, таким чином: введемо нові змінні

$$y_1 = y',\ y_2 = y'',\ \ldots,\ y_{n-1} = y^{(n-1)}, \tag{5.5}$$

за допомогою яких звичайне диференціальне рівняння n-го порядку (5.2) зводиться до еквівалентної системи n рівнянь першого порядку.

Система рівнянь (5.4) має множину розв'язків, яка в загальному випадку залежить від n параметрів $c_1, c_2, \ldots, c_n$ і може бути записана у формі

$$\bar{y} = \bar{y}(x, \bar{c}). \tag{5.6}$$

Щоб виділити єдиний розв'язок системи (5.4), необхідно на функції $y_k(x)$ накласти n додаткових умов. Задача Коші для системи рівнянь (5.4) полягає в тому, щоб знайти функції $y_1, y_2, \ldots, y_n$, які задовольняють цю систему і початкові умови

$$y_1(x_0) = y_{10},\ y_2(x_0) = y_{20},\ \ldots,\ y_n(x_n) = y_{n0}. \tag{5.7}$$

Ці умови можна розглядати як задання координат початкової точки $(x_0, y_{10}, \ldots, y_{n0})$ інтегральної кривої в $(n+1)$-вимірному просторі $(x_1, y_1, y_2, \ldots, y_n)$. Звичайно розв'язок потрібно знайти на деякому відрізку $x_0 \le x \le x_n$; тому точку x_0 можна вважати початком відрізку.

Якщо праві частини (5.4) неперервні й обмежені в деякому околі початкової точки $(x_0, y_{10}, \ldots, y_{n0})$, то задача Коші (5.4), (5.7) має розв'язок, але, загалом, не єдиний. Якщо праві частини не тільки неперервні, а й задовольняють умову Ліпшіца по змінних y_k

$$|f_i(x, y_1', y_2', \ldots, y_n') - f(x, y_1'', y_2'', \ldots, y_n'')| \le$$
$$\le L\{|y_1' - y_1''| + |y_2' - y_2''| + \ldots + |y_n' - y_n''|\}, \tag{5.8}$$

для будь-яких точок $(x, y_1', \ldots, y_n')$ і $(x, y_1'', \ldots, y_n'')$ околу, що розглядається, то існує єдиний розв'язок задачі Коші

$$y_1 = y_1(x),\ y_2 = y_2(x),\ \ldots,\ y_n = y_n(x), \tag{5.9}$$

який неперервно залежить від координат початкової точки (5.7), тобто задача Коші поставлена коректно. Якщо, крім того, праві частини мають неперервні похідні по всіх аргументах до q-го порядку включно, то розв'язок $\bar{y}(x)$ має $q+1$ похідну по x.

Якщо розглядається задача Коші для системи нелінійних звичайних диференціальних рівнянь, а саме:

$$\bar{y}' = \bar{f}(x, \bar{y});\quad x \in [0, l];\quad \bar{y}_0 = \bar{d},$$

то виникає питання, чи існує розв'язок цієї задачі і чи він єдиний. У цьому випадку існування розв'язку задачі Коші на всьому проміжку інтегрування [0,1] не завжди має місце. Крім того, навіть для неперервних і достатньо гладких на всьому проміжку [0, 1] вектор-функцій $\bar{f}(x, \bar{y})$ розв'язок задачі Коші може не існувати. Це можна показати на такому прикладі:

$$\frac{dy}{dx} = y^2; \quad x \in [0, 1]; \quad y(0) = y_0.$$

Розв'язок цього рівняння має вигляд: $-\frac{1}{y} = x + C$, і, задовольняючи початкову умову, дістанемо

$$y(x) = -y_0/(1 - y_0 x).$$

Як бачимо, для будь-якого проміжку [0, 1] існує ряд значень y_0, при яких розв'язання задачі Коші перетворюється в нескінченність в середині цього проміжку, тобто розв'язку на цьому проміжку не існує.

Якщо розв'язок задачі Коші існує, то можна встановити, що він є єдиним, від супротивного.

При дослідженні чисельних методів для задачі Коші припускаємо, що її розв'язок існує, єдиний і має необхідні якості гладкості при всіх значеннях початкового вектора, яким ми будемо користуватись.

Перший крок на шляху чисельного розв'язання полягає в розбитті відрізку $[a, b]$ на скінченну кількість частин введенням вузлових точок $a = x_1 < x_2 < \ldots < x_n = b$. Хоч нерівномірне розбиття відрізку не призводить до особливих труднощів, для простоти викладу й аналізу будемо вважати, що вузлові точки ділять відрізок на рівні частини. Якщо позначити через h відстань між вузлами (крок сітки), то $h = (b - a)n$ і $x_k = a + kh$ $(k = 1, 2, \ldots, n)$, де n — число відрізків розбиття. Далі будемо через $y(x_k)$ позначати значення точного розв'язку (5.4) в точці x_k, а через y_k — відповідне наближене значення, побудоване за допомогою чисельного методу, що розглядається.

З урахуванням сказаного поставимо задачу Коші: в точках $x_0, x_1, \ldots, x_n$ потрібно знайти наближення y_n для значень точного розв'язку $y(x_n)$.

При чисельному розв'язанні задач Коші для звичайних диференціальних рівнянь методи розв'язання призначені для наближеного визначення функцій, що є розв'язком цих задач. «Функція» може характеризуватися тільки скінченною кількістю параметрів і деяким правилом, яке будемо називати *алгоритмом*, що дозволяє визначити якості шуканого розв'язку, що нас цікавить, з деяким ступенем точності, що вимагається. Задача пошуку найбільш зручних чисельних виразів і її зв'язок з методами покращення таких наближень відіграє важливу роль. Враховуючи особливості конкретної задачі, можна значно прискорити процес розв'язання, а також зменшити вимоги до об'єму оперативної пам'яті машини. У той же час такий підхід до розв'язування задач допомагає розв'язати питання про

те, чи можна взагалі виконати розрахунки. Ми будемо розглядати лише так звані *дискретні методи*, тобто методи, що визначають розв'язок для дискретних значень незалежної змінної. Характерна особливість дискретних методів для розв'язання рівняння

$$y' = f(x, y) \tag{5.11}$$

з початковими умовами $y|_{x=x_0} = y_0$ полягає в тому, що вони дозволяють знайти y_n за відомими (раніше визначеними) значеннями y_{n-k}, $k = 1, 2, \ldots, n$ і $f(x_k, y_k)$.

Вибір одного з багатьох існуючих наближених методів і загальний план обчислень залежать, очевидно, від величини кроку й потрібної точності. Якщо розв'язок потрібно знайти тільки на короткому проміжку і з помірною точністю, можна застосувати один з менш точних методів, що описані в § 5.2. Для знаходження розв'язку на проміжку більшої довжини (або на короткому проміжку, але з більшою точністю) обчислення необхідно проводити дуже точно і використовувати один із більш точних методів. При цьому точність проміжних розрахунків повинна бути значно вища, ніж потрібна точність кінцевого результату (додають, наприклад, два або три запасних знаки).

Більшість з наближених методів спираються на рівняння

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x))dx, \tag{5.12}$$

яке одержується інтегруванням рівняння (5.11); при цьому інтеграл у правій частині (5.12) замінюється тим чи іншим наближеним виразом.

Існують дві групи чисельних методів розв'язку задачі Коші: однокрокові методи типу Рунге—Кутта та багатокрокові різницеві методи.

§ 5.2. ЗАГАЛЬНІ ЗАУВАЖЕННЯ ЩОДО ОЦІНКИ ПОХИБКИ ПРИ РОЗВ'ЯЗАННІ ЗАДАЧ КОШІ

Після того як тим чи іншим способом в точках x_k знайдені наближені значення y_k для розв'язання задач Коші з початковими умовами дуже важливо оцінити величину похибки $\varepsilon_k = y_k - y(x_k)$ $(k = 1, 2, \ldots)$. Бажано мати можливість оцінити, у всякому разі, порядок похибок ε_k, щоб знати, на які десяткові знаки результату можна покладатись. При інтегруванні звичайних диференціальних рівнянь для практичної оцінки точності одержаних результатів рекомендується користуватись індуктивними прийомами, що дають приблизне уявлення про дійсну похибку. Розглянемо ці прийоми.

Порівняємо два наближення, одержані з різним кроком інтегрування. Знайдемо наближений розв'язок при кроці h і припустимо $x_p = x_0 + p\bar{h}$;

потім проведемо розрахунки з іншим кроком $\bar{h}$, узявши $\bar{h} = 2h$ і припустимо $\bar{x}_p = x_0 + p\bar{h}$. Порівнюючи відповідні наближені значення $y_2, y_4, \ldots, y_{20}$ зі значеннями $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p$, знайдемо різниці $\varepsilon_p = y_{2p} - \bar{y}_p$. Визначимо поняття порядку точності наближеного методу. Припустимо, що значення $y_1, y_2, \ldots, y_p$ збігаються з точними, тобто $y_k = y(x_k)$ при $k = 1, 2, \ldots, p$. Нехай y_{p+1} обчислено за допомогою даного наближеного методу. Розкладемо $y(x_{p+1})$ і y_{p+1} в ряди Тейлора з центром у точці x_p

$$y(x_{p+1}) = y(x_p) + \frac{h}{1!} y'(x_p) + \frac{h^2}{2!} y''(x_p) + \ldots;$$

$$y_{p+1} = y(x_p) + \frac{h}{1!} a_1 + \frac{h^2}{2!} a_2 + \ldots.$$

Існує останній член, в якому ще збігаються коефіцієнти при h^k. Показник степеня h у цьому члені і називається *порядком точності наближеного методу*. Таким чином, якщо порядок точності є m, то $a_1 = y'(x)$, $a_2 = y''(x), \ldots, a_m = y^{(m)}(x)$, але $a_{m+1} \neq y^{(m+1)}(x)$.

Для грубої оцінки похибки методу часто використовують принцип Рунге. Наводимо міркування з цього приводу. Для наближеного підрахунку похибки можна припустити, що на кожному кроці довжиною h зроблено похибку, яка пропорційна h^{k+1}. Таким чином, в точці $x = x_0 + 2nh$ припущено сумарну похибку $Ah^{k+1}2n$. Маємо

$$y(x) = y_{2n} + A2nh^{k+1}. \tag{5.13}$$

Припускається також, що при розрахунках з кроком $2h$ до цієї ж точки буде n разів зроблено похибку, пропорційну $(2h)^{k+1}$ з тим же коефіцієнтом пропорційності, а кінцева похибка дорівнює $An(2h)^{k+1}$, тоді

$$y(x) = y_n + An(2h)^{k+1}. \tag{5.14}$$

З (5.13) і (5.14) одержимо принцип Рунге у вигляді

$$\varepsilon = |y_{2n} - y(x_{2n})| \approx \frac{|y_{2n} - y_n|}{2^k - 1}. \tag{5.15}$$

У випадку систем рівнянь модуль заміняється нормою $\| y_{2n} - y_n \|$.

Одержані значення y_n і y_{2n} використовуються для порівняння досягнутої точності з кроком h. Якщо значення ε більше заданої точності $\bar{\varepsilon}$, то крок h зменшується вдвоє і процедура повторюється; якщо ε значно менше $\bar{\varepsilon}$ (вказується границя), то крок h збільшується вдвоє; якщо ε менше $\bar{\varepsilon}$, але незначно, то продовжуємо розрахунковий процес з тим же кроком h.

Таким чином, оцінювати точність можна за зменшенням різниці розв'язків в двох послідовних наближеннях.

Оцінку точності можна проводити й шляхом зіставлення значень приблизного розв'язку задачі з точним розв'язком, відомим для деяких особливих випадків.

§ 5.3. МЕТОД ЕЙЛЕРА І ЙОГО МОДИФІКАЦІЇ

Розглянемо три однокрокові чисельні методи, які застосовуються тільки тоді, коли не вимагається велика точність і кількість кроків невелика. Метод Ейлера полягає в тому, що при обчисленні по одному наближеному значенню y_k наступного значення y_{k+1}. Згідно з рівнянням (5.12), інтеграл у правій частині заміняється простіше — через похідні початкового значення підінтегральної функції на крок h

$$y_{k+1} = y_k + hf(x_k, y_k), \quad y_0 = \bar{y}, \; k = 0, 1, \ldots, n-1. \tag{5.16}$$

Зображення методу Ейлера є очевидним. З розвинення Тейлора функції y в околі точки x_k маємо

$$y(x_{k+1}) = y(x_k) + hy'(x_k) + \frac{h^2}{2} y''(z_k) =$$

$$= y(x_k) + hf(x_k, y(x_k)) + \frac{h^2}{2} y''(z_k),$$

де z_k лежить в середині відрізка $[x_k, x_{k+1}]$. Якщо похідна y'' обмежена, а крок h малий, останній член можна відкинути і наближено написати

$$y(x_{k+1}) \approx y(x_k) + hf(x_k, y_k).$$

Геометричний зміст методу Ейлера полягає в апроксимації розв'язку на відрізку $[x_k, x_{k+1}]$ відрізком дотичної L_1, проведеної до графіку розв'язку в точці x_k (рис.4).

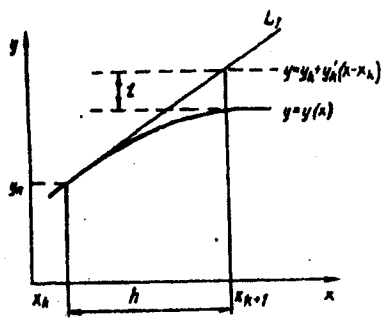


Рис. 4

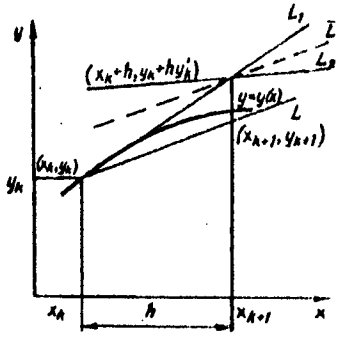


Рис. 5

Помилка при $x = x_{k+1}$ показана у вигляді відрізку l. Метод Ейлера дуже простий для реалізації на ЕОМ: на кроці k обчислюється значення $f(x_k, y_k)$, яке потім підставляється в (5.16). Таким чином, всі необхідні операції зводяться до обчислення $f(x_k, y_k)$. Очевидно, знайдений таким

чином наближений розв'язок має перший порядок точності, оскільки він узгоджується з розвиненням в ряд Тейлора до членів порядку h включно.

Крім того, цей метод дуже часто виявляється нестійким: мала помилка, що є наслідком обмеження наближення, або ж закладена у вхідних даних, збільшується зі зростанням x.

Для обчислення значення y_{k+1} метод Ейлера використовує нахил дотичної тільки в точці x_k, y_k. Цей метод можна поліпшити багатьма різноманітними способами. З цих способів ми тут розглянемо два: метод Ейлера—Коші і модифікований метод Ейлера, що мають другий порядок точності.

У методі Ейлера—Коші ми знаходимо середній тангенс кута нахилу дотичної для двох точок: x_k, y_k і $x_k + h$, $y_k + hy_k'$. Остання точка є тією, яка в методі Ейлера позначалась x_{k+1}, y_{k+1}. Геометрично цей процес знаходження точки x_{k+1}, y_{k+1} можна простежити за рис. 5. За допомогою методу Ейлера знаходиться також точка $x_k + h$, $y_k + hy_k'$, що лежить на прямій L_1. У цій точці знову визначається тангенс кута нахилу дотичної. На рисунку цьому значенню відповідає пряма L_2. Знайдене середнє значення двох тангенсів дає пряму $\bar{L}$. Нарешті, через точку x_k, y_k проводимо пряму L, паралельну $\bar{L}$. Точка, в якій пряма L перетнеться з ординатою, одержаною з $x = x_{k+1} = x_k + h$, і буде шуканою точкою x_{k+1}, y_{k+1}.

Тангенс кута нахилу прямої $\bar{L}$ і прямої L дорівнює

$$\Phi(x_k, y_k, h) = 0{,}5\,[f(x_k, y_k) + f(x_k + h, y_k + hy_k')\,], \qquad (5.17)$$

де

$$y_k' = f(x_k, y_k). \qquad (5.18)$$

Рівняння лінії L при цьому записується у вигляді

$$y = y_k + (x - x_k)\Phi(x_k, y_k, h),$$

тому

$$y_{k+1} = y_k + h\Phi(x_k, y_k, h). \qquad \text{ü}(5.19)$$

Співвідношення (5.17), (5.18) і (5.19) описують метод Ейлера—Коші.

Щоб визначити, наскільки добре цей метод узгоджується з розвиненням в ряд Тейлора, згадаємо, що розвинення в ряд функції $f(x, y)$ можна записати таким чином:

$$f(x, y) = f(x_k, y_k) + (x - x_k)\frac{\partial f}{\partial x} + (y - y_k)\frac{\partial f}{\partial y} + \ldots, \qquad (5.20)$$

де часткові похідні розраховуються при $x = x_k$ і $y = y_k$. Підставляючи в формулу (5.20) $x = x_k + h$, $y = y_k + hy_k'$ і використовуючи вираз (5.18) для y_k', дістанемо

$$f(x_k + h, y_k + hy_k') = f + hf_x' + hf_y' + O(h^2),$$

де знову функція f і її похідні визначаються в точці x_k, y_k. Підставляючи результат в (5.17) і виконуючи необхідні перетворення, матимемо

$$\Phi(x_k, y_k, h) = f + 0{,}5h(f_x' + ff_y') + O(h^2).$$

Підставивши останнє співвідношення в (5,19), прийдемо до формули, яку можна безпосередньо порівняти з розвиненням в ряд Тейлора

$$y_{k+1} = y_k + hf + 0{,}5h^2(f_x' + ff_y') + O(h^3).$$

Як бачимо, метод Ейлера—Коші узгоджується з розвиненням в ряд Тейлора до членів степеня h^2 включно, і, отже, є методом другого порядку.

У виправленому методі Ейлера—Коші усереднювались нахили дотичних. Можна піти іншим шляхом і усереднити точки іншим чином. Розглянемо рис. 6, де початкова побудова виконана, як і для попереднього методу: через точку x_k, y_k проведено пряму L_1 з тангенсом кута нахилу, що дорівнює $f(x_k, y_k)$. Але цього разу ми беремо точку, що лежить на перетині прямої і ординати, одержаної з точки $x = x_k + h/2$. На рисунку її позначено через P, а ордината дорівнює $y = y_k + (h/2)y_k'$. Обчислимо тангенс кута нахилу дотичної в цій точці

$$\Phi(x_k, y_k, h) = f(x_k + h/2,\ y_k + h/2y_k'), \tag{5.21}$$

де

$$y_k' = f(x_k, y_k). \tag{5.22}$$

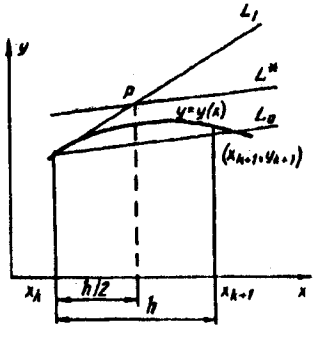


Рис. 6

Пряма з таким нахилом, що проходить через P, позначена через L^*. Проведемо через точку x_k, y_k пряму, паралельну L^* і позначимо її через L_0. Перетин цієї прямої з ординатою $x = x_k + h$ і дає шукану точку x_{k+1}, y_{k+1}. Рівняння прямої L_0 можна записати у вигляді

$$y = y_k + (x - x_k)\Phi(x_k, y_k, h),$$

де Φ задається формулою (5.21). Тому

$$y_{k+1} = y_k + h\Phi(x_k, y_k, h). \tag{5.23}$$

Співвідношення (5.21), (5.22) і (5.23) описують модифікований метод Ейлера. За аналогією з попереднім методом можна показати, що останній також погоджується з розвиненням в ряд Тейлора до членів степеня h^2 включно, тобто є методом другого порядку.

Оцінка похибки наближеного розв'язку y_n з кроком h в точці x_n може бути одержана за допомогою подвійного розрахунку: розрахунок повторюють з кроком $h/2$ і більш точне значення позначають y_n^*. Згідно з (5.15)

похибка приблизного розв'язку одержаного з кроком $h/2$, в $(2^k - 1)$ разів менше двох наближених значень y_n^* і y_n

$$|y_n^* - y(x_n)| \le \frac{1}{2^k - 1} |y_n^* - y_n|, \qquad (5.24)$$

де $y(x_n)$ — точний розв'язок; k — порядок методу.

Додержуючись (5.24), похибку наближених значень за методом Ейлера можна оцінити таким чином:

$$|y_n^* - y(x_n)| \le |y_n^* - y_n|.$$

За методом Ейлера—Коші і модифікованим методом Ейлера маємо

$$|y_n^* - y(x_n)| \le \frac{1}{3} |y_n^* - y_n|.$$

Приклад. Застосовуючи метод Ейлера і його модифікації скласти на відрізку [0, 1] таблицю значень розв'язку рівняння

$$y' = y - 2x/y \qquad (5.25)$$

з початковою умовою $y(0) = 1$, вибравши крок $h = 0{,}2$. Для цієї задачі відомий точний розв'язок $y = \sqrt{2x + 1}$.

Розв'язання. Застосовуючи описані алгоритми розв'язку задачі Коші, обчислюємо дискретні значення функції $y(x)$ і заносимо результати в табл. 7.

Таблиця 7

x	Метод Ейлера	Метод Ейлера—Коші	Модифікований метод Ейлера	Точний розв'язок
0,2	1,2000	1,1867	1,1836	1,1832
0,4	1,3733	1,3483	1,3426	1,3416
0,6	1,5294	1,4937	1,4850	1,4832
0,8	1,6786	1,6279	1,6152	1,6124
1,0	1,8237	1,7543	1,7362	1,7320

Як видно з таблиці, числові розв'язки помітно відрізняються від точного. При використанні наближених методів головним є оцінка точності наближених значень y_k. Взагалі, існує два джерела похибки цих наближень:

1) похибка дискретизації, що виникає в результаті заміни диференціального рівняння (5.11) різницевою апроксимацією вигляду (5.16), (5.19) або (5.23);

2) похибка округлення, що накопичилась при виконанні арифметичних операцій за відповідними розрахунковими алгоритмами. Як уже було

показано, помилка дискретизації для методів Ейлера, Ейлера—Коші і модифікованого методу Ейлера становить h і h^2 відповідно. Тобто похибка дискретизації прагне до нуля при прагненні до нуля h. Отже, за рахунок зменшення кроку h помилку дискретизації можна зробити скільки завгодно малою. Проте чим менше h, тим більше буде вимагатися кроків за наближеним методом і, загалом, тим більше відіб'ються на одержаному розв'язку похибки округлення. На практиці, коли при виконанні арифметичних операцій використовуються слова фіксованої довжини, завжди існує така величина кроку h, менше якої похибки округлення починають домінувати в сумарній похибці. Ця ситуація схематично зображена на рис.7, де h_0 визначає мінімальний крок, який можна використовувати при розрахунках.

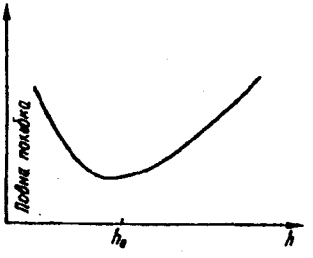


Рис. 7

Цю мінімальну величину дуже важко встановити спочатку, але в задачах, де дуже висока точність не потрібна, необхідний крок звичайно буде значно більшим від цього мінімуму, і основний внесок в повну похибку буде мати похибка дискретизації. Така поведінка характерна й для інших методів, хоча мінімальний розмір кроку буде мінятись від методу до методу і від задачі до задачі.

§ 5.4. МЕТОД РУНГЕ—КУТТА

Метод Рунге—Кутта найбільше використовується серед однокрокових методів підвищеної точності наближеного розв'язку задачі Коші для диференціального рівняння

$$\frac{dy}{dx} = f(x, y), \tag{5.26}$$

з початковими умовами

$$y\,|_{x=x_0} = y_0. \tag{5.27}$$

Ідея цього методу має багато спільного з ідеєю вищевикладених методів Ейлера і його модифікацій і полягає в підгонці ряду Тейлора.

Визначимо шуканий розв'язок $y(x)$ задачі Коші (5.26), (5.27) в околі кожної точки $x = x_n$ ($n = 0, 1, 2, \ldots$) за формулою Тейлора, обчислимо коефіцієнти розвинення безпосередньо за правою частиною рівняння (5.26), використовуючи умови (5.27). Вказане розвинення запишеться у вигляді

$$y(x) = y_n + h\frac{dy}{dx} + \frac{h^2}{2!}\frac{d^2y}{dx^2} + \frac{h^3}{3!}\frac{d^3y}{dx^3} + \dots, \qquad (5.28)$$

де значення похідних узяті при $x = x_k$. В залежності від того, скількома членами розвинення ми задовольнимося в формулі (5.28), дістанемо ту чи іншу точність наближеного розв'язку. У методі Рунге—Кутта обмежимося чотирма або п'ятьма членами розвинень (утримуються члени з степенями до h^3 або h^4 включно).

Розглянемо метод Рунге—Кутта третього порядку точності:

$$y(x) \approx y_n + h\frac{dy}{dx} + \frac{h^2}{2!}\frac{d^2y}{dx^2} + \frac{h^3}{3!}\frac{d^3y}{dx^3}. \qquad (5.29)$$

Припускаємо, що

$$y_{n+1} = y_n + \lambda_n, \qquad (5.30)$$

де

$$\lambda_n = h\frac{dy}{dx} + \frac{h^2}{2}\frac{d^2y}{dx^2} + \frac{h^3}{6}\frac{d^3y}{dx^3}. \qquad (5.31)$$

Величини λ_n визначаємо за допомогою лінійних комбінацій вигляду

$$\lambda_n = \alpha k_1 + \beta k_2 + \gamma k_3 + \delta k_4, \qquad (5.32)$$

де $\alpha, \beta, \gamma, \delta$ — невизначені коефіцієнти, а k_1, k_2, k_3, k_4 — числа, що визначаються рівняннями

$$k_1 = hf(x_n, y_n);$$

$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right);$$

$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right);$$

$$k_4 = hf(x_n + h, y_n + k_2). \qquad (5.33)$$

Ці числа мають простий геометричний зміст.

Нехай крива MCM_1 (рис. 8) — розв'язок задачі Коші (5.26), (5.27), C — точка цієї кривої, що лежить на прямій, паралельній осі y, і ділить відрізок $[x_n, x_{n+1}]$ навпіл; B і G — точки перетину дотичної до MCM_1 у точці M з ординатами AB і N_1G_1. Тоді число k_1 з точністю до множника h ($h = x_{n+1} - x_n$) буде кутовим коефіцієнтом дотичної в точці M інтегральної кривої MCM_1 $k_1 = hy_M' = hf(x_m, y_m)$. Число k_2 з точністю до множника h буде кутовим коефіцієнтом дотичної до інтегральної кривої в точці B (BF — відрізок дотичної). Проведемо через точку M пряму, паралельну відрізку BF, і позначимо точки перетину D і E з ординатами AB і N_1G_1. Тоді

з точністю до множника h числа k_3 і k_4 будуть кутовими коефіцієнтами дотичних до інтегральних кривих відповідно в точках D і E.

Для визначення коефіцієнтів $\alpha, \beta, \gamma, \delta$ виразимо похідні, що входять в рівняння (5.31), через праву частину рівняння (5,26)

$$\frac{d^2y}{dx^2} = \frac{\partial f}{\partial x} + y' \frac{\partial f}{\partial y} = \frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y}.$$

Для простоти викладення надалі введемо оператор D

$$D = \frac{\partial}{\partial x} + f \frac{\partial}{\partial y},$$

тоді

$$\frac{d^2y}{dx^2} = Df;$$

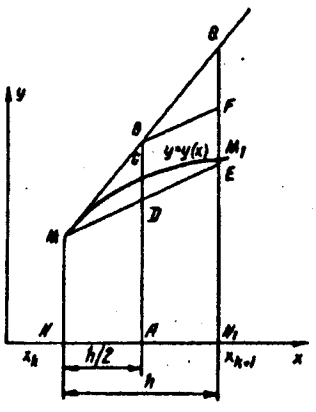


Рис. 8

$$\frac{d^3y}{dx^3} = D(Df) = \frac{\partial^2 f}{\partial x^2} + 2f \frac{\partial^2 f}{\partial x \partial y} + f^2 \frac{\partial^2 f}{\partial y^2} + \frac{\partial f}{\partial y}\left(\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y}\right) =$$

$$= D^2 f + \frac{\partial f}{\partial y} Df.$$

Підставляючи знайдені значення похідних у рівняння (5.31), знаходимо

$$\lambda_n = hf + \frac{h^2}{2} Df + \frac{h^3}{6}\left(D^2 f + \frac{\partial f}{\partial y} Df\right). \qquad (5.34)$$

Виразимо k_2, k_3, k_4 як функції двох змінних за формулою Тейлора. Маємо:

$$k_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)h = \left[f + \left(\frac{h}{2}\frac{\partial}{\partial x} + \frac{k_1}{2}\frac{\partial}{\partial y}\right)f + \right.$$

$$\left. + \frac{1}{2}\left(\frac{h}{2}\frac{\partial}{\partial x} + \frac{k_1}{2}\frac{\partial}{\partial y}\right)^2 f + \ldots\right]h.$$

Обмежимося третіми степенями h і, враховуючи, що $k_1 = hf$, дістанемо:

$$k_2 = \left[f + \frac{h}{2}\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y}\right)f + \frac{h^2}{8}\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y}\right)^2 f\right] =$$

$$= hf + \frac{h^2}{2} Df + \frac{h^3}{8} D^2 f.$$

Аналогічно для k_3 і k_4

$$k_3 = \left[f + \left(\frac{h}{2}\frac{\partial}{\partial x} + \frac{k_2}{2}\frac{\partial}{\partial y}\right)f + \frac{1}{2}\left(\frac{h}{2}\frac{\partial}{\partial x} + \frac{k_2}{2}\frac{\partial}{\partial y}\right)^2 f\right]h =$$

$$= \left[f + \frac{h}{2}\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y} + \frac{h}{2}Df\frac{\partial}{\partial y}\right)f + \frac{h^2}{8}\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y} + \ldots\right)^2 f\right]h =$$

$$= hf + \frac{h^2}{2} Df + \frac{h^2}{4}\frac{\partial f}{\partial y} Df + \frac{h^3}{8} D^2 f;$$

$$k_4 = \left[f + \left(h\frac{\partial}{\partial x} + k_2\frac{\partial}{\partial y}\right)f + \frac{1}{2}\left(h\frac{\partial}{\partial x} + k_2\frac{\partial}{\partial y}\right)^2 f\right]h =$$

$$= \left[f + h\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y} + \frac{h}{2}Df\frac{\partial}{\partial y}\right)f + \frac{h^2}{2}\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y} + \ldots\right)^2 f\right]h =$$

$$= hf + h^2 Df + \frac{h^3}{2}\frac{\partial f}{\partial y} Df + \frac{h^3}{2} D^2 f.$$

Знайдемо тепер суму $\lambda_n = \alpha k_1 + \beta k_2 + \gamma k_3 + \delta k_4$.

Порівнюючи коефіцієнти при однакових степенях h в останньому рівнянні і виразі (5.34) для визначення $\alpha, \beta, \gamma, \delta$, дістанемо систему рівнянь

$$\alpha + \beta + \gamma + \delta = 1 \quad (\text{при } hf);$$

$$\frac{\beta}{2} + \frac{\gamma}{2} + \delta = \frac{1}{2} \quad (\text{при } h^2 Df);$$

$$\frac{\beta}{8} + \frac{\gamma}{8} + \frac{\delta}{2} = \frac{1}{6} \quad (\text{при } h^3 D^2 f);$$

$$\frac{\gamma}{4} + \frac{\delta}{2} = \frac{1}{6} \quad \left(\text{при } h^3 \frac{\partial f}{\partial y} Df\right).$$

Ця система має розв'язок

$$\alpha = \delta = \frac{1}{6}, \quad \beta = \gamma = \frac{1}{3}.$$

Таким чином,

$$\lambda_n = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \tag{5.35}$$

Для обчислення в точці x_{n+1} маємо формулу

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \tag{5.36}$$

де k_1, k_2, k_3, k_4 — числа, що визначаються рівняннями (5.33).

У випадку метода Рунге—Кутта четвертого порядку точності, тобто коли в розвиненні (5.28) утримуються члени з степенями від h до h^4 включно, процес інтегрування слід проводити таким же чином, тільки змінюються числа k_1, k_2, k_3, k_4

$$k_1 = hf(x_n, y_n);$$

$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right);$$

$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right);$$

$$k_4 = hf(x_n + h, y_n + k_3). \tag{5.37}$$

Аналогічно можна побудувати формули вищих степенів.

Заради спрощення ми розглядали методи чисельного розв'язання диференціальних рівнянь першого порядку. Узагальнення на систему рівнянь першого порядку не дає нічого нового. Достатньо інтегрувати $\bar{y}$ у відповідних формулах як вектор, а відповідні функції — як вектор-функції. Наприклад, для системи

$$y_1' = f_1(x_1, y_1, y_2, \ldots, y_n);$$

$$y_2' = f_2(x_1, y_1, y_2, \ldots, y_n);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$y_n' = f_n(x_1, y_1, y_2, \ldots, y_n)$$

стандартні формули Рунге—Кутта четвертого степеня мають вигляд

$$y_j(x_{i+1}) = y_j(x_i) + \frac{1}{6}[k_{1j}^{(i)} + 2k_{2j}^{(i)} + 2k_{3j}^{(i)} + k_{4j}^{(i)}] + O(h^5),$$

$$j = 1, 2, \ldots, n;\ i = 1, 2, \ldots, n, \tag{5.38}$$

де

$$k_{1j}^{(i)} = hf_j(x_i, y_1^{(i)}, \ldots, y_n^{(i)});$$

$$k_{2j}^{(i)} = hf_j\left(x_i + \frac{h}{2}, y_1^{(i)} + \frac{k_{1,1}^{(i)}}{2}, y_2^{(i)} + \frac{k_{1,2}^{(i)}}{2}, \ldots, y_n^{(i)} + \frac{k_{1,n}^{(n)}}{2}\right);$$

$$k_{3j}^{(i)} = hf_j\left(x_i + \frac{h}{2},\ y_1^{(i)} + \frac{k_{2,1}^{(i)}}{2},\ y_2^{(i)} + \frac{k_{2,2}^{(i)}}{2},\ \ldots,\ y_n^{(i)} + \frac{k_{2,n}^{(n)}}{2}\right);$$

$$k_{4j}^{(i)} = hf_j(x_i + h,\ y_1^{(i)} + k_{3,1}^{(i)},\ y_2^{(i)} + k_{3,2}^{(i)},\ \ldots,\ y_n^{(i)} + k_{3n}^{(i)});$$

$$j = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, n.$$

Для рівняння (5.26) за допомогою розвинення за формулою Тейлора була одержана оцінка похибки методу Рунге—Кутта четвертого порядку точності

$$|y_1 - y(x_1)| \leq \frac{6MN\,|x_1 - x_0|^5\,|N^5 - 1|}{|N - 1|},$$

де M і N — константи такі, що в околі $|x - x_0| < a$, $|y - y_0| < b$ повністю виконуються нерівності

$$|f(x, y)| < M;$$

$$\left|\frac{\partial^{i+k} f}{\partial x^i \partial y^k}\right| < \frac{N}{M^{k-1}} \quad (i + k \leq 3);$$

$$|x - x_0|N < 1, \quad aM < b, \quad h \leq a.$$

У загальному випадку вказати точні оцінки похибки для методу Рунге—Кутта досить важко. Тому часто доводиться користуватись приблизною оцінкою похибки цього методу, виходячи з різних побічних роздумів. Намагаючись підвищити точність розв'язку за рахунок зменшення кроку, на практиці користуються таким прийомом. Знаходять різниці $|k_2 - k_3|$ і $|k_1 - k_2|$ і вимагають, щоб перша з них не перевищувала кількох відсотків останньої.

Якщо цю умову не виконано, то крок розбиття зменшується. Відповідне число $\dfrac{|k_2 - k_3|}{|k_1 - k_2|}$ є свого роду мірою чутливості або показником кроку. При реалізації методу Рунге—Кутта на ЕОМ з автоматичним вибором кроку звичайно в кожній точці x_i обчислення виконують двічі: спочатку з кроком h, а потім з кроком $h/2$. Якщо одержані при цьому значення y_i розрізняються в межах припустимої точності, то крок h для наступної точки x_{i+1} збільшують вдвічі; в іншому випадку беруть половинний крок. Грубу оцінку похибки можна отримати у відповідності з виразом (5.15)

$$|y_n^* - y(x_n)| \approx \frac{|y_n^* - y_n|}{15},$$

де $y(x_n)$ — значення точного розв'язку рівняння в точці x_n, а y_n^* і y_n приблизні значення, одержані з кроком $h/2$ і h.

Зупинимося на деяких особливостях методів Рунге—Кутта і модифікаціях методів Ейлера, які можна розглядати як методи Рунге—Кутта першого й другого порядків точності.

1. Вони одноступінчасті: щоб знайти приблизне значення y_{m+1}, потрібна інформація про попередню точку x_m, y_m.

2. Узгоджуються з рядом Тейлора до членів порядку h^k включно; де степінь k різна для різних порядків і називається *порядком методу.*

3. Не потребують обчислення похідних від $f(x, y)$, а тільки обчислення самої функції. Завдяки цій особливості методи Рунге—Кутта більш зручні для практичних обчислень, порівняно з рядом Тейлора. Проте для обчислення однієї наступної точки розв'язку нам доведеться обчислювати $f(x, y)$ кілька разів при різних x і y, і складніша права частина пов'язана зі значною обчислювальною роботою.

4. Схема Рунге—Кутта дає хорошу точність, якщо права частина диференціального рівняння обмежена й неперервна разом зі своїми похідними до четвертого порядку. Якщо ж права частина не має вказаних похідних, то граничний порядок точності цієї схеми може бути не реалізованим. Тоді не гірші, ніж за схемою Рунге—Кутта, і, напевне, не кращі результати дають схеми меншого порядку точності, що дорівнюють порядку відомих похідних.

Приклад. У табл. 8 для задачі (5.25) порівнюються результати, одержані за методом Рунге—Кутта четвертого порядку точності з кроком h і зменшеним вдвоє, з точним розв'язком $y = \sqrt{2x + 1}$.

Таблиця 8

x	Точний розв'язок	Метод Рунге—Кутта	
		$h = 0,4$	$h = 0,2$
0,2	1,1832160		1,1832292
0,4	1,3416407	1,342066	1,3416668
0,6	1,4832397		1,4832847
0,8	1,6124516	1,613449	1,6125186
1,0	1, 7320508		1,7321483
1,2	1,8439089	1,846000	1,8440490

§ 5.5. БАГАТОКРОКОВІ МЕТОДИ

Серед відомих приблизних методів найбільш точними є різницеві методи, які відносяться до класу багатокрокових. Перейдемо до різницевих методів розв'язання звичайних диференціальних рівнянь, застосування яких потребує тільки одноразового обчислення правої частини на кожно-

му кроці. Обмежимося випадком одного рівняння першого порядку. Нехай потрібно знайти розв'язок рівняння

$$y' = f(x, y), \quad a \le x \le b, \tag{5.39}$$

що задовольняє початкову умову $y'(a) = y_0$. У попередніх параграфах цієї глави значення y_{k+1} залежало тільки від інформації в попередній точці x_k. Здається досить імовірним, що можна досягти більшої точності, якщо використовувати інформацію про деякі попередні точки x_k, x_{k-1}, ... Саме так і роблять при використанні багатокрокових методів.

Розглянемо один з підходів багатокрокових методів. Якщо підставити в (5.39) точний розв'язок $y(x)$ і проінтегрувати це рівняння на відрізку $[x_k, x_{k+1}]$, то дістанемо

$$y(x_{k+1}) - y(x_k) = \int_{x_k}^{x_{k+1}} y'(x)dx = \int_{x_k}^{x_{k+1}} f(x, y(x))dx \approx \int_{x_k}^{x_{k+1}} p(x)dx, \tag{5.40}$$

де в останньому члені припускаємо, що $p(x)$ — поліном, що апроксимує $f(x, y(x))$. Для побудови цього полінома, припустимо, як завжди, що y_k, y_{k-1}, ... , y_{k-N} — наближення до розв'язку в точках x_k, x_{k-1}, ... , x_{k-N}. Як і раніше, вважаємо, що вузли x_i розташовані рівномірно з кроком h. Тоді $f_i = f(x_i, y_i)$ $(i = k, k-1, \dots, k-N)$ є наближеннями до $f(x, y(x))$ у точках x_k, x_{k-1}, ... , x_{k-N}, і замість p візьмемо інтерполяційний поліном для набору даних (x_i, f_i) $(i = k, k-1, \dots, k-N)$. Таким чином, p — поліном степеня N, що задовольняє умови $p(x_i) = f_i$ $(i = k, k-1, \dots, k-N)$. Взагалі, можна проінтегрувати цей поліном, що приводить до методу

$$y_{k+1} = y_k + \int_{x_k}^{x_{k+1}} p(x)dx. \tag{5.41}$$

У найпростішому випадку, коли $N = 0$, поліном p є константою, що дорівнює f_k, і (5.41) перетворюється на звичайний метод Ейлера. Якщо $N = 1$, то p — лінійна функція, яка проходить через точки (x_{k-1}, f_{k-1}) і (x_k, f_k), тобто

$$p(x) = -\frac{(x - x_k)}{h} f_{k-1} + \frac{(x - x_{k-1})}{h} f_k.$$

Інтегруючи цей поліном від x_k до x_{k+1}, дістанемо

$$y_{k+1} = y_k + \frac{h}{2}(3f_k - f_{k-1}), \tag{5.42}$$

який є двокроковим, оскільки використовує інформацію в двох точках x_k і x_{k-1}. Аналогічно, якщо $N = 2$, то p є квадратичним поліномом, що інтерполює дані (x_{k-2}, f_{k-2}), (x_{k-1}, f_{k-1}) і (x_k, f_k), а відповідний метод має вигляд

$$y_{k+1} = y_k + \frac{h}{12}(23f_k - 16f_{k-1} + 5f_{k-2}). \qquad (5.43)$$

Якщо $N = 3$, то інтерполяційний поліном є кубічним, а відповідний метод визначається за формулою

$$y_{k+1} = y_k + \frac{h}{24}(55f_k - 59f_{k-1} + 37f_{k-2} - 9f_{k-3}). \qquad (5.44)$$

Метод (5.43) трикроковий, а (5.44) — чотирикроковий. Формули (5.42) — (5.44) відомі як методи Адамса—Башфорта і є відповідно методами другого, третього й четвертого порядків.

Цей процес можна продовжити і, використовуючи все більшу кількість попередніх точок, а звідси, й інтерполяційний поліном p вищого степеня, отримати методи Адамса—Башфорта скільки завгодно високого порядку. При цьому із зростанням N формули стають більшими й незграбними, але принцип залишається тим самим.

Багатокрокові методи породжують проблему, яка не виникала при використанні однокрокових методів. Ця проблема стає зрозумілою, якщо, наприклад, розглянути метод Адамса—Башфорта четвертого порядку (5.44). Дано початкове значення y_0, або при $k = 0$ для обчислення за формулою (5.44) необхідна інформація в точках x_{-1}, x_{-2}, x_{-3}, яка, зрозуміло, відсутня. Складність полягає в тому, що багатокрокові методи спочатку потребують допомоги. Використовувати (5.44) при $k < 3$ або (5.43) при $k < 2$ ми не можемо. Виходом з цього становища є використання одного з однокрокових методів того ж порядку точності (наприклад, Рунге—Кутта) доти, поки не буде одержано достатньо значень для роботи багатокрокового методу. Крім того, на першому кроці можна використовувати однокроковий метод, на другому — двокроковий і так далі, поки не буде одержано достатньо стартових значень. При цьому, проте, суттєво, щоб ці стартові значення були обчислені з тим же ступенем точності, з якою буде працювати кінцевий метод. Оскільки стартові методи звичайно мають більш низький порядок, на початку обчислення потрібно вести з меншим кроком і використовувати більше проміжних точок.

Методи Адамса—Башфорта використовують уже обчислені значення в точці x_k і в попередніх точках. Взагалі, при побудові інтерполяційного полінома ми можемо використовувати і точки x_{k+1}, x_{k+2} і т. д. Найпростіший випадок полягає у використанні точок $x_{k+1}, x_k, \ldots, x_{k-N}$ і побудові інтерполяційного полінома степеня $N + 1$, що задовольняє умови $p(x_i) = f_i$ $(i = k+1, k, \ldots, k-N)$. При цьому виникає клас методів, відомих як методи Адамса—Моултона. Якщо $N = 1$, то p — лінійна функція, що проходить через точки (x_k, f_k) і (x_{k+1}, f_{k+1}); і відповідний метод

$$y_{k+1} = y_k + \frac{h}{2}(f_{k+1} + f_k) \qquad (5.45)$$

є методом Адамса—Моултона другого порядку. Якщо $N = 3$, то p є кубічним поліномом, побудованим за точками (x_{k+1}, f_{k+1}), (x_k, f_k), (x_{k-1}, f_{k-1}), (x_{k-2}, f_{k-2}), і відповідний метод

$$y_{k+1} = y_k + \frac{h}{24}(9f_{k+1} + 19f_k - 5f_{k-1} + f_{k-2}) \tag{5.46}$$

є методом Адамса—Моултона четвертого порядку.

Зазначимо, що в формулах (5.45) і (5.46) значення f_{k+1} є невідомим. Уся справа в тому, що для обчислення $f(x_{k+1}, y_{k+1}) = f_{k+1}$ потрібно знати значення y_{k+1}, яке поки що невідоме. Отже, методи Адамса—Моултона визначають y_{k+1} тільки не явно. Так, співвідношення (5.45) дійсно є рівнянням

$$y_{k+1} = y_k + \frac{h}{2}[f(x_{k+1}, y_{k+1}) + f_k], \tag{5.47}$$

відносно невідомого значення y_{k+1}. Те ж саме справедливо й відносно (5.46). Через це методи Адамса—Моултона називаються *неявними*. У той же час методи Адамса—Башфорта називаються *явними*, оскільки вони для знаходження значення y_{k+1} не потребують розв'язку ніяких рівнянь.

На практиці звичайно не розв'язують безпосередньо рівняння (5.47), а використовують разом явну і неявну формули, що призводить до методу *прогнозу* і *корекції*. Одним з методів прогнозу і корекції, який широко використовується, є об'єднання методів Адамса четвертого порядку (5.44) і (5.46):

$$y_{k+1}^{(p)} = y_k + \frac{h}{24}(55f_k - 59f_{k-1} + 37f_{k-2} - 9f_{k-3});$$

$$f_{k+1}^{(p)} = f(x_{k+1}, y_{k+1}^{(p)});$$

$$y_{k+1} = y_k + \frac{h}{24}(9f_{k+1}^{(p)} + 19f_k - 5f_{k-1} + f_{k-2}). \tag{5.48}$$

У цілому цей метод є явним. Спочатку за формулою Адамса—Башфорта обчислюється значення $y_{k+1}^{(p)}$, що є «прогнозом» для y_{k+1}. Потім $y_{k+1}^{(p)}$ використовується для наближеного значення f_{k+1}, яке в свою чергу використовується в формулі Адамса—Моултона. Таким чином, формула Адамса—Моултона «коригує» наближення, що надається формулою Адамса—Башфорта.

Розглянемо більш загальний випадок вибору інтерполяційного полінома $p(x)$, з якого наведені вище формули Адамса—Башфорта і Адамса—Моултона випливають як частинні. Виберемо в (5,40) в значенні полінома $p(x)$ поліном Ньютона інтерполяції назад, який в точках $x_{k-N}, x_{k-N+1}, \ldots, x_k$ набуває значень $f_{k-N}, f_{k-N+1}, \ldots, f_k$. Цей поліном має вигляд

$$p(x) = f_k + t\nabla f_k + \frac{t(t+1)}{2!}\nabla^2 f_k + \ldots + \frac{1}{N!}(t(t+1)\ldots(t+N-1)\nabla^N f_k$$

$$\left(\nabla f_k = f_k - f_{k-1},\ t = \frac{x - x_k}{h}\right), \tag{5.49}$$

а відповідний йому залишковий член R задовольняє нерівність

$$|R| \le \frac{t(t+1)\dots(t+N)}{(N+1)!} h^{N+1} M_{N+1},\ M_{N+1} = \max_{x_{k-N} \le x' \le x_k} |f^{(N+1)}(x')|.$$

Підставляючи (5.48) в (5.40) і перейшовши до змінної інтегрування t, дістанемо

$$y_{k+1} = y_k + hf_k + h\nabla f_k \int_0^1 t dt + \dots + \frac{1}{N!} h\nabla^N f_k \int_0^1 t(t+1)\dots(t+p-1)dt, \tag{5.50}$$

або в компактній формі

$$y_{k+1} = y_k + h \sum_{n=0}^{N} \alpha_n \nabla^n f_k \quad (k = N, N+1, N+2, \dots), \tag{5.51}$$

де $\alpha_0 = 1$, $\alpha_n = \frac{1}{n!} \int_0^1 t(t+1)\dots(t+n-1)dt$ $(n = 1, 2, \dots, N)$.

Обчислення наближених значень шуканої функції за формулою (5.51) є екстраполяційним методом Адамса.

В окремому випадку при $N = 3$ дістанемо найбільш уживану формулу екстраполяційного методу Адамса

$$y_{k+1} = y_k + h\left(f_k + \frac{1}{2}\nabla f_k + \frac{5}{12}\nabla^2 f_k + \frac{3}{8}\nabla^3 f_k\right). \tag{5.52}$$

При розрахунках значення шуканої функції y_{k+1}, а також f_k і різниці ∇f_k, $\nabla^2 f_k$, $\nabla^3 f_k$ зручно записувати у вигляді табл. 9.

Таблиця 9

x_k	y_k	hf_k	∇f_k	$\nabla^2 f_k$	$\nabla^3 f_k$
x_0	y_0	f_0			
$x_0 + h$	y_1	f_1	∇f_1		
$x_0 + 2h$	y_2	f_2	∇f_2	$\nabla^2 f_2$	
$x_0 + 3h$	y_3	f_3	∇f_3	$\nabla^2 f_3$	$\nabla^3 f_3$

Таблиця 10

N	n			
	0	1	2	3
1	$\frac{3}{2}$	$-\frac{1}{2}$		
2	$\frac{23}{12}$	$-\frac{16}{12}$	$\frac{5}{12}$	
3	$\frac{55}{24}$	$-\frac{59}{24}$	$\frac{37}{24}$	$-\frac{9}{24}$

Інколи рівняння (5.51) приводять в трохи інший формі, виражаючи різниці $\nabla^n f_k$ через значення функції f_k у вузлах інтерполяції. Перетворена формула має вигляд

$$y_{k+1} = y_k + h \sum_{n=0}^{N} \alpha_n^* f_{k-n}, \tag{5.53}$$

де α_n^* — коефіцієнти, що визначаються з тотожності

$$\sum_{n=0}^{N} \alpha_n \nabla^n f_k = \sum_{n=0}^{N} \alpha_n^* f_{k-n}.$$

Числові значення кількох перших коефіцієнтів α_n^* наведені в таблиці 10. Для контролю служить рівняння $\sum_{n=0}^{N} \alpha_n^* = 1$.

Як видно з наведеного, з (5.53) випливають формули Адамса—Башфорта (5.42)—(5.44). У рівнянні (5.40) у ролі полінома $p(x)$ візьмемо поліном Ньютона інтерполяції назад, побудований за значенням функції f у вузлах інтерполяції $x_{k-N+1}, \ldots, x_k, x_{k+1}$. Запишемо його у вигляді

$$p(x) = f_{k+1} + (t-1)\nabla f_{k+1} + \frac{(t-1)t}{2!}\nabla^2 f_{k+1} + \ldots +$$

$$+ \frac{(t-1)t \ldots (t+N-2)}{N!}\nabla^N f_{k+1}$$

$$\left(t = \frac{x - x_k}{h}\right). \qquad (5.54)$$

Залишковий член інтерполяції, що відповідає цьому поліному,

$$R = \frac{(t-1)t \ldots (t+N-1)}{(N+1)!} h^{N+1} f^{(N+1)}(x');$$

$$x_{k-N+1} \leq x' \leq x_{k+1}.$$

Після інтегрування (5.40) з урахуванням (5.54) дістанемо

$$y_{k+1} = y_k + h \sum_{n=0}^{N} \beta_n \nabla^n f_{k+1}, \qquad (5.55)$$

де $\beta_0 = 1$, $\beta_n = \frac{1}{n!}\int_0^1 (t-1)t \ldots (t+n-2)dt \quad (n = 1, 2, \ldots, N)$.

Обчислення приблизних значень шуканої функції за формулою (5.55) і є інтерполяційним методом Адамса. Так, при $N = 3$ отримаємо розрахункову формулу

$$y_{k+1} = y_k + h\left(f_{k+1} - \frac{1}{2}\nabla f_k - \frac{1}{12}\nabla^2 f_k - \frac{1}{24}\nabla^3 f_k\right). \qquad (5.56)$$

Формулу (5.56), як і (5.53), можна подати не за різницями $\nabla^n f_{k+1}$, а за значеннями функції у вузлах інтерполяції, при цьому вона набуває вигляду

$$y_{k+1} = y_k + h \sum_{n=0}^{N} \beta_n^* \nabla^n f_{k-n+1}.$$

Числові значення кількох перших коефіцієнтів β_n^* наведені в табл. 11. Для контролю має місце рівняння $\sum_{n=0}^{N} \beta_n^* = 1$. У такому випадку приходимо до методів Адамса—Моултона вигляду (5.45), (5.46).

Таблиця 11

N	n				
	0	1	2	3	4
1	$\frac{1}{2}$	$\frac{1}{2}$			
2	$\frac{5}{12}$	$\frac{8}{12}$	$-\frac{1}{12}$		
3	$\frac{9}{24}$	$\frac{19}{24}$	$-\frac{5}{24}$	$\frac{1}{24}$	
4	$\frac{251}{720}$	$\frac{646}{720}$	$-\frac{264}{720}$	$\frac{106}{720}$	$-\frac{19}{720}$

У розрахункових формулах (5.52) і (5.56) використовуються різниці до третього порядку. Похибка в цих випадках грубо дорівнює першому відкинутому члену і становить

$$\frac{251}{720} h^5 \left.\frac{d^4 f(x, y(x))}{dx^4}\right|_{x=\xi}; \qquad -\frac{19}{720} h^5 \left.\frac{d^4 f(x, y(x))}{dx^4}\right|_{x=\eta}$$

відповідно для (5.52) і (5.56).

Формула (5.56) є алгебраїчним або трансцендентним рівнянням відносно y_{k+1}. Для розв'язання цього рівняння потрібно знати початкове наближення $y_{k+1}^{(0)}$, знайдене за екстраполяційною формулою Адамса, і потім провести уточнення за формулою

$$y_{k+1}^{(i+1)} = y_k + h\left(f_{k+1}^{(i)} - \frac{1}{2}\nabla f_{k+1}^{(i)} - \frac{1}{12}\nabla^2 f_{k+1}^{(i)} - \frac{1}{24}\nabla^3 f_{k+1}^{(i)}\right),$$

$$f_{k+1}^{(i)} = f(x_k, y_k^{(i)}) \qquad (5.57)$$

до збігу двох послідовних наближень із заданою точністю. Тут i — номер ітерації.

Використання формул (5.52), (5.56) в парі (прогнозування y_{k+1} (5.52) і уточнення, коригування за (5.56) (метод прогноз-корекція) дозволяє отримати оцінку похибки розв'язку без додаткових обчислень

$$y(x_{k+1}) = y_{k+1}^{c} + \frac{251}{720} h^5 \left.\frac{d^4 f(x, y(x))}{dx^4}\right|_{x=\xi};$$

$$y(x_{k+1}) = y^{\prime}_{k+1} - \frac{19}{720} h^5 \left.\frac{d^4 f(x, y(x))}{dx^4}\right|_{x=\eta};$$

$$|y(x_{k+1}) - y^{\prime}_{k+1}| = \frac{19}{720} |y^{\prime}_{k+1} - y^{c}_{k+1}| \approx \frac{1}{14} |y^{\prime}_{k+1} - y^{c}_{k+1}|.$$

Якщо $\frac{1}{14} |y^{\prime}_{k+1} - y^{c}_{k+1}|$ менше заданої точності, крок потрібно зменшити. Якщо ви[illegible]а різниця не набагато перевищує припустиму похибку розрахунків, то крок h вважається вибраним вірно і розрахунки продовжують з вибраним кроком.

Розглянемо похибку дискретизації і, щоб не ускладнювати викладення, зупинимось лише на методі Адамса—Башфорта (5.42).

Визначимо локальну похибку дискретизації у точці x як

$$L(x, y) = \frac{1}{h}\left\{y(x + h) - y(x) - \frac{h}{2}[3f(x, y(x)) - f(x - h, y(x - h))]\right\}, \tag{5.58}$$

де $y(x)$ — точний розв'язок диференціального рівняння, оскільки $y'(x) = f(x, y(x))$, можемо переписати (5.58) у термінах y і y' і потім розкласти їх у ряд Тейлора в околі точки x. В результаті дістанемо

$$L(x, y) = \frac{1}{h}\left\{y(x + h) - y(x) - \frac{h}{2}[3f(x, y(x)) - f(x - h, y(x - h))]\right\} =$$

$$= \frac{1}{h}\left\{hy'(x) + \frac{h^2}{2} y''(x) + \frac{h^3}{6} y'''(x) + \frac{h^4}{24} y^4(z_1) - \frac{h}{2}[3y'(x) - \right.$$

$$\left. - y'(x) + hy''(x) - \frac{h^2}{2} y'''(x) + \frac{h^3}{6} y^4(z_2)]\right\} =$$

$$= \frac{5}{12} h^2 y'''(x) + \frac{h^3}{24} y^4(z_1) - \frac{h^4}{12}(z_2), \tag{5.59}$$

де z_1 і z_2 — проміжні точки, що входять в залишкові члени формули Тейлора. Припускаючи тепер, що четверта похідна розв'язку обмежена (а звідси обмежені й всі молодші похідні), бачимо, що локальна похибка дискретизації співвідношення

$$L(h) = \max_{a \le x \le b - h} |L(x, h)| = O(h^2), \tag{5.60}$$

яка показує, що цей метод має другий порядок точності.

Можна було б визначити локальну похибку дискретизації окремо для кожного згаданого в цьому параграфі методу. Проте всі ці методи є окремими випадками лінійних багатокрокових методів і описуються загальною формулою

$$y_{k+1} = \sum_{i=1}^{m} \delta_i y_{k+1-i} + h \sum_{i=0}^{m} \gamma_i f_{k+1-i}, \tag{5.61}$$

де, як завжди, $f_i = F(x_i, y_i)$; m — деяке фіксоване ціле. Методи (5.61) називаються *лінійними*, тому що y_{k+1} є лінійною комбінацією y_i і f_i. Якщо $\gamma_0 \neq 0$, то метод виявляється неявним. У всіх методах Адамса $\delta_1 = 1$ і $\delta_i = 0$ $(i > 1)$; в екстраполяційних методах Адамса $\gamma_0 = 0$; в інтерполяційних методах Адамса $\gamma_0 \neq 0$.

Для загального лінійного багатокрокового методу (5.61) визначимо локальну похибку дискретизації у точці x як

$$L(x, h) = \frac{1}{h}\left[y(x + h) - \sum_{i=1}^{m} \delta_i y(x - (i - 1)h)\right] -$$

$$- \sum_{i=0}^{m} \gamma_i f(x, y(x - (i - 1)h)) = \frac{1}{h}\left[y(x + h) - \sum_{i=1}^{m} \delta_i y(x - (i - 1)h)\right] -$$

$$- \sum_{i=0}^{m} \gamma_i y'(x - (i - 1)h) \tag{5.62}$$

і локальну похибку дискретизації як

$$L(h) = \max_{a \le x \le b - h} |L(x, h)|. \tag{5.63}$$

Для будь-якого конкретного методу, тобто для будь-яких заданих значень m і констант δ_i і γ_i, можна за допомогою розвинень Тейлора для функцій y і y' у точці x обчислити локальну похибку дискретизації. Наприклад, при відповідних припущеннях про диференційовність розв'язку можна показати, що екстраполяційні методи Адамса (5.44) і (5.52), і інтерполяційні методи Адамса (5.46) і (5.56) мають четвертий порядок точності.

Після того як локальна похибка дискретизації знайдена, постає задача оцінки глобальної похибки дискретизації, яка, як і у випадку однокрокових методів, визначається виразом

$$\max_{1 \le x \le N} |y(x_k) - y_k| = \varepsilon.$$

У загальному випадку ця задача достатньо складна, але при відповідних пропозиціях відносно функції f і розв'язку y можна показати, що для всіх методів цього параграфа $\varepsilon = O(h^p)$, якщо $L(h) = O(h^p)$.

На основі методів Адамса розроблено цілий ряд програм на ЕОМ. Методи Адамса в них реалізовані таким чином, що дають можливість міняти не тільки величину кроку, а й порядок самого методу.

Методи Адамса (як інтерполяційний, так і екстраполяційний), а також метод «прогноз-корекцію» можна застосувати і у випадку систем диференціальних рівнянь з будь-якою кількістю рівнянь.

Нехай потрібно, наприклад, розв'язати систему рівнянь

$$y' = f(x, y, z), \quad z' = g(x, y, z)$$

при початкових умовах

$$y|_{x = x_0} = y_0, \quad z|_{x = x_0} = z_0.$$

Визначимо яким-небудь чином додаткові значення приблизних розв'язків $y_1, y_2, \ldots, y_N$, $z_1, z_2, \ldots, z_N$ у точках $x_1 = x_0 + h$, $x_2 = x_0 + 2h$, $\ldots, x_N = x_0 + Nh$. Потім у випадку екстраполяційного методу подальші обчислення можна вести за допомогою формул, аналогічних (5.51)

$$\begin{cases} y_{k+1} = y_k + h \sum\limits_{n=0}^{N} \alpha_n \nabla^n f_k, \\ z_{k+1} = z_k + h \sum\limits_{n=0}^{N} \alpha_n \nabla^n g_k, \end{cases} \tag{5.64}$$

де α_n такі ж, як і в (5.51).

При $N = 3$ ці формули мають вигляд

$$\begin{cases} y_{k+1} = y_k + h\left(f_k + \frac{1}{2}\nabla f_k + \frac{5}{12}\nabla^2 f_k + \frac{3}{8}\nabla^3 f_k\right); \\ z_{k+1} = z_k + h\left(g_k + \frac{1}{2}\nabla g_k + \frac{5}{12}\nabla^2 g_k + \frac{3}{8}\nabla^3 g_k\right). \end{cases} \tag{5.65}$$

Заміняючи різниці $\nabla^n f_k$, $\nabla^n g_k$ на значення функцій f і g у вузлах інтерполяції, формули (5.64) для $N = 3$ можна подати за допомогою методу Адамса—Башфорта (5.44) у вигляді

$$\begin{cases} y_{k+1} = y_k + \frac{h}{24}(55f_k - 59f_{k-1} + 37f_{k-2} - 9f_{k-3}); \\ z_{k+1} = z_k + \frac{h}{24}(55g_k - 59g_{k-1} + 37g_{k-2} - 9g_{k-3}), \end{cases} \tag{5.66}$$

де $f_{k-N} = f(x_{k-N}, y_{k-N}, z_{k-N})$, $g_{k-N} = g(x_{k-N}, y_{k-N}, z_{k-N})$.

В інтерполяційному методі Адамса при розв'язанні системи двох рівнянь використовуються формули

$$\begin{cases} y_{k+1} = y_k + h \sum\limits_{n=0}^{N} \beta_n \nabla^n f_{k+1}; \\ z_{k+1} = z_k + h \sum\limits_{n=0}^{N} \beta_n \nabla^n g_{k+1} \end{cases} \tag{5.67}$$

аналогічно формулам (5.55) з тими ж коефіцієнтами β_n. При $N = 3$ ці формули мають вигляд

$$\begin{cases} y_{k+1} = y_k + h\left(f_{k+1} - \frac{1}{2}\nabla f_k - \frac{1}{12}\nabla^2 f_k - \frac{1}{24}\nabla^3 f_k\right); \\ z_{k+1} = z_k + h\left(g_{k+1} - \frac{1}{2}\nabla g_k - \frac{1}{12}\nabla^2 g_k - \frac{1}{24}\nabla^3 g_k\right). \end{cases} \tag{5.68}$$

Якщо замінити різниці $\nabla^n f_k$ і $\nabla^n g_k$ на значення функцій f і g у вузлах інтерполяції, формули (5.67) для $N = 3$ можна подати за допомогою методу Адамса—Моултона (5.46) у вигляді

$$\begin{cases} y_{k+1} = y_k + \frac{h}{24}(9f_{k+1} + 19f_k - 5f_{k-1} + f_{k-2}); \\ z_{k+1} = z_k + \frac{h}{24}(9g_{k+1} + 19g_k - 5g_{k-1} + g_{k-2}). \end{cases} \tag{5.69}$$

Процес ітерації для методу (5.68) виконується за формулами

$$\begin{cases} y_{k+1}^{(i)} = y_k + h\left(f_{k+1}^{(i)} - \frac{1}{2}\nabla f_{k+1}^{(i)} - \frac{1}{12}\nabla^2 f_{k+1}^{(i)} - \frac{1}{24}\nabla^3 f_{k+1}^{(i)}\right); \\ z_{k+1}^{(i)} = z_k + h\left(g_{k+1}^{(i)} - \frac{1}{2}\nabla g_{k+1}^{(i)} - \frac{1}{12}\nabla^2 g_{k+1}^{(i)} - \frac{1}{24}\nabla^3 g_{k+1}^{(i)}\right), \end{cases}$$

де за початкове наближення $f_{k+1}^{(0)}(x_{k+1}, y_{k+1}^{(e)}, z_{k+1}^{(e)})$, $g_{k+1}^{(0)}(x_{k+1}, y_{k+1}^{(e)}, z_{k+1}^{(e)})$ вибираються значення $y_{k+1}^{(e)}$, $z_{k+1}^{(e)}$, знайдені за екстраполяційним методом (5.65).

Для методу (5.69) значення $f_{k+1}(x_{k+1}, y_{k+1}, z_{k+1})$, $g_{k+1}(x_{k+1}, y_{k+1}, z_{k+1})$ знаходяться за допомогою методу прогнозу (5.66), і обчислювальна схема корекції має вигляд

$$\begin{cases} y_{k+1} = y_k + \frac{h}{24}\left(9f_{k+1}^{(p)} + 19f_k - 5f_{k-1} + f_{k-2}\right); \\ z_{k+1} = z_k + \frac{h}{24}\left(9g_{k+1}^{(p)} + 19g_k - 5g_{k-1} + g_{k-2}\right). \end{cases}$$

Приклад. Розглянемо рівняння (5.25) і розв'яжемо задачу Коші на відрізку [0; 1] з використанням методів Адамса четвертого порядку точності. При цьому вважаємо, що в перших трьох точках $x = 0{,}2;\ 0{,}4;\ 0{,}6$ значення функції обчислені за методом Рунге—Кутта.

Точні розв'язки задачі за методом Рунге—Кутта наведені в табл. 12 і продовжено таблицю значень для $x = 0{,}8;\ 1{,}0$ за методом Адамса—Башфорта (5.48) (верхній рядок) і з корекцією за методом Адамса—Моултона (нижній рядок). У дужках подано різницю між точним і приблизним розв'язками.

Таблиця 12

k	x	Точний розв'язок	Метод Рунге—Кутта	Метод Адамса
0	0	1	1	
1	0,2	1,1832160	1,832292 (0,0000132)	
2	0,4	1,3416407	1,3416668 (0,0000261)	
3	0.6	1,4832397	1,4832847 (0,0000450)	
4	0,8	1,6124516	1,6125186 (0,0000670)	1,6114342 (0,0010174)
				1,612418 (0,0000336)
5	1,0	1, 7320508	1,7321483 (0,0000975)	1,7315654 (0,0005829)
				1,7319612 (0,0000896)

Розв'язок задачі із застосуванням екстраполяційного методу Адамса (5.52) і уточненням одержаного приблизного значення за допомогою ітераційного процесу за інтерполяційним методом Адамса (5.57) наведено в табл. 13.

Таблиця 13

k	x	y	f_k	∇f_k	$\nabla^2 f_k$	$\nabla^3 f_k$
0	0					
1	0,2	1,1832291	0,8451713	−0,1548287		
2	0,4	1,3416668	0,7453936	−0,0997777	0,0550511	
3	0,6	1,4832847	0,6742695	−0,0711241	0,0286536	−0,0263974
		1,6114359	0,6185327	−0,0557368	0,0153873	−0,0132663
4	0,8	1,612419	0,6201211	−0,0541484	0,0169757	−0,0116779
		1,6125381	0,6203135	−0,0539561	0,0171681	−0,0114855
		1,6125166	0,6202788	−0,0539907	0,0171334	−0,0115202
5	1,0	1,7317371	0,5768274	−0,0434514	0,0105393	−0,0065941
		1,7321065	0,5774431	−0,0428357	0,0111551	−0,0059784
		1,7321525				

§ 5.6. ПОНЯТТЯ СТІЙКОСТІ Й ЖОРСТКИХ РІВНЯНЬ ПРИ РОЗВ'ЯЗАННІ ЗАДАЧ

Проблема стійкості — одне з центральних питань наукового програмування. Цей термін використовується досить часто і в залежності від контексту може мати різні значення. Обговоримо кілька аспектів проблеми стійкості в тому сенсі, як вона розуміється при чисельному розв'язанні звичайних диференціальних рівнянь.

Розглянемо диференціальне рівняння другого порядку

$$y'' - 10y' - 11y = 0 \tag{5.70}$$

з початковими умовами

$$y(0) = 1, \quad y'(0) = -1. \tag{5.71}$$

Розв'язком задачі (5.70), (5.71) є функція $y(x) = e^{-x}$. Припустимо, що перша початкова умова змінена на малу величину ε так, що початкові умови набудуть вигляду

$$y(0) = 1 + \varepsilon, \quad y'(0) = -1. \tag{5.72}$$

Тоді, як легко впевнитися безпосередньо підставленням, розв'язком рівняння (5.70) з початковими умовами (5.72) буде функція

$$y(x) = \left(1 + \frac{11}{12}\right)e^{-x} + \frac{\varepsilon}{12}e^{11x}. \tag{5.73}$$

Звідси при будь-якому скільки завгодно малому $\varepsilon > 0$ другий член в (5.73) приводить до того, що розв'язок прямує до нескінченності при $x \to \infty$. Ці дві задачі показано на рис. 9.

Можемо стверджувати, що розв'язок $y(x) = e^{-x}$ задачі (5.70), (5.71) нестійкий. Це означає, що як завгодно малі зміни початкових умов можуть викликати як завгодно великі зміни розв'язку при $x \to \infty$. У чисельному аналізі такі задачі називаються *погано обумовленими*. У цьому випадку дуже складно отримати вказаний розв'язок чисельно, оскільки похибки округлення й відтинання впливають точно так, як і зміна початкових умов, і призводять до того, що розв'язок прямує до нескінченності.

Ще різкіше нестійкість може виявитись у нелінійних рівняннях. Наприклад, задача

$$y' = xy(y - 2), \quad y(0) = 2 \tag{5.74}$$

має розв'язок $y(x) \equiv 2$, який є нестійким. Дійсно, при початковій умові $y(0) = y_0$ розв'язок задається формулою

$$y(x) = 2y_0/(y_0 + (2 - y_0)e^{x^2}).$$

Тому, якщо $y_0 < 2$, то $y(x) \to 0$ при $x \to \infty$, а якщо $y_0 > 2$, то розв'язок зростає й особливістю його є $y_0 + (2 - y_0)e^{x^2} = 0$. Характерну поведінку розв'язків показано на рис. 10.

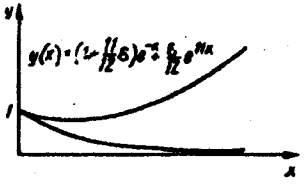


Рис. 9

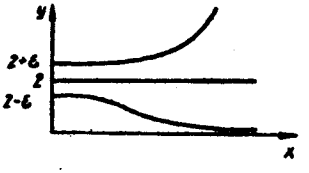


Рис. 10

Два попередні приклади були ілюстрацією нестійкості розв'язків самих диференціальних рівнянь. Звернемося до можливої нестійкості чисельних методів. Розглянемо багатокроковий метод

$$y_{n+1} = y_{n-1} + 2hf_n, \tag{5.75}$$

що має другий порядок точності. Застосуємо метод (5.75) до задачі

$$y' = -2y + 1, \quad y(0) = 1, \tag{5.76}$$

точний розв'язок якої має вигляд

$$y(x) = 0{,}5e^{-2x} + 0{,}5. \tag{5.77}$$

Цей розв'язок є стійким. Дійсно, якщо замінити початкову умову на $y(0) = 1 + \varepsilon$, то розв'язок набуде вигляду

$$y(x) = (0{,}5 + \varepsilon)e^{-2x} + 0{,}5$$

і зміниться тільки на εe^{-2x}.

В застосуванні до задачі (5.76) метод (5.75) визначається формулою

$$y_{n+1} = y_{n-1} + 2h(-2y_n + 1), \qquad y_0 = 1, \tag{5.78}$$

де в ролі y_0 береться початкова умова.

Оскільки метод (5.75) багатокроковий, для обчислення необхідно знати значення y_1. У ролі y_1 візьмемо значення точного розв'язку (5.77) при $x = h$, тобто

$$y = 0{,}5e^{-2x} + 0{,}5. \tag{5.79}$$

Поведінку послідовності $\{y_n\}$, що породжується формулою (5.78), можна досить легко проаналізувати. Будемо розглядати (5.78) як різницеве рівняння. Теорія різницевих рівнянь має багато паралелей з теорією диференціальних рівнянь, і ми коротко зупинимося на основних елементах цієї теорії у випадку лінійних рівнянь порядку m з постійними коефіцієнтами. Такі рівняння мають вигляд

$$y_{n+1} = a_m y_n + .. + a_1 y_{n-m+1} + a_0;$$

$$n = m - 1, m, m + 1, \ldots, \tag{5.80}$$

де $a_0, a_1, \ldots, a_m$ — задані константи, а однорідна частина рівняння (5.80)

$$y_{n+1} = a_m y_n + \ldots + a_1 y_{n-m+1}. \tag{5.81}$$

За аналогією з диференціальними рівняннями спробуємо знайти для рівняння (5.81) розв'язки експоненціального типу; тільки в цьому випадку замість експоненти візьмемо вираз $y_k = \lambda^k$ з деякою невідомою константою λ. Як бачимо, якщо λ задовольняє рівняння

$$\lambda^m - a_m\lambda^{m-1} - \ldots - a_1 = 0, \tag{5.82}$$

яке є характеристичним для (5.81), то $y_k = \lambda^k$ дійсно є розв'язком (5.81). Якщо припустити, що всі m коренів $\lambda_1, \lambda_2, \ldots, \lambda_m$ рівняння (5.82) різні, то послідовності $\lambda_1^k, \lambda_2^k, \ldots, \lambda_m^k$ створюють фундаментальну схему розв'язків і загальний розв'язок рівняння (5.81) можна записати у вигляді

$$y_k = \sum_{i=1}^{m} c_i\lambda_i^k, \quad k = 0, 1, \ldots, \tag{5.83}$$

де c_i — довільні константи. Якщо $1 - a_m - a_{m-1} - \ldots - a_1 \neq 0$, то, як легко перевірити, істинний розв'язок (5.80) виражається формулою

$$y_k = a_0/(1 - a_1 - \ldots - a_m). \tag{5.84}$$

Звідси загальний розв'язок рівняння (5.80) є сума (5.83) і (5.84):

$$y_k = \sum_{i=1}^{m} c_i\lambda_i^k + a_0/(1 - a_1 - \ldots - a_m), \quad k = 0, 1, \ldots. \tag{5.85}$$

Як у випадку диференціальних рівнянь, будь-які константи в (5.85) визначаються з додаткових умов, що накладаються на розв'язок. Так, якщо задані початкові значення

$$y_1, y_2, \ldots, y_{m-1}, \tag{5.86}$$

то з (5.85) випливають умови

$$\sum_{i=1}^{m} c_i\lambda_i^k + a_0/(1 - a_1 - \ldots - a_m) = y_k, \quad k = 0, 1, \ldots, m-1, \tag{5.87}$$

що є системою m лінійних рівнянь відносно m невідомих $c_1, \ldots, c_m$, яку можна використовувати для визначення значень c_i.

Застосуємо цю теорію до різницевого рівняння (5.78), переписавши його таким чином:

$$y_{n+1} = -4hy_n + y_{n-1} + 2h, \quad y_0 = 1;$$

$$y_1 = 0{,}5e^{-2h} + 0{,}5. \tag{5.88}$$

Відповідне характеристичне рівняння $\lambda^2 - 4h\lambda - 1 = 0$ має корені

$$\lambda_1 = -2h + \sqrt{1 + 4h^2}, \quad \lambda_2 = -2h - \sqrt{1 + 4h^2}. \tag{5.89}$$

Розв'язуючи тоді відносно c_1 і c_2 умови (5.87), які в цьому випадку мають вигляд

$$c_1 + c_2 + \frac{1}{2} = y_0 = 1;$$

$$c_1\lambda_1 + c_2\lambda_2 + \frac{1}{2} = y_1 = 0{,}5e^{-2h} + 0{,}5,$$

знаходимо

$$c_1 = \frac{1}{4} + \frac{y_1 - \frac{1}{2} + h}{2\sqrt{1 + 4h^2}}; \quad c_2 = \frac{1}{4} - \frac{y_1 - \frac{1}{2} + h}{2\sqrt{1 + 4h^2}}. \tag{5.90}$$

Таким чином, для розв'язання рівняння (5.88) дістанемо вираз

$$y_n = c_1\left(-2h + \sqrt{1 + 4h^2}\right)^n + c_2\left(-2h - \sqrt{1 + 4h^2}\right)^n + 0{,}5. \tag{5.91}$$

Хоча такий вигляд розв'язку може здаватися трохи громіздким, але він дозволяє легко визначити поведінку y_n при $n \to \infty$. Дійсно, при будь-якій фіксованій величині кроку $h > 0$ очевидно, що

$$0 < -2h + \sqrt{1 + 4h^2} < 1, \quad 2h + \sqrt{1 + 4h^2} > 1.$$

Звідси при $n \to \infty$ перший член в (5.91) прямує до нуля, а другий, осциліюючи, прямує до нескінченності. Оскільки розв'язок (5.77) задачі (5.76) прямує до 0,5 при $n \to \infty$, ясно, що похибка приблизного розв'язку $\{y_n\}$ прямує до нескінченності і метод (5.78) виявляється нестійким. Зауважимо, що це зростання похибки ніяк не пов'язане з похибками округлень. Оскільки формула (5.91) є точним математичним виразом для y_n, і якби послідовність (5.88) обчислювалася в точній арифметиці, значення, що отримуються, повністю б збіглися зі значеннями, що надаються формулою (5.91).

Наведений приклад показує, наскільки важливо, щоб метод був в деякій мірі стійким. Найбільш фундаментальне визначення стійкості можна сформулювати в термінах загального методу:

$$y_{n+1} = \sum_{i=1}^{m} a_i y_{n+1-i} + h\varphi(x_{n+1}, \ldots, x_{n+1-m}, y_{n+1}, \ldots, y_{n+1-m}). \tag{5.92}$$

Метод (5.92) стійкий, якщо всі нулі λ_i полінома

$$p(\lambda) \equiv \lambda^m - a_1\lambda^{m-1} - \ldots - a_m \qquad (5.93)$$

задовольняють умову $|\lambda_i| \leq 1$ і будь-який нуль такий, що $|\lambda_i| = 1$, є простим. Якщо до цього ще й $m-1$ нулів полінома (5.93) такі, що $|\lambda_i| < 1$, метод (5.92) є строго стійким.

Будь-який метод, що має щонайменше перший порядок точності, повинен задовольняти умову $\sum_{i=1}^{m} a_i = 1$ і, отже, й повинно бути нулем відповідного полінома (5.93). У цьому випадку для будь-якого строго стійкого методу поліном (5.93) буде мати один нуль, що дорівнює 1, а всі інші нулі за модулем будуть строго менші за 1. Оскільки методи Рунге—Кутта однокрокові, то для них $p(\lambda) = \lambda - 1$. Цей поліном не має ніяких інших нулів, крім $\lambda = 1$, і тому методи Рунге—Кутта завжди стійкі. У випадку m-крокового методу Адамса $p(\lambda) = \lambda^m - \lambda^{m-1}$, так що всі інші $m-1$ нулів (5.93) дорівнюють нулю, і такі методи теж строго стійкі.

Для методу (5.88) поліном (5.93) набуває вигляду $p(\lambda) = \lambda^2 - 1$ і має два нулі: +1 і -1. Звідси випливає, що цей метод стійкий, але не строго стійкий. Саме відсутність строгої стійкості і призводить до нестійкої поведінки послідовності $\{y_k\}$, що породжується формулою (5.88). Це можна пояснити таким чином: різницеве рівняння (5.88) має другий порядок і, отже, два фундаментальних розв'язки λ_1^n і λ_2^n, де λ_1 і λ_2 — корені характеристичного рівняння, що визначається формулами (5.89). Послідовність $\{y_k\}$, що отримується за методом (5.88), будується з метою апроксимації розв'язку диференціального рівняння першого порядку (5.76), яке має один фундаментальний розв'язок. Цей розв'язок апроксимується послідовністю λ_1^n, послідовність же λ_2^n є «паразитною» і повинна швидко прямувати до нуля. Проте $|\lambda_2| > 1$ при будь-якому $h > 0$ і, звідси, λ_2^n прямує до нескінченності, а не до нуля; саме це і спричиняє нестійкість. Зазначимо, що при $h \to 0$ значення λ_1 і λ_2 прямують до нулів полінома стійкості (5.93). Дійсно, цей поліном є граничним при $h \to 0$ для характеристичного полінома $\lambda^2 + 4h\lambda - 1$ рівняння (5.88). Поняття строгої стойкості стає більш очевидним. Якщо всі, за винятком одного, нулі полінома стійкості за абсолютною величиною менші за 1, то при достатньому малому h всі, крім одного, корені характеристичного рівняння методу, що розглядається, будуть за абсолютною величиною менші за 1. Звідси, степені цих коренів, що є «паразитними» фундаментальними розв'язками різницевого рівняння, будуть прагнути до нуля і не призводитимуть до виникнення нестійкості.

Теорія стійкості, яку ми тільки що обговорили, стосується саме стійкості границі при $h \to 0$. Наведений вище приклад нестійкості показує, що може відбуватись при скільки завгодно малому h, якщо метод стійкий, але не строго стійкий. Проте навіть строго стійкі методи можуть вести себе нестійко, якщо h дуже велике. І хоч цю перепону можна подолати за рахунок зменшення кроку h, це може призвести до неприпустимо великих

затрат машинного часу. Така ситуація виникає при розв'язанні диференціальних рівнянь, які називають *жорсткими.*

Розглянемо рівняння

$$y' = -100y + 100, \quad y(0) = y_0. \tag{5.94}$$

Точним розв'язком цієї задачі є функція

$$y(x) = (y_0 - 1)e^{-100x} + 1. \tag{5.95}$$

Очевидно, що цей розв'язок стійкий. Дійсно, якщо ми замінимо початкову умову на $y_0 + \varepsilon$, то розв'язок зміниться на e^{-100x}. Метод Ейлера, застосований до задачі (5.94), набуває вигляду

$$y_{n+1} = y_n + h(-100y_n + 100) = (1 - 100h)y_n + 100h, \tag{5.96}$$

і точний розв'язок цього різницевого рівняння першого порядку виражається формулою

$$y_n = (y_0 - 1)(1 - 100h)^n + 1. \tag{5.97}$$

Припустимо, що $y_0 = 2$. Тоді точні розв'язки (5.95) і (5.97) матимуть вигляд

$$y(x) = e^{-100x} + 1; \tag{5.98}$$

$$y_n = (1 - 100h)^n + 1. \tag{5.99}$$

Функція $y(x)$ дуже швидко зменшується від $y_0 = 2$ до свого граничного значення 1. Так, $y(0{,}1) \approx 1 + 5 * 10^{-5}$. Тому на початковому етапі ми чекаємо, що для точного обчислення розв'язку потрібно буде вести обчислення з малим кроком h. Проте після, нехай, $x = 0{,}1$, розв'язок змінюється повільно і, загалом, дорівнює 1. І, як бачимо, метод Ейлера повинен дати хорошу точність при порівняно великому кроці h. Але з (5.99) видно, якщо $h > 0{,}02$, то $|1 - 100h| > 1$, і значення y_n зі збільшенням номера кроку починають швидко зростати, що говорить про нестійкість. Із порівняння точних розв'язків (5.98) і (5.99) випливає, що часткові розв'язки рівнянь (5.94) і (5.96) тотожно збігаються (і дорівнюють 1). Величина $(1 - 100h)^n$ слугує апроксимацією експоненціального члена e^{-100x} і дійсно є добрим наближенням при малих h. Це наближення швидко стає нестійким, коли h досягає значення 0,02. І хоч цей експоненціальний член після $x = 0{,}1$ практично не впливає на розв'язок, для збереження стійкості метод Ейлера, як і раніше, вимагає, щоб цей член апроксимувався досить точно. Ця ситуація характерна для жорстких рівнянь: розв'язок має член, внесок якого дуже малий, але звичайні методи для збереження стійкості потребують, щоб цей член апроксимувався досить точно.

Ця проблема часто виникає при розв'язку систем рівнянь і пов'язана з різноманітністю процесів, що описуються даною системою.

Розглянемо, наприклад, рівняння другого порядку

$$y'' + 10y' + 100y = 0. \qquad (5.100)$$

Це рівняння, як показано в § 5.1, можна перетворити в еквівалентну систему двох рівнянь першого порядку, але для нашого випадку достатньо розглянути його в первісній формі. Загальний розв'язок (5.100) має вигляд

$$y(x) = c_1 e^{-100x} + c_2 e^{-x}.$$

Тому розв'язком, що задовольняє початкові умови $y(0) = 0{,}1$, $y'(0) = -2$, є функція

$$y(x) = \frac{1}{100} e^{-100x} + e^{-x}. \qquad (5.101)$$

Після того, як x досягне порядку 0,1, внесок першого члена в розв'язок буде дуже малим. Якщо все ж таки, застосуємо до відповідної рівнянню (5.100) системи першого порядку метод Ейлера, то ми зіткнемося з тією ж самою проблемою, що і в попередньому прикладі: доведеться вибрати крок достатньо малим, щоб точно апроксимувати член e^{-100x}, незважаючи на те, що його внесок у розв'язок дуже малий. Аналогічні складності виникають і при розв'язанні будь-якої системи звичайних диференціальних лінійних рівнянь

$$\frac{d\bar{y}}{dx} = A\bar{y} + \bar{f}, \quad 0 \le x \le 1, \quad \bar{y}(0) = y_0, \qquad (5.102)$$

якщо матриця цієї системи має великий діапазон власних чисел. У (5.102) $\bar{y}, \bar{f}$ — вектори; A — матриця, у якої всі власні значення λ_k мають від'ємні дійсні частини такі, що

$$\mathrm{Re}\,\lambda_k \le \lambda_0 < 0, \quad 1 \le i \le n;$$

$$\min|\mathrm{Re}\,\lambda_k| \approx 1, \quad \max|\mathrm{Re}\,\lambda_k| >> 1, \quad \max|\mathrm{Im}\,\lambda_k| \approx 1. \qquad (5.103)$$

Тут виникає та ж ситуація, що й в описаних вище прикладах: умова стійкості в явній схемі Ейлера не дає можливості інтегрувати з кроком h, який визначається точністю доданків, що повільно змінюються для тих значень x, де $\exp(\lambda_i x)$ з $\mathrm{Re}\lambda_i << 1$ мало відрізняються від нуля.

Для загальних систем нелінійних диференціальних рівнянь поняття жорстких рівнянь вводиться за аналогією з наведеними вище прикладами. Нехай маємо задачу

$$\frac{dy_i}{dx} = f_i(x, y_1, \ldots, y_n), \quad 0 \le x \le 1;$$

$$y_i(0) = y_i^{(0)}, \quad 1 \le i \le n,$$

точний розв'язок якої $y(x) = (y_1(x), \ldots, (y_n(x))$.

Обчислюємо якобіан

$$A(x) = \left(\frac{\partial f_i}{\partial y_j} (x, y_1(x), \dots, y_n(x) \right), \quad 1 \le i, j \le n.$$

Якщо матриця $A(x)$ при деяких $x \in [0, 1]$ має властивість (5.103), то початкова система рівнянь належить до класу жорстких.

Сформулюємо визначення жорсткої системи рівнянь. Розглянемо систему (5.102) з постійною, тобто систему, що не залежить від x, матрицею A. Система диференціальних рівнянь (5.102) з постійною матрицею $A(n * n)$ називається *жорсткою*, якщо:

1) $\mathrm{Re}\lambda_k < 0, \quad k = 1, 2, \dots, m$ (тобто система асимптотично стійка за Ляпуновим);

2) відношення

$$S = \frac{\max\limits_{1 \le k \le n} |\mathrm{Re}\lambda_k|}{\min\limits_{1 \le k \le n} |\mathrm{Re}\lambda_k|}$$

досить велике.

Число S називається *числом жорсткості* системи (5.102). Друга вимога не вказує кордонів для S, починаючи з яких система стає жорсткою.

Якщо матриця A залежить від x, то $\lambda_k = \lambda_k(x)$, $k = 1, 2, \dots, n$. При кожному x можна визначити число жорсткості

$$S(x) = \frac{\max\limits_{1 \le k \le n} |\mathrm{Re}\lambda_k(x)|}{\min\limits_{1 \le k \le n} |\mathrm{Re}\lambda_k(x)|}.$$

Система (5.102) називається *жорсткою на інтервалі* $[0, x]$, якщо $\mathrm{Re}\lambda_k(x) < 0$, $k = 1, 2, \dots, n$ для всіх $x \in [0, X]$ і число $\sup\limits_{x \in [0, X]} S(x)$ досить велике. У цьому випадку властивість жорсткості може залежати від довжини відрізка.

Не випадково в цьому параграфі інтервал інтегрування строго фіксовано одиничним $0 \le x \le 1$. Велика довжина інтервалу $a \le x \le b$ може привести до того, що система рівнянь повинна розглядатись як жорстка, хоч $S(x)$ і не великий на $[a, b]$. Це легко зрозуміти з простого прикладу. Нехай маємо задачу

$$\frac{dy}{dx} = -y, \quad y(0) = y^{(0)}, \quad 0 \le x \le b.$$

Коефіцієнт жорсткості $S(x) = 1$ на всьому інтервалі $[0, b]$. Але якщо привести заміною змінної $x_1 = x/b$ цю задачу до нормованого інтервалу $0 \le x_1 \le 1$, то дістанемо

$$\frac{dy}{dx_1} = -by, \quad y(0) = y^{(0)}, \quad 0 \le x_1 \le 1,$$

при $b >> 1$ жорстке рівняння.

Звідси доцільно зробити практичний висновок: інтегрування задачі Коші на великих інтервалах може призвести до явища жорсткості, яке потрібно враховувати при виборі методу розв'язання.

Таким чином, розв'язок жорсткої системи має як елементи, що зменшуються швидко, так і елементи, що зменшуються повільно. Починаючи з деякого $x > 0$, розв'язок системи майже повністю визначається елементом, що зменшується повільно. Але в обчислювальній схемі необхідно враховувати і елементи, що зменшуються швидко, які вже практично не вносять свій внесок до розв'язку, і при використанні явних різницевих методів погано впливають на стійкість, що змушує брати крок інтегрування надто малим.

Звичайний підхід до розв'язання проблеми жорсткості полягає у використанні неявних методів. Розглянемо простий випадок, що полягає у використанні для розв'язання рівняння $y' = f(x, y)$ формули

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}), \tag{5.104}$$

що називається *неявним методом Ейлера.*

Застосуємо (5.104) до задачі (5.94)

$$y_{n+1} = y_n + h(-100y_{n+1} + 100),$$

яке можна записати таким чином:

$$y_{n+1} = (1 + 100h)^{-1}(y_n + 100h). \tag{5.105}$$

Точний розв'язок рівняння (5.105) задається виразом

$$y_n = (y_0 - 1)(1 + 100h)^{-n} + 1, \tag{5.106}$$

яке при початковій умові $y_0 = 2$ набуває вигляду

$$y_n = (1 + 100h)^{-n} + 1. \tag{5.106}$$

Як бачимо, розв'язок буде стійким при будь-якій величині кроку. Проте вибір кроку обумовлюється вимогою доброго наближення до точного розв'язку.

Наявний метод Ейлера має тільки перший порядок точності; тому до кращих результатів привело б використання методів Адамса—Моултона вищих порядків,

Використання неявного методу для розв'язку рівняння (5.94) виглядає досить простим тому, що це рівняння є лінійним і відповідне різницеве рівняння (5.105) легко розв'язується відносно y_{n+1}. Якщо б диференціальне рівняння було нелінійним, то для визначення y_{n+1} потрібно було б на кожному кроці розв'язувати нелінійне рівняння. У загальному випадку системи диференціальних рівнянь на кожному кроці доводиться розв'язувати систему рівнянь. Дійсно, це пов'язано з великими витратами машинного часу, але для ефективного розв'язку жорстких рівнянь чисельний метод повинен мати ті чи інші неявні елементи.

ГЛАВА 6

МЕТОДИ РОЗВ'ЯЗАННЯ ЛІНІЙНИХ КРАЙОВИХ ЗАДАЧ ДЛЯ ЗВИЧАЙНИХ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ

§ 6.1. ПОСТАНОВКА ЗАДАЧІ. ЗВЕДЕННЯ ДО ЗАДАЧІ КОШІ

Сформулюємо лінійну крайову задачу для одного звичайного диференціального рівняння n-го порядку. Розглянемо двоточкову крайову задачу, яка описується звичайним диференціальним рівнянням зі змінними коефіцієнтами:

$$L(y) = r(t) \quad (a \le t \le b), \tag{6.1}$$

де

$$L(y) = \sum_{k=0}^{n} f_k(t) y^{(n-k)} =$$

$$= f_0(t)y^{(n)} + f_1(t)y^{(n-1)} + f_2(t)y^{(n-2)} + \ldots + f_{n-1}(t)y' + f_n(t)y; \tag{6.2}$$

з граничними умовами:

$$U_k[y] = \gamma_k \quad (k = 1, 2, 3, \ldots, n), \tag{6.3}$$

де

$$U_k[y] = \sum_{s=0}^{n-1} [\alpha_{ks} y^{(s)}(a) + \beta_{ks} y^{(s)}(b)], \tag{6.4}$$

$r(t)$, $f_k(t)$ — задані неперервні функції, γ_k, α_{ks}, β_{ks} — сталі.

Розв'язати крайову задачу (6.1), (6.3) означає знайти таку функцію $y(t)$, яка задовольняє диференціальне рівняння (6.1) і граничну умову (6.3).

Розглянемо крайову задачу для диференціального рівняння другого порядку

$$y'' + f_1(t)y' + f_2(t)y = r(t) \quad (a \le t \le b) \tag{6.5}$$

з граничними умовами

$$\alpha_{10}y(a) + \alpha_{11}y'(a) = \gamma_{11}; \quad \beta_{10}y(b) + \beta_{11}y'(b) = \gamma_{12}. \tag{6.6}$$

Розв'язок крайової задачі (6.5), (6.6) шукаємо у вигляді

$$y(t) = Cy_1(t) + y_2(t), \tag{6.7}$$

де $y_1(t)$ — розв'язок задачі Коші для однорідного рівняння (6.5) (тобто при $r(t) = 0$); $y_2(t)$ — розв'язок задачі Коші для неоднорідного рівняння (6.5), при $r(t) \not\equiv 0$ C — довільна стала,

Вимагатимемо, щоб розв'язок (6.7) крайової задачі задовольняв граничну умову (6.6) при довільному C. Маємо:

$$C\alpha_{10}y_1(a) + \alpha_{10}y_2(a) + C\alpha_{11}y_1'(a) + \alpha_{11}y_2'(a) = \gamma_{11}, \qquad (6.8)$$

або

$$C[\alpha_{10}y_1(a) + \alpha_{11}y_1'(a)] + \alpha_{10}y_2(a) + \alpha_{11}y_2'(a) = \gamma_{11}.$$

Звідси випливає, що для задоволення рівності (6.8) повинні виконуватись рівності

$$\alpha_{10}y_1(a) + \alpha_{11}y_1'(a) = 0; \qquad (6.9)$$

$$\alpha_{10}y_2(a) + \alpha_{11}y_2'(a) = \gamma_{11}. \qquad (6.10)$$

Для забезпечення рівностей (6.9), (6.10) достатньо, наприклад, покласти

$$y_1(a) = \alpha_{11}m, \quad y_1'(a) = -\alpha_{10}m \quad (m \neq 0); \qquad (6.11)$$

$$y_2(a) = \gamma_{11}/\alpha_{10}, \quad y_2'(a) = 0, \qquad (6.12)$$

якщо $\alpha_{10} \neq 0$, або

$$y_2(a) = 0, \quad y_2' = \gamma_{11}/\alpha_{11}, \qquad (6.13)$$

якщо $\alpha_{11} \neq 0$.

Як видно, $y_1(t)$ — розв'язок задачі Коші для однорідного рівняння (6.5) з початковими умовами (6.11), а $y_2(t)$ — розв'язок задачі Коші для неоднорідного рівняння (6.5) з початковими умовами (6.12) або (6.13).

Виберемо сталу C таким чином, щоб функція (6.7) задовольнила другу граничну умову (6.6). Підставляючи (6.6) в другу умову (6.7), маємо

$$C[\beta_{10}y_1(b) + \beta_{11}y_1'(b)] + \beta_{10}y_2(b) + \beta_{11}y_2'(b) = \gamma_{12},$$

звідки

$$C = \frac{\gamma_{12} - \beta_{10}y_2(b) - \beta_{11}y_2'(b)}{\beta_{10}y_1(b) + \beta_{11}y_1'(b)}, \qquad (6.14)$$

де знаменник $\beta_{10}y_1(b) + \beta_{11}y_1'(b) \not\equiv 0$, що випливає з вимоги існування розв'язку.

Таким чином, крайова задача (6.5), (6.6) звелась до двох задач Коші для функцій $y_1(t)$ і $y_{2(0)}$ з відповідними початковими умовами.

Розглянемо крайову задачу для диференціального рівняння n-го порядку. У цьому випадку розв'язок крайової задачі (6.1), (6.3) шукаємо у вигляді

$$y(t) = \sum_{j=1}^{n} C_j y_j(t) + y_{n+1}(t), \tag{6.15}$$

де $y_j(t)$ $(j = 1, 2, \ldots, n)$ — лінійно незалежні розв'язки задач Коші для однорідного рівняння (6.1), тобто при $r(t) = 0$, $y_{n+1}(t)$ — розв'язок задачі Коші для неоднорідного рівняння (6.1) при $r(t) \neq 0$ C_k — довільні сталі.

Для заданих задач Коші можна сформулювати такі початкові умови:

$$y_j^{(s)}(a) = \begin{cases} 0, & s \neq j-1 \quad (s = 0, 1, 2, \ldots, n-1); \\ 1, & s = j-1 \quad (s = 1, 2, 3, \ldots, n); \end{cases} \tag{6.16}$$

$$y_{n+1}^{(s)} = 0 \quad (s = 0, 1, 2, 3, \ldots, n-1). \tag{6.17}$$

Розв'язавши задачі Коші для функцій $y_j(t)$ $(j = 1, 2, \ldots, n+1)$ і задовольнивши умови (6.16), (6.17), підставимо розв'язок (6.15) в граничні умови (6.3) і дістанемо систему n рівнянь для визначення невідомих сталих C_j $(j = 1, 2, \ldots, n)$. Знайшовши ці сталі з розв'язання отриманої системи рівнянь і підставивши у вираз (6.15), матимемо шуканий розв'язок вихідної задачі (6.1), (6.3).

Розглянемо випадок, коли крайова задача описується системою диференціальних рівнянь у нормальній формі Коші у вигляді

$$\frac{d\bar{y}}{dt} = A(t)\bar{y} + \bar{f}(t), \quad a \leq t \leq b \tag{6.18}$$

з граничними умовами

$$B_1\bar{y}(a) = \bar{b}_1; \tag{6.19}$$

$$B_2\bar{y}(b) = \bar{b}_2, \tag{6.20}$$

де $\bar{y} = \{y_1, y_2, \ldots, y_n\}$ — вектор-стовпець; $\bar{f}$ — вектор-стовпець правої частини; $A(t)$ — задана квадратна матриця порядку n; b_1, b_2 — задані вектори.

Лінійну крайову задачу для одного диференціального рівняння (6.1) можна звести до системи рівнянь вигляду (6.18) з відповідними граничними умовами, і навпаки.

У випадку, що розглядається, розв'язок крайової задачі (6.18)—(6.20) шукатимемо у вигляді

$$\bar{y}(t) = \sum_{j=1}^{m} C_j\bar{y}_j(t) + \bar{y}_{m+1}(t), \tag{6.21}$$

де $m = \min(k, n-k)$ (для визначеності покладемо $m = n-k$), $\bar{y}_j$ $(j = 1, 2, \ldots, m)$ — розв'язки задач Коші для системи рівнянь (6.18) при $\bar{f} = 0$ з початковими умовами, що задовольняють граничну умову на кінці інтервалу (6.19) при $\bar{b}_1 = 0$; $\bar{y}_{m+1}$ — розв'язок задачі Коші для системи (6.18) з

початковими умовами, що задовольняють граничні умови (6.19); m — число граничних умов на правому кінці інтервалу інтегрування,

Виконаємо зазначені вимоги, подавши їх у точці $t = a$ в розгорнутому вигляді,

$$b_{11}y_1 + b_{12}y_2 + \ldots + b_{1k}y_k + b_{1,k+1}y_{k+1} + \ldots + b_{1n}y_n = b_1;$$

$$b_{21}y_1 + b_{22}y_2 + \ldots + b_{2k}y_k + b_{2,k+1}y_{k+1} + \ldots + b_{2n}y_n = b_2;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$b_{k1}y_1 + b_{k2}y_2 + \ldots + b_{kk}y_k + b_{k,k+1}y_{k+1} + \ldots + b_{kn}y_n = b_k. \quad (6.22)$$

Вважаючи, що коефіцієнти перших k стовпчиків у рівностях (6.22) утворюють неособливу матрицю, перенесемо інші стовпчики в праву частину. Запишемо умови (6.22)

$$b_{11}y_1 + b_{12}y_2 + \ldots + b_{1k}y_k = b_1 - b_{1,k+1}y_{k+1} - \ldots - b_{1n}y_n;$$

$$b_{21}y_1 + b_{22}y_2 + \ldots + b_{2k}y_k = b_2 - b_{2,k+1}y_{k+1} - \ldots - b_{2n}y_n;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$b_{k1}y_1 + b_{k2}y_2 + \ldots + b_{kk}y_k = b_k - b_{k,k+1}y_{k+1} - \ldots - b_{kn}y_n. \quad (6.23)$$

Надаючи компонентам $y_{k+1}, y_{k+2}, \ldots, y_n$ послідовно значення стовпчиків одиничної матриці та покладаючи $b_1 = b_2 = \ldots = b_k = 0$, визначимо початкові умови для $\bar{y}_j$ $(j = 1, 2, \ldots, m)$; при $y_{k+1} = y_{k+2} = \ldots = y_n = 0$ знаходимо початкові умови для $\bar{y}_{m+1}$. Одержавши розв'язки сформульованих задач Коші, вираз (6.21) підставляємо в граничні умови (6.20) і отримуємо систему m рівнянь, з якої знаходимо невідомі сталі C_j $(j = 1, 2, \ldots, m)$. Маючи розв'язки задач Коші $\bar{y}_j(t)$ $(j = 1, 2, \ldots, m+1)$ і значення сталих C_j, де $(j = 1, 2, \ldots, m)$, за допомогою виразу (6.21) визначимо розв'язок крайової задачі (6.18)—(6.20).

Поряд з двоточковою задачею розглянемо ще триточкову для диференціального рівняння третього порядку:

$$y''' + f_1(t)y'' + f_2(t)y' + f_3(t)y = r(t) \quad (6.24)$$

з граничним умовами

$$y'(a) = 0, \quad y(b) = 0, \quad y'(c) = 0. \quad (6.25)$$

Розв'язок крайової задачі (6.24), (6.25) шукаємо у вигляді

$$y(t) = y_0(t) + C_1y_1(t) + C_2y_2(t), \quad (6.26)$$

маючи на увазі, що в точці $t = 0$ недостає ще двох початкових умов. Для знаходження загального розв'язку крайової задачі сформулюємо три задачі Коші:

$$y_0''' + f_1(t)y_0'' + f_2(t)y_0' + f_3(t)y_0 = r(t),$$

$$y_0(0) = 0, \; y_0'(0) = 0, \; y_0''(0) = 0; \quad (6.27)$$

$$y_1''' + f_1(t)y_1'' + f_2(t)y_1' + f_3(t)y_1 = 0,$$

$$y_1(0) = 0,\ y_1'(0) = 0,\ y_1''(0) = 0; \tag{6.28}$$

$$y_2''' + f_1(t)y_2'' + f_2(t)y_2' + f_3(t)y_2 = 0,$$

$$y_2(0) = 0,\ y_2'(0) = 0,\ y_2''(0) = 0. \tag{6.29}$$

Після розв'язання задач Коші (6.27)—(6.29), задовольняючи дві останні граничні умови (6.25), знаходимо сталі з системи рівнянь

$$y_0(b) + C_1y_1(b) + C_2y_2(b) = 0;$$

$$y_0'(c) + C_1y_1'(c) + C_2y_2'(c) = 0. \tag{6.30}$$

Це система двох алгебраїчних рівнянь відносно невідомих C_1 і C_2, коефіцієнти якої знаходяться з розв'язання задач Коші (6.27), (6.29).

Як приклад розглянемо деформацію рівномірно навантаженої тришарової балки, кут зсуву якої описується звичайним диференціальним рівнянням

$$\psi''' - k^2\psi' + a = 0 \quad (0 \le t \le 1), \tag{6.31}$$

де k^2 і a — механічні параметри, що залежать від пружних властивостей шарів.

Кінці балки шарнірно оперті, що виражається такими граничними умовами:

$$\psi'(0) = 0, \quad \psi'(1) = 0. \tag{6.32}$$

Ще одна умова випливає з симетрії задачі і має вигляд

$$\psi(1/2) = 0. \tag{6.33}$$

Таким чином, рівняння (6.31) і граничні умови (6.32), (6.33) визначають триточкову крайову задачу. Шукаємо її розв'язок:

$$\psi(t) = \psi_0(t) + C_1\psi_1(t) + C_2\psi_2(t). \tag{6.34}$$

Для визначення сталих C_1 і C_2 сформулюємо три задачі Коші:

$$\psi_0''' - k^2\psi_0' + a = 0,\ \psi_0(0) = 0,\ \psi_0''(0) = 0,\ \psi_0'(0) = 0; \tag{6.35}$$

$$\psi_1''' - k^2\psi_1' = 0,\ \psi_1(0) = 1,\ \psi_1'(0) = 0,\ \psi_1''(0) = 0; \tag{6.36}$$

$$\psi_2''' - k^2\psi_2'' = 0,\ \psi_2(0) = 0,\ \psi_2'(0) = 0,\ \psi_2''(0) = 1. \tag{6.37}$$

Задовольняючи після розв'язання граничну умову (6.33) і другу граничну умову (6.32), отримуємо систему

$$\psi_0(1/2) + C_1\psi_1(1/2) + C_2\psi_2(1/2) = 0, \tag{6.38}$$

$$\psi_0'(1) + C_1\psi_1'(1) + C_2\psi_2'(1) = 0.$$

Знайшовши C_1 і C_2 і підставивши їх у вираз (6.34), маємо розв'язок вихідної крайової задачі (6.31)—(6.33). Значення функції ψ при $a = 1$, $k = 5$; 10 наведені в табл. 14. Одержані значення розв'язку крайової задачі відповідають точному розв'язку цієї задачі.

Таблиця 14

k	t	$\psi(t)$
5,0	0,0	−0,0121
	0,2	−0,0092
	0,4	−0,0033
	0,6	0,0033
	0,8	0,0092
	1,0	0,0121
10,0	0,0	−0,0040
	0,2	−0,0029
	0,4	−0,0010
	0,6	0,0010
	0,8	0,0029
	1,0	0,0040

§ 6.2. МЕТОД ДИФЕРЕНЦІАЛЬНОЇ ПРОГОНКИ

Викладений у попередньому параграфі підхід до розв'язання крайових задач для лінійних диференціальних рівнянь, що зводиться до задач Коші, не завжди приводить до надійного результату в зв'язку з тим, що в жорстких системах має місце нестійкість обчислень,

Це зумовлено тим, що за певних умов власні значення диференціального рівняння чи матриці системи диференціальних рівнянь істотно відрізняються за величиною дійсної частини, і при інтегруванні зі зростанням аргументу за рахунок втрати значущих цифр система розв'язків задач Коші стає майже лінійно залежною. Тому у зв'язку з цим не можна з достатньою точністю при задоволенні граничних умов на другому кінці інтервалу визначити сталі інтегрування й самі шукані функції. Може статися, що в розв'язку не залишиться жодного правильного знака.

Для уникнення вказаних труднощів, розроблено ряд методів, за допомогою яких чисельне розв'язання крайових задач зводиться до стійкого обчислювального процесу. Одним з них є метод диференціальної прогонки, який дозволяє побудувати алгоритм стійкого обчислювання розв'язку крайової задачі для жорсткого диференціального рівняння (див. гл. 5).

Суть методу розглянемо спочатку на прикладі крайової задачі для лінійного диференціального рівняння другого порядку:

$$y'' - p(t)y = q(t) \quad (a \le t \le b) \tag{6.39}$$

з такими граничними умовами:

$$y'(a) - \alpha_{00}y(a) = \alpha_{10}, \tag{6.40}$$

$$y'(b) - \beta_{00}y(b) = \beta_{10}, \tag{6.41}$$

де $p > 0$, $\alpha_{00} > 0$.

Повне диференціальне рівняння з членом, що утримує y', не важко звести до рівняння (6.39).

Ідея методу полягає в тому, що замість диференціального рівняння другого порядку розглядається диференціальне рівняння першого порядку, тобто рівняння, розв'язок якого утримує один параметр. При цьому ставиться вимога, щоб невідомі коефіцієнти при функції та вільний член вибрати у вигляді таких функцій, щоб розв'язок цього рівняння задовольняв вихідне рівняння (6.39) і граничні умови (6.40), (6.41). За рахунок обчислення задачі Коші для знаходження цих двох невідомих функцій і розв'язання задачі Коші для побудованого диференціального рівняння першого порядку можна знайти розв'язок вихідної крайової задачі при стійкому обчислювальному процесі.

Отже, розглянемо диференціальне рівняння першого порядку

$$y' - \alpha_0(t)y = \alpha_1(t), \tag{6.42}$$

в якому треба вибрати $\alpha_0(t)$ і $\alpha_1(t)$ так, щоб функція $y(t)$ задовольняла рівняння (6.39).

Продиференціюємо диференціальне рівняння (6.42), і отримаємо

$$y'' - \alpha_0(t)y' + \alpha_0'(t)y = \alpha_1'(t). \tag{6.43}$$

Замінюючи похідну y' її виразом з (6.42) дістанемо

$$y'' - \alpha_0(t)[\alpha_0(t)y + \alpha_1(t)] + \alpha_0'(t)y = \alpha_1'(t),$$

або

$$y''' - [\alpha_0'(t) + \alpha_0^2(t)]y = \alpha_1'(t) + \alpha_0(t)\alpha_1(t). \tag{6.44}$$

З порівняння рівняння (6.44) з рівнянням (6.39) випливає, що

$$\alpha_0' + \alpha_0^2 = p(t); \tag{6,45}$$

$$\alpha_1' + \alpha_0\alpha_1 = q(t), \tag{6.46}$$

тобто отримуємо два диференціальних рівняння. Порівнюючи (6.42) в точці $t = a$, з рівністю (6.40) знаходимо, що

$$\alpha_0(a) = \alpha_{00}; \quad \alpha_1(a) = \alpha_{10}. \tag{6.47}$$

Таким чином, диференціальні рівняння (6.45) і (6.46) разом з початковими умовами (6.47) утворюють дві задачі Коші, які можна розв'язати яким-небудь чисельним методом, викладеним в гл. 5.

Розв'язавши згадані задачі Коші, знаходимо $\alpha_0(b)$ і $\alpha_1(b)$.

Підставляючи одержані значення в рівняння (6.42), отримаємо

$$y'(b) + \alpha_0(b)y(b) = \alpha_1(b). \tag{6.48}$$

Тепер можна розглянути граничну умову в точці $t = b$ (6.41) разом з рівністю (6.48) як систему алгебраїчних рівнянь відносно $y'(b)$ і $y(b)$. Розв'язуючи цю систему рівнянь, матимемо

$$y(b) = \frac{\beta_{10} - \alpha_1(b)}{\alpha_0(b) - \beta_{00}}, \tag{6.49}$$

$$y'(b) = \frac{\beta_{10}\alpha_0(b) - \beta_{00}\alpha_1(b)}{\alpha_0(b) - \beta_{00}}. \tag{6.50}$$

Після цього можна було б знайти розв'язок вихідної крайової задачі (6.39)—(6.41), розв'язавши задачу Коші для диференціального рівняння (6.39) з початковими умовами (6.49), (6.50). Але ж це може привести до значної втрати точності, що обумовлено тими ж умовами, про які йшлося вище. І тепер вже не потрібно розв'язувати задачу для вихідного диференціального рівняння (6.39). Для знаходження розв'язку вихідної крайової задачі можна використати рівняння (6.42) з початковою умовою (6.49), тобто розв'язати задачу Коші назад від точки $t = b$ до точки $t = a$.

Таким чином, процес розв'язання крайової задачі (6.39)—(6.41) полягає в послідовному розв'язанні двох задач Коші для рівнянь (6.45), (6.46) з умовами (6.47), на лівому кінці інтервалу в точці $t = a$, що є прямим ходом прогонки і задачі Коші для рівняння (6.42) з умовою (6.49) на правому кінці інтервалу в точці $t = b$, що є зворотним ходом прогонки. Тобто при прямому ході гранична умова з лівого кінця інтегрування переганяється на правий кінець. При такому підході до розв'язання задачі обчислення розв'язків усіх задач Коші виконується за допомогою стійких процесів. Необхідно звернути увагу на те, що при реалізації прямого ходу доцільно зберігати значення функцій $\alpha_0(t)$ і $\alpha_1(t)$ на інтервалі $[a, b]$ для їх використання на зворотному ході при знаходженні розв'язку $y(t)$.

Характерна властивість методу прогонки полягає в тому, що розв'язки задач Коші при прямому ході зростають повільно у порівнянні зі зростанням аргументу. Те ж саме має місце при зворотному ході прогонки.

Приклад. Треба знайти розв'язок крайової задачі для диференціального рівняння

$$y'' - k^2 y = 0 \quad (0 \le t \le b) \tag{6.51}$$

з граничними умовами

$$y'(0) = -k, \quad y'(b) = -ke^{-kb}. \tag{6.52}$$

Розв'язання. Відповідно до методу диференціальної прогонки при прямому ході маємо дві задачі Коші:

$$\alpha_0' + \alpha_0^2 = k^2, \quad \alpha_0(0) = 0; \tag{6.53}$$

$$\alpha_1' + \alpha_0(t)\alpha_1 = 0, \quad \alpha_1(0) = -k, \tag{6.54}$$

після розв'язання яких на зворотному ході розв'язуємо задачу Коші:

$$y^{(1)} - \alpha_0(t) = \alpha_1(t), \quad y(b) = -\frac{\alpha_1(b) + ke^{-kb}}{\alpha_0(b)}. \tag{6.55}$$

Розглянемо застосування методу диференціальної прогонки до розв'язання крайових задач для диференціальних рівнянь третього порядку. Спочатку зупинимось на крайовій задачі диференціального рівняння на інтервалі $[0, 1]$ з однією граничною умовою в точці $t = 0$ і двома в точці $t = 1$, тобто маємо рівняння

$$y''' - p(t)y' - q(t)y = r(t) \tag{6.56}$$

і граничні умови

$$y''(0) - \alpha_{00}y'(0) - \alpha_{10}y(0) = \alpha_{20}; \tag{6.57}$$

$$y''(1) - \beta_{00}y'(1) - \beta_{10}y(1) = \beta_{20}; \tag{6.58}$$

$$y''(1) - \gamma_{00}y'(1) - \gamma_{10}y(1) = \gamma_{20}. \tag{6.59}$$

Побудуємо рівняння другого порядку з невідомими функціональними коефіцієнтами

$$y'' - \alpha_0(t)y' - \alpha_1(t)y = \alpha_2(t). \tag{6.60}$$

Після диференціювання рівняння (6.60) і виключення y'' за допомогою (6.60), отримуємо

$$y''' - (\alpha_0' + \alpha_0^2 + \alpha_1)y' - (\alpha_1' + \alpha_0\alpha_1)y = \alpha_2' + \alpha_0\alpha_2. \tag{6.61}$$

Як і у випадку рівняння другого порядку, вибираємо функції $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$ так, щоб функція $y(t)$ задовольнила рівняння (6.56). Для цього маємо такі рівняння:

$$\begin{aligned} \alpha_0' + \alpha_0^2 + \alpha_1 &= p(t); \\ \alpha_1' + \alpha_0\alpha_1 &= q(t); \\ \alpha_2' + \alpha_0\alpha_2 &= r(t). \end{aligned} \tag{6.62}$$

Початкові умови для цих рівнянь запишемо у вигляді

$$\alpha_0(0) = \alpha_{00}, \quad \alpha_1(0) = \alpha_{10}, \quad \alpha_2(0) = \alpha_{20}. \tag{6.63}$$

Після інтегрування задач Коші для рівнянь (6.62) з умовами (6.63) на інтервалі $[0, 1]$ знаходимо $\alpha_0(1)$, $\alpha_1(1)$, $\alpha_2(1)$. Для $t = 1$ з (6.60) дістанемо

$$y''(1) - \alpha_0(1)y'(1) - \alpha_1(1)y(1) = \alpha_2(1). \tag{6.64}$$

Три рівності (6.58), (6.59) і (6.64) задовольняють систему для знаходження $y(1)$, $y'(1)$ і $y''(1)$. Знайшовши ці значення, можна проінтегрувати рівняння (6.56) від $t = 1$ до $t = 0$, розв'язуючи задачу Коші.

Розглянемо крайову задачу для диференціального рівняння третього порядку з граничними умовами в точках інтервалу інтегрування. Отже, нехай крайова задача описується рівнянням

$$y''' - p(t)y' - q(t)y = r(t) \tag{6.65}$$

з граничними умовами

$$y'(0) - \alpha_{00}y''(0) - \alpha_{10}y(0) = \alpha_{20}; \tag{6.66}$$

$$y(b) - \beta_{00}y''(b) - \beta_{10}y'(b) = \beta_{20}; \tag{6.67}$$

$$y'(c) - \gamma_{00}y''(c) - \gamma_{10}y(c) = \gamma_{20}. \tag{6.68}$$

Побудуємо два диференціальних рівняння другого порядку. Введемо рівняння

$$\alpha_0(t)y'' - y' + \alpha_1(t)y + \alpha_2(t) = 0. \tag{6.69}$$

Диференціюючи рівняння (6.69) і виключаючи y'' за допомогою (6.69), отримаємо рівняння

$$\alpha_0^2(t)y'' - [1 - \alpha_0'(t) - \alpha_0(t)\alpha_1(t)]y' - [\alpha_1(t)\alpha_0'(t) - \alpha_1(t) - \alpha_0(t)\alpha_1'(t)]y = -[\alpha_2(t) - \alpha_2(t)\alpha_0'(t) + \alpha_0(t)\alpha_2'(t)]. \tag{6.70}$$

Порівнюючи рівняння (6.70) з (6.65), дістанемо систему рівнянь:

$$\alpha_0'(t) + \alpha_0(t)\alpha_1(t) + p(t)\alpha_0^2(t) = 1; \tag{6.71}$$

$$\alpha_0'(t) + \alpha_1^2(t) + p(t)\alpha_0(t)\alpha_1(t) + q(t)\alpha_0(t) = 0; \tag{6.72}$$

$$\alpha_2'(t) + \alpha_1(t)\alpha_2(t) + p(t)\alpha_0(t)\alpha_2(t) + r(t)\alpha_0(t) = 0. \tag{6.73}$$

Додаючи до цієї системи початкові умови

$$\alpha_0(0) = \alpha_{00}, \quad \alpha_1(0) = \alpha_{10}, \quad \alpha_2(0) = \alpha_{20}, \tag{6.74}$$

приходимо до задачі Коші.

Введемо друге диференціальне рівняння, що відповідає граничній умові при $t = b$:

$$\beta_0(t)y'' + \beta_1(t)y' - y = -\beta_2(t). \tag{6.75}$$

Диференціюючи (6.75) і виключаючи y'' за допомогою (6.75), отримуємо рівняння

$$\beta_0^2y''' - (\beta_0 - \beta_0\beta_1' + \beta_1\beta_0' + \beta_1^2)y' + (\beta_0' + \beta_1)y = -(\beta_2\beta_0' + \beta_1\beta_2 - \beta_0\beta_2'). \tag{6.76}$$

Порівнюючи рівняння (6.76) з вихідним рівнянням (6.65), дістанемо систему рівнянь:

$$\beta_0' + \beta_1 + \beta_0^2 q(t) = 0; \tag{6.77}$$

$$\beta_1' + \beta_0\beta_1 q(t) + \beta_0 p(t) = 1; \tag{6.78}$$

$$\beta_2' + \beta_0\beta_2 q(t) + \beta_0 r(t) = 0. \tag{6.79}$$

Граничні умови мають вигляд

$$\beta_0(b) = \beta_{00}, \quad \beta_1(b) = \beta_{10}, \quad \beta_2(b) = \beta_{20}. \tag{6.80}$$

Задачі Коші (6.71)—(6.74) і (6.77)—(6.80) треба інтегрувати до точки $t = c$ — кінцевої точки інтервалу. Зокрема, отримаємо значення таких величин: $\alpha_0(c)$, $\alpha_1(c)$, $\alpha_2(c)$; $\beta_0(c)$, $\beta_1(c)$, $\beta_2(c)$. При $t = c$ з рівнянь (6.69) і (6.75) матимемо

$$\alpha_0(c)y''(c) - y'(c) + \alpha_1(c)y(c) + \alpha_2(c) = 0; \tag{6.81}$$

$$\beta_0(c)y''(c) + \beta_1(c)y'(c) - y(c) + \beta_2(c) = 0. \tag{6.82}$$

Третє рівняння одержуємо з граничної умови (6.68) у вигляді

$$\gamma_{00}y''(c) - y'(c) + \gamma_{10}y(c) + \gamma_{20} = 0. \tag{6.83}$$

З системи рівнянь знаходимо $y(c)$, $y'(c)$, $y''(c)$, що визначають граничні умови на кінці $t = c$.

Цим завершується прямий хід прогонки.

Використовуючи отримані вище значення функції і двох похідних при $t = c$ як початкові умови і інтегруючи від точки $t = c$ до точки $t = 0$, знаходимо розв'язок крайової задачі (6.65)—(6.68), що буде закінченням зворотного ходу прогонки.

Як приклад розглянемо задачу про деформацію тришарової балки, що наведена в § 6.1.

Крайова задача описується диференціальним рівнянням

$$\psi''' - k^2\psi' + a = 0 \quad (0 \le t \le 1) \tag{6.84}$$

з граничними умовами

$$\psi'(0) = 0, \quad \psi(1/2) = 0, \quad \psi'(1) = 0. \tag{6.85}$$

Порівнюючи (6.84) і (6.85) з (6.65)—(6.68), маємо

$$p = k^2, \quad q = 0, \quad r = -a;$$

$$\alpha_{00} = \alpha_{10} = \alpha_{20} = 0, \quad \beta_{00} = \beta_{10} = \beta_{20} = 0, \quad \gamma_{00} = \gamma_{10} = \gamma_{20} = 0;$$

$$b = 1/2, \quad c = 1.$$

Дві задачі Коші (6.71)—(6.74) і (6.77)—(6.80) набувають такого вигляду:

$$\alpha_0' + \alpha_0\alpha_1 + k^2\alpha_2^0 = 1;$$

$$\alpha_1' + \alpha_1^2 + k^2\alpha_0\alpha_1 = 0;$$

$$\alpha_2' + \alpha_1\alpha_2 + k^2\alpha_0\alpha_2 - a\alpha_0 = 0;$$

$$\alpha_0(0) = \alpha_1(0) = \alpha_2(0) = 0; \qquad (6.86)$$

$$\beta_0' + \beta_1 = 0, \quad \beta_1' + k^2\beta_0 = 1 \quad \beta_2' - a\beta_0 = 0;$$

$$\beta_0(1/2) = \beta_1(1/2) = \beta_2(1/2) = 0. \qquad (6.87)$$

Покладемо $k = 5$, $a = 1$. Розв'язки α_i і β_i задач Коші (6.86) і (6.87) наведені в табл. 15.

Таблиця 15

x	α_0	α_2	β_0	β_1	β_2
0,0	0,0000	0,0000	0,0000	0,0000	0,0000
0,2	0,1523	0,0141	−0,0051	0,1042	−0,0002
0,4	0,1928	0,0294	−0,0217	0,2350	−0,0014
0,6	0,1990	0,0360	−0,0541	0,4259	−0,0050
0,8	0,1999	0,0385	−0,1105	0,7254	−0,0130
1,0	0,2000	0,0395	−0,2053	1,2100	−0,0284

Примітка. Для α_1 усі значення дорівнюють 0,00.

Визначимо всі значення початкових умов при $t = 1$. Систему рівнянь (6.81)—(6.83) для розглядуваної задачі запишемо у вигляді

$$\psi'(1) = 0, \quad \alpha_0(1)\psi''(1) + \alpha_1(1)\psi(1) + \alpha_2(1)) = 0;$$

$$\beta_0(1)\psi''(1) - \psi(1) + \beta_2(1) = 0;$$

звідки

$$\psi(1) = \frac{\alpha_0(1)\beta_2(1) - \beta_0(1)\alpha_2(1)}{\alpha_1(1)\beta_0(1) + \alpha_0(1)};$$

$$\psi'(1) = 0, \quad \psi''(1) = -\frac{\alpha_1(1)\beta_2(1) + \alpha_2(1)}{\alpha_1(1)\beta_0(1) + \alpha_0(1)}. \qquad (6.88)$$

Обчислюючи вирази в правих частинах рівностей (6.88), знаходимо $\psi(1) = 0{,}0121$, $\psi'(1) = 0$, $\psi''(1) = -0{,}1973$.

Одержані значення можна використати як початкові умови для зворотної прогонки для рівняння (6.84) при інтегруванні від $t = 1$ до $t = 0$. Після цього знаходимо розв'язок вихідної крайової задачі, який повністю збігається з наведеним у табл.14 при $k = 5$.

§ 6.3. МЕТОД РІЗНИЦЕВОЇ ПРОГОНКИ

Розглянемо застосування методу скінченних різниць для апроксимації похідних, що містяться в диференціальних рівняннях та граничній умові крайової задачі. Тобто рівняння крайової задачі та граничні умови перетворюються в систему лінійних алгебраїчних рівнянь. Розв'язання цієї системи рівнянь дає значення залежної змінної на дискретній множині незалежної змінної. Для розв'язання одержаної скінченнорізницевої системи рівнянь застосуємо метод різницевої прогонки або метод факторизації.

Лінійне диференціальне рівняння другого порядку зі змінними коефіцієнтами має вигляд

$$y'' + p(t)y' + q(t)y = r(t) \quad (0 \le t \le l), \tag{6.89}$$

з граничними умовами

$$y(0) = \alpha, \quad y(l) = \beta. \tag{6.90}$$

Запишемо рівняння (6.89) в скінченнорізницевій формі. Нехай точки системи визначаються таким чином:

$$t_0 = 0, \quad t_n = t_{n-1} + h, \quad n = 1, 2, \ldots, N,$$

де N — число інтервалів і $t_N = l$.

Значення змінної y і її похідних в точці t_n задаються співвідношеннями

$$y = y_n, \quad y' = \frac{y_{n+1} - y_{n-1}}{2h},$$

$$y'' = \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}. \tag{6.90}$$

На основі одержаних виразів рівняння (6.89) набуває вигляду

$$\frac{1}{h^2}(y_{n+1} - 2y_n + y_{n-1}) + \frac{p(t_n)}{2h}(y_{n+1} - y_{n-1}) + q(t_n)y_n = r(t_n),$$

або

$$a_n y_{n-1} + b_n y_n + c_n y_{n+1} = d_n, \tag{6.91}$$

де

$$a_n = 1 - \frac{h}{2}p(t_n); \quad b_n = h^2 q(t_n) - 2;$$

$$c_n = 1 + \frac{h}{2}p(t_n); \quad d_n = h^2 r(t_n),$$

а граничні умови

$$y_0 = \alpha, \quad y_N = \beta. \tag{6.92}$$

У векторно-матричній формі рівняння (6.91) і (6.92) запишемо у вигляді

$$A\bar{y} = \bar{s}, \tag{6.93}$$

де

$$\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{N-1} \end{bmatrix}, \quad \bar{s} = \begin{bmatrix} s_1 \\ s_2 \\ \dots \\ s_{N-1} \end{bmatrix} = \begin{bmatrix} d_1 - \alpha_1 \\ d_2 \\ \dots \\ d_{N-1} - \beta c_{N-1} \end{bmatrix};$$

$$A = \begin{bmatrix} b_1 & c_1 & & & 0 \\ a_2 & b_2 & c_2 & & \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ & & & a_{N-2} & b_{N-2} \quad c_{N-2} \\ & 0 & & & a_{N-1} \quad b_{N-1} \end{bmatrix}.$$

У матриці A відрізняються від нуля лише діагональні та найближчі до них зліва й справа елементи, тобто вона тридіагональна. Для розв'язання рівняння (6.93) можна використати дуже ефективну процедуру факторизації.

Покладемо, що A — невироджена матриця і може бути факторизована, тобто записана у вигляді добутку двох матриць, а саме:

$$A = LU, \tag{6.94}$$

де

$$L = \begin{bmatrix} \beta_1 & & & & 0 \\ \alpha_2 & \beta_2 & & & \\ & \alpha_3 & \beta_3 & & \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ & & & \alpha_{N-2} & \beta_{N-2} \\ & 0 & & & \alpha_{N-1} \quad \beta_{N-1} \end{bmatrix}; \tag{6.95}$$

$$U = \begin{bmatrix} 1 & \gamma_1 & & & 0 \\ & 1 & \gamma_2 & & \\ & & 1 & \gamma_3 & \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ & & & 1 & \gamma_{N-2} \\ & 0 & & & 1 \end{bmatrix}. \tag{6.96}$$

З рівняння (6.94) знаходимо такі співвідношення:

$$\beta_1 = b_1, \quad \beta_1\gamma_1 = c_1;$$

$$\beta_n = b_n - a_n\gamma_{n-1}, \quad n = 2, 3, \ldots, N-1;$$

$$\beta_n\gamma_n = c_n, \quad n = 2, 3, \ldots, N-2. \qquad (*)$$

Враховуючи (6.94), рівняння (6.93) можна записати у вигляді

$$L\bar{U}\bar{y} = \bar{s}. \qquad (6.97)$$

Поклавши

$$\bar{U}\bar{y} = \bar{z}, \qquad (6.98)$$

перетворимо рівняння (6.97)

$$L\bar{z} = \bar{s}, \qquad (6.99)$$

або

$$\begin{bmatrix} \beta_1 & & & & 0 \\ \alpha_2 & \beta_2 & & & \\ & \alpha_3 & \beta_3 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & \alpha_{N-2} & \beta_{N-2} \\ & 0 & & & \alpha_{N-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \cdots \\ \cdots \\ z_{N-2} \\ z_{N-1} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \cdots \\ \cdots \\ s_{N-2} \\ s_{N-1} \end{bmatrix}. \qquad (6.100)$$

Звідси невідомі компоненти вектора $\bar{z}$ визначаються формулами

$$z_1 = s_1/\beta_1, \quad z_n = (s_n - a_n z_{n-1})/\beta_n, \quad n = 2, 3, \ldots, N-1. \qquad (6.101)$$

Цим завершується прямий хід прогонки.

Після визначення компонент вектора $\bar{z}$, враховуючи, що елементи матриці U вже відомі, можна знайти розв'язок системи рівнянь (6.98). З (6.98) маємо

$$\begin{bmatrix} 1 & \gamma_1 & & & 0 \\ & 1 & \gamma_2 & & \\ & & 1 & \gamma_3 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & 1 & \gamma_{N-2} \\ & 0 & & & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ \cdots \\ y_{N-2} \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \cdots \\ \cdots \\ z_{N-2} \\ z_{N-1} \end{bmatrix}. \qquad (6.102)$$

Звідси знаходимо розв'язок вихідної системи рівнянь (6.93):

$$y_{N-1} = z_{N-1}, \quad y_n = z_n - \gamma_n y_{n+1}, \quad n = N-2, N-3, \ldots, 2, 1. \qquad (6.103)$$

Це зворотний хід прогонки.

Формулами (6.103) визначається аналог диференціального рівняння першого порядку (6.42), яке вводиться при диференціальній прогонці замість рівняння другого порядку.

Розглядається випадок, коли граничні умови (6.90) не містять похідних, але в іншому випадку граничні умови теж можна апроксимувати скінченнорізницевими виразами, й тоді в систему рівнянь (6.93) додаються ще деякі рівняння, що не впливає на процес її розв'язання. Метод різницевої прогонки приводить до стійкого обчислювального процесу і припускає просту реалізацію на комп'ютерах.

Покажемо застосування цього методу на прикладі крайової задачі:

$$y'' - k^2 y = 0 \quad (0 \le t \le b);$$

$$y(0) = 1, \quad y(b) = e^{-kb}. \tag{6.104}$$

Розв'яжемо задачу при $k = 5$, $b = 2$. Маємо:

$$a_n = 1, \ b_n = -(k^2h^2 + 2), \ c_n = 1, \ d_n = 0 \quad (n = 1, 2, \dots, N-1);$$

$$s_1 = -1, \ s_n = 0 \quad (n = 2, 3, \dots, N-2), \ s_{N-1} = -e^{-kb}.$$

За елементами a_n, b_n, c_n отримаємо елементи β_n, γ_n за формулами (*).

Таблиця 16

t	$y(t)$			
	$N = 20$	$N = 40$	$N = 80$	Точний розв'язок
0,1	0,6096	0,6073	0,6067	0,6065
0,2	0,3716	0,3688	0,3681	0,3679
0,3	0,2266	0,2240	0,2234	0,2231
0,4	0,1381	0,1360	0,1355	0,1353
0,5	0,08419	0,08262	0,08222	0,08209
0,6	0,05132	0,05018	0,04988	0,04979
0,7	0,03129	0,03047	0,03027	0,03020
0,8	0,01907	0,01851	0,01836	0,01832
0,9	0,01163	0,01124	0,01114	0,01111
1,0	0,007088	0,006826	0,006760	0,006738
1,1	0,004321	0,004145	0,004101	0,004087
1,2	0,002634	0,002518	0,002488	0,002479
1,3	0,001608	0,001529	0,001510	0,001503
1,4	0,0009787	0,0009285	0,0009160	0,0009119
1,5	0,0005964	0,0005638	0,0005558	0,0005531
1,6	0,0003631	0,0003423	0,0003372	0,0003355
1,7	0,0002207	0,0002077	0,0002045	0,0002035
1,8	0,0001334	0,0001259	0,0001240	0,0001234
1,9	0,00007947	0,00007599	0,00007514	0,00007485

Далі знаходимо компоненти вектора $\bar{z}$ за формулою (6.101), а потім за формулою (6.103) компоненти вектора $\bar{y}$. В табл. 16 наведені результати розв'язання задачі методом різницевої прогонки і точні значення розв'язку.

§ 6.4. МЕТОД ДИСКРЕТНОЇ ОРТОГОНАЛІЗАЦІЇ

Як уже зазначалось в § 6.2, при розв'язанні деяких класів крайових задач для лінійних диференціальних рівнянь виникають труднощі при реалізації обчислювального процесу, які обумовлені жорсткістю диференціальних рівнянь, тобто одночасно мають місце розв'язки рівняння, які ростуть повільно і дуже швидко. У деяких випадках навіть на комп'ютерах з великим числом значущих цифр не можна отримати надійного результату. Для таких крайових задач метод зведення крайових задач до задачі Коші уже не можна застосувати. В цих випадках пропонуються інші підходи, що дозволяють уникнути зазначених труднощів. Деякі методи розв'язання таких задач наведені в § 6.2 і 6.3.

У цьому параграфі наведемо ще один метод, що дозволяє у зазначених випадках побудувати стійкий обчислювальний процес розв'язання крайової задачі у вигляді системи диференціальних рівнянь:

$$\frac{d\bar{y}}{dt} = A(t)\bar{y} + \bar{f}(t), \quad (a \le t \le b) \tag{6.105}$$

з граничними умовами

$$B_1\bar{y}(a) = \bar{b}_1; \tag{6.106}$$

$$B_2\bar{y}(b) = \bar{b}_2, \tag{6.107}$$

де $\bar{y} = \{y_1, y_2, \ldots, y_n\}^T$ — вектор-стовпець; $\bar{f}$ — вектор-стовпець правої частини; $A(t)$ — задана квадратна матриця порядку n; B_1, B_2 — задані прямокутні матриці відповідно порядків $k \times n$ і $(n-k) \times n$ $-(k < n)$; $\bar{b}_1$, $\bar{b}_2$ — задані вектори.

Розглянемо суть методу. Розв'язок крайової задачі (6.105)—(6.107) шукатимемо у вигляді

$$\bar{y}(t) = \sum_{j=1}^{m} C_j y_j(t) + \bar{y}_{m+1}(t), \tag{6.108}$$

де $m = \min(k, n-k)$ (для визначеності покладемо $m = n-k$); $\bar{y}_j$ —розв'язки задач Коші для системи рівнянь (6.105) при $\bar{f} = 0$ з початковими умовами, що задовольняють граничні умови на лівому кінці інтервалу (6.106) при

$\bar{b}_1 = 0$; $\bar{y}_{m+1}$ — розв'язок задачі Коші для системи (6.105) з початковими умовами, що задовольняють граничні умови (6.106); m дорівнює числу граничних умов на правому кінці інтервалу інтегрування.

Метод дискретної ортогоналізації дає можливість одержати стійкий обчислювальний процес за рахунок ортогоналізації векторів-розв'язків задач Коші в скінченному числі точок інтервалу зміни аргументу.

Розіб'ємо весь інтервал $[a, b]$ на малі відрізки точками інтегрування t_s $(s = 0, 1, \dots, N)$ так, що $t_0 = a$, $t_N = b$. Серед цих точок беремо точки ортогоналізації T_i $(i = 0, 1, \dots, M)$. Це залежить від ступеня необхідної точності розв'язку задачі і не залежить від інших вимог.

Нехай у точці T_i будь-яким чисельним методом, наприклад Рунге—Кутта, знайдені розв'язки задач Коші, які позначимо через $\bar{u}_r(T_i)$, де $(r = 1, 2, \dots, m+1)$.

Таким чином, у точці T_i до ортогоналізації маємо вектори

$$\bar{u}_1(T_i),\ \bar{u}_2(T_i),\ \dots,\ \bar{u}_m(T_i),\ \bar{u}_{m+1}(T_i). \tag{6.109}$$

Проортонормуємо вектори $\bar{u}_j(T_i)$ $(j = 1, 2, \dots, m)$ у точці T_i й позначимо їх через

$$\bar{z}_1(T_i),\ \bar{z}_2(T_i),\ \dots,\ \bar{z}_m(T_i). \tag{6.110}$$

Вектори $\bar{z}_i$ виражаються через вектори $\bar{u}_i$ таким чином:

$$\bar{z} = \frac{1}{\omega_{rr}}\left(\bar{u}_r - \sum_{j=1}^{r-1} \omega_{rj}\bar{z}_j\right),\quad r = 1, 2, \dots, m, \tag{6.111}$$

де

$$\omega_{rj} = (\bar{u}_r, \bar{z}_j),\quad j < r;$$

$$\omega_{rr} = \sqrt{(\bar{u}_r, \bar{u}_r) - \sum_{j=1}^{r-1} \omega_{rj}^2}\ .$$

Вектор $\bar{z}_{m+1}$ не нормується й обчислюється за формулою

$$\bar{z}_{m+1} = \bar{u}_{m+1} - \sum_{j=1}^{m} \omega_{m+1\,j}\bar{z}_j;\quad \omega_{m+1\,j} = (\bar{u}_{m+1}, \bar{z}_j). \tag{6.112}$$

Згідно з (6.111) та (6.112) при $t = T_j$ дістанемо

$$\begin{aligned}
\omega_{11}\bar{z}_1 &= \bar{u}_1;\\
\omega_{22}\bar{z}_2 &= \bar{u}_2 - \omega_{21}\bar{z}_1;\\
\omega_{33}\bar{z}_3 &= \bar{u}_3 - \omega_{31}\bar{z}_1 - \omega_{32}\bar{z}_2;\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
\omega_{mm}\bar{z}_m &= \bar{u}_m - \omega_{m1}\bar{z}_1 - \omega_{m2}\bar{z}_2 - \dots - \omega_{m,m-1}\bar{z}_{m-1};\\
1\cdot\bar{z}_{m+1} &= \bar{u}_{m+1} - \omega_{m+1,1}\bar{z}_1 - \omega_{m+1,2}\bar{z}_2 - \dots - \omega_{m+1,m-1}\bar{z}_{m-1} - \omega_{m+1,m}\bar{z}_m.
\end{aligned} \tag{6.113}$$

Після перетворень із (6.113) отримаємо матричну рівність

$$\begin{bmatrix} \bar{u}_1(T_l) \\ \bar{u}_2(T_l) \\ \cdots \\ \cdots \\ \bar{u}_m(T_l) \\ \bar{u}_{m+1}(T_l) \end{bmatrix} = \Omega_l \begin{bmatrix} \bar{z}_1(T_l) \\ \bar{z}_2(T_l) \\ \cdots \\ \cdots \\ \bar{z}_m(T_l) \\ \bar{z}_{m+1}(T_l) \end{bmatrix}; \tag{6.114}$$

$$\Omega_l = \begin{bmatrix} \omega_{11}(T_l) & 0 & 0 & \cdots & 0 \\ \omega_{21}(T_l) & \omega_{22}(T_l) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \omega_{m1}(T_l) & \omega_{m2}(T_l) & \omega_{m3}(T_l) & \cdots & 0 \\ \omega_{m+1,1}(T_l) & \omega_{m+1,2}(T_l) & \omega_{m+1,3}(T_l) & \cdots & 1 \end{bmatrix} \tag{6.115}$$

Вектори $\bar{z}_r(T_l)$ є початковими значеннями задач Коші для однорідної ($r = 1, 2, .., m$) та неоднорідної ($r = m + 1$) систем диференціальних рівнянь (6.105) в інтервалі $T_l \le t \le T_{l+1}$.

У кожній точці ортогоналізації T_l розв'язок системи рівнянь (6.105), що задовольняє граничні умови на лівому кінці інтервалу (6.106), можна записати у вигляді двох виразів:

до ортогоналізації

$$\bar{y}(T_l) = \sum_{j=1}^{m} C_j^{(l-1)} \bar{u}_j(T_l) + \bar{u}_{m+1}(T_l); \tag{6.116}$$

після ортогоналізації

$$\bar{y}(T_l) = \sum_{j=1}^{m} C_j^{(l)} \bar{z}_j(T_l) + \bar{z}_{m+1}(T_l). \tag{6.117}$$

Розв'язок системи рівнянь (6.105) на відрізку $T_l \le t \le T_{l+1}$ можна подати у вигляді

$$\bar{y}(t) = \sum_{j=1}^{m} C_j^{(l)} \bar{z}_j(t) + \bar{z}_{m+1}(t). \tag{6.118}$$

Після виконання інтегрування на останньому відрізку $T_{m-1} \le t \le T_m$ і ортогоналізації в точці T_M за формулою (6.117) маємо

$$\bar{y}(T_M) = \sum_{j=1}^{m} C_j^{(M)} \bar{z}_j(T_M) + \bar{z}_{m+1}(T_M). \tag{6.119}$$

Задовольняючи граничні умови на правому кінці інтервалу інтегрування, тобто підставляючи (6.119) в (6.107), дістанемо систему лінійних алгебраїчних рівнянь для визначення невідомих $C_j^{(M)}$. Після знаходження $C_j^{(M)}$ розв'язок крайової задачі (6.105)—(6.106) в точці $t = T_M$ дається формулою (6.119). На цьому закінчується прямий хід розв'язання задачі.

При зворотному ході за значеннями сталих $C_j^{(i)}$ $(j = 1, 2, \ldots, m)$ визначаються сталі $C_j^{(i-1)}$, визнаючи $i = M$. Для цього прирівняємо праві частини виразів (6.116) і (6.117):

$$\sum_{j=1}^{m} C_j^{(i-1)} \bar{u}_j(T_i) + \bar{u}_{m+1}(T_i) = \sum_{j=1}^{m} C_j^{(i)} \bar{z}_j(T_i) + \bar{z}_{m+1}(T_i). \qquad (6.120)$$

Підставляючи замість $\bar{u}_j$ їхні значення з (6.114), при $t = T_i$ маємо

$$C_1^{(i-1)}\omega_{11}\bar{z}_1 + C_2^{(i-1)}(\omega_{21}\bar{z}_1 + \omega_{22}\bar{z}_2) + C_3^{(i-1)}(\omega_{31}\bar{z}_1 + \omega_{32}\bar{z}_2 + \omega_{33}\bar{z}_3) +$$

$$+ \ldots + C_m^{(i-1)}(\omega_{m1}\bar{z}_1 + \omega_{m2}\bar{z}_2 + \ldots + \omega_{mm}\bar{z}_m) + 1 \cdot (\omega_{m+1,1}\bar{z}_1 + \omega_{m+1,2}\bar{z}_2 + \ldots +$$

$$+ \omega_{m+1,m}\bar{z}_m + 1 \cdot \bar{z}_{m+1}) = C_1^{(i)}\bar{z}_1 + C_2^{(i)}\bar{z}_2 + \ldots + C_m^{(i)}\bar{z}_m + 1 \cdot \bar{z}_{m+1}.$$

Прирівнюючи коефіцієнти при векторах $\bar{z}_j$ $(j = 1, 2, \ldots, m+1)$ у (6.121), знаходимо

$$\Omega^{\mathrm{T}} C^{(i-1)} = \bar{C}^{(i)}, \quad i = 1, 2, \ldots, M \qquad (6.122)$$

або

$$\bar{C}^{(i-1)} = [\Omega_i^{\mathrm{T}}]^{-1} \bar{C}^{(i)},$$

де Ω_i^{T} — транспонована матриця (6.115); $\bar{C}^{(i)}$ — вектор-стовпчик з компонентами $C_1^{(i)}, C_2^{(i)}, \ldots, C_m^{(i)}, 1$.

Таким чином, за допомогою рівності (6.122) можна знайти значення сталих $C_j^{(i)}$ в усіх точках, починаючи з $i = M$. За формулою (6.118) обчислюються розв'язки $\bar{y}$ крайової задачі.

При реалізації даного алгоритму на ЕОМ необхідно зберігати інформацію про матриці Ω_i і вектори $\bar{z}_i$ $(i = 1, 2, \ldots, m+1)$.

Практично одержана інформація в усіх точках ортогоналізації звичайно не використовується, а при розв'язанні задач обмежується тільки значеннями шуканих функцій у ряді точок — так званих точок видачі результатів, яких часто значно менше, ніж точок ортогоналізації.

Використовуючи такий підхід, можна суттєво скоротити обсяг результатів.

З рівності (6.122) можна отримати, що

$$\Omega_i^{\mathrm{T}} \Omega_{i-1}^{\mathrm{T}} \ldots \Omega_{i-(p-1)}^{\mathrm{T}} \bar{C}^{(i-p)} = \bar{C}^{(i)} \qquad (6.123)$$

або

$$\left(\prod_{j=0}^{p-1} \Omega_{i-j}^{\mathrm{T}} \right) \bar{C}^{(i-p)} = \bar{C}^{(i)}, \qquad (6.124)$$

звідки

$$\overline{C}^{(l-p)} = \left(\prod_{j=0}^{p-1} \Omega_{l-j}^{\mathrm{T}}\right)^{-1} \overline{C}^{(l)}. \qquad (6.125)$$

Таким чином, для знаходження вектора $C^{(l-p)}$ необхідно зберігати

Таблиця 17

t	$y(t)$		
	$\Delta t = 0{,}1$	$\Delta t = 0{,}01$	Точний розв'язок
0,1	0,6068	0,6065	0,6065
0,2	0,3682	0,3679	0,3679
0,3	0,2234	0,2231	0,2231
0,4	0,1355	0,1353	0,1353
0,5	0,08225	0,08208	0,08209
0,6	0,4991	0,4979	0,4979
0,7	0,03028	0,03020	0,03020
0,8	0,01837	0,01832	0,01832
0,9	0,01115	0,01111	0,01111
1,0	0,006705	0,006738	0,006738
1,1	0,004105	0,004087	0,004087
1,2	0,002491	0,002479	0,002479
1,3	0,001511	0,001503	0,001503
1,4	0,0009169	0,0009119	0,0009119
1,5	0,0005563	0,0005531	0,0005531
1,6	0,0003375	0,0003355	0,0003355
1,7	0,0002048	0,0002035	0,0002035
1,8	0,0001242	0,0001234	0,0001234
1,9	0,00007520	0,00007485	0,00007485
2,0			

інформацію про добуток матриць $\prod_{j=0}^{p-1} \Omega_{l-j}^{\mathrm{T}}$, що дає значну економію.

Побудову задачі Коші для векторів $\overline{y}_j(t)$ $(j = 1, 2, \ldots, m, m+1)$, де перші m розв'язків задовольняють однорідні рівняння (6.105) і однорідні граничні умови (6.106), а розв'язок $\overline{y}_{m+1}$ — неоднорідне рівняння (6.105) і неоднорідні граничні умови (6.106) на лівому кінці інтервалу інтегрування подано в § 6.1. Продемонструємо метод дискретної ортогоналізації на прикладі.

Розглянемо крайову задачу

$$y'' - k^2 y = 0 \quad (0 \le t \le b) \qquad (6.126)$$

з граничними умовами

$$y(0) = 1, \quad y(b) = e^{-kb}. \tag{6.127}$$

Результати розв'язання задачі при $k = 10$, $b = 3$ наведені в табл.17.

§ 6.5. МЕТОДИ РІТЦА І БУБНОВА—ГАЛЬОРКІНА

Розглянемо варіаційні задачі, які зв'язані з лінійними звичайними диференціальними рівняннями, і з'ясуємо основні факти і методи застосування варіаційного числення до розв'язання крайових задач.

Розв'яжемо крайову задачу для диференціального рівняння

$$L(y) = \frac{d}{dx}(py') - qy = f \quad (0 \le x \le l) \tag{6.128}$$

з граничними умовами

$$y(0) = y_0, \quad y(l) = y_1. \tag{6.129}$$

У варіаційному численні показано, що крайовій задачі (6.128), (6.129) можна поставити у відповідність задачу про мінімум інтеграла

$$I(y) = \int_0^1 [p(y')^2 + qy^2 + 2fy]dx \tag{6.130}$$

при тих же граничних умовах (6.129).

Функція $y(x)$, яка задовольняє умови (6.129) і дає мінімум інтеграла (6.130), і є розв'язком крайової задачі (6.128), (6.129).

Метод Рітца полягає в тому, що вибирається функція y у вигляді

$$y = \Phi(x, a_1, a_2, \dots, a_n), \tag{6.131}$$

тобто ця функція залежить від n параметрів. Підставивши функцію (6.131) у вираз (6.130) і виконавши диференціювання та інтегрування, дістанемо функцію n змінних, тобто

$$I(y) = I(a_1, a_2, \dots, a_n). \tag{6.132}$$

Для знаходження мінімуму цієї функції n змінних необхідно задовольнити умови

$$\frac{\partial I}{\partial a_k} = 0 \quad (k = 1, 2, \dots, n). \tag{6.133}$$

Розв'язавши цю систему рівнянь, знаходимо значення параметрів a_k і після підстановки їх у вираз (6.131) отримаємо шуканий розв'язок.

Функцію $\Phi(x, a_1, a_2, \dots, a_n)$, можна шукати у вигляді

$$\Phi = \varphi_0(x) + \sum_{k=1}^{n} a_k \varphi_k(x), \qquad (6.134)$$

де $\varphi_0(x)$ задовольняє граничні умови.

Метод Бубнова—Гальоркіна полягає в тому, що розв'язок крайової задачі (6.128), (6.129) шукатимемо у вигляді

$$y = \varphi_0(x) + \sum_{k=1}^{n} c_k \varphi_k(x), \qquad (6.135)$$

де $\varphi_0(x)$ задовольняє граничні умови (6.129), $\varphi_k(x)$ — задані базисні функції. Підставимо вираз (6.135) у диференціальне рівняння (6.128) і отримаємо відхил

$$\varepsilon(x, c_1, c_2, \ldots, c_n) \equiv L(y) - f = L(\varphi_0) + \sum_{k=1}^{n} c_k L(\varphi_k) - f(x).$$

Для того щоб функція y з виразу (6.135) була точним розв'язком крайової задачі (6.128), (6.129), треба, щоб відхил ε дорівнював тотожно нулю. Ця вимога рівносильна вимозі, щоб відхил $\varepsilon = L(y) - \varphi$ був ортогональним до всіх базисних функцій $\varphi_k(x)$ $(k = 1, 2, \ldots, n, \ldots)$. Але ж коли у виразі (6.135) для y є тільки n сталих $c_1, c_2, \ldots, c_n$, можна задовольнити тільки n умов ортогональності відхилу ε до базисних функцій $\varphi_k(x)$ $(k = 1, 2, \ldots, n, \ldots)$, тобто маємо

$$\int_0^l [L(y(x)) - f(x)]\varphi_k(x)dx = 0 \quad (k = 1, 2, \ldots, n). \qquad (6.136)$$

З цієї системи рівнянь знаходимо коефіцієнти c_k $(k = 1, 2, \ldots, n)$, і, підставляючи їх у вираз (6.135), отримуємо шуканий розв'язок.

На прикладі крайової задачі (6.128), (6.129) розглянемо, в якому співвідношенні знаходяться метод Рітца і метод Бубнова—Гальоркіна. Для зручності вважатимемо, що граничні умови одержані, тобто

$$y(0) = y(b) = 0. \qquad (6.137)$$

Цього можна досягти за рахунок заміни функцій

$$y = z + \frac{x}{l} y_l + \frac{l - x}{l} y_0.$$

Вважаємо, що в інтервалі $0 \le x \le l$ мають місце нерівності $p(x) > 0$, $q(x) \ge 0$. Розв'язок крайової задачі (6.128), (6.137) шукаємо у вигляді

$$y_n(x) = \sum_{k=1}^{n} a_k \varphi_k(x). \qquad (6.138)$$

Після підставлення $y_n(x)$ у вираз для інтервалу (6.130) отримуємо

$$I(y_n) = \int_0^l [p(y_n')^2 + qy_n^2 + 2fy_n]dx =$$

$$= \int_0^l \left[p\left(\sum_{k=1}^n a_k\varphi_k'\right)^2 + q\left(\sum_{k=1}^n a_k\varphi_k\right)^2 + 2f\sum_{k=1}^n a_k\varphi_k \right] dx = \sum_{k,s=1}^n \alpha_{k,s}a_k a_s +$$

$$+ 2\sum_{k=1}^n \beta_k a_k, \qquad (6.139)$$

де

$$\alpha_{k,s} = \alpha_{s,k} = \int_0^l (p\psi_k'\psi_s' + q\psi_k\psi_s)dx; \quad \beta_k = \int_0^l f\psi_k dx. \qquad (6.140)$$

Після диференціювання по a_s виразу (6.139) отримуємо систему рівнянь для визначення сталих a_k у вигляді

$$\frac{1}{2}\frac{\partial I(y_n)}{\partial a_s} = \int_0^l (py_n'\psi_s' + qy_n\psi_s + f\psi_s)dx = 0 \quad (s = 1, 2, \dots, n) \qquad (6.141)$$

або у вигляді

$$\sum_{k=1}^n \alpha_{k,s}a_k + \beta_s = 0 \quad (s = 1, 2, \dots, n). \qquad (6.142)$$

Розв'язок системи рівнянь (6.142) позначимо через $a_k^{(n)}$ $(k = 1, 2, \dots, n)$, тоді наближений розв'язок за методом Рітца в n-му наближенні має вигляд

$$y_n(x) = \sum_{k=1}^n a_k^{(n)}\varphi_k(x). \qquad (6.143)$$

Зазначимо, що систему рівнянь (6.142) можна записати в іншому вигляді, який в деяких випадках полегшує її складання. Інтегруючи перший член рівностей (6.141) за частинами і беручи до уваги, що φ_k і y_n дорівнюють нулю на кінцях інтервалу, знаходимо:

$$\int_0^l py_n'\psi_s'dx = [py_n'\psi_s]\Big|_0^l - \int_0^l (py_n')'\psi_s dx = -\int_0^l (py_n')'\psi_s dx.$$

Підставляючи цей вираз у (6.141) замість першого члена, змінюючи знак і враховуючи вигляд рівняння (6.128), маємо

$$\int_0^l [(py_n')' - qy_n - f]\psi_s dx = \int_0^l [L(y_n) - f]\psi_s dx = 0$$

$$(s = 1, 2, \ldots, n). \qquad (6.144)$$

У такому записі отримані рівняння мають вигляд рівнянь методу Бубнова—Гальоркіна (6.136).

Таким чином, для даної крайової задачі застосування методу Бубнова—Гальоркіна для її розв'язання приводить до одного й того ж наближеного розв'язку. Цим підтверджується зв'язок методу Бубнова—Гальоркіна з варіаційними методами.

Приклад 1. Знайдемо розв'язок крайової задачі

$$y'' + y + x = 0 \quad (0 \le x \le 1);$$

$$y(0) = y(1) = 0.$$

Розв'язання: Варіаційна задача зводиться до знаходження лінійного інтеграла

$$I(y) = \int_0^1 [(y')^2 - y^2 - 2xy]dx.$$

Згідно з методом Рітца і Бубнова—Гальоркіна наближений розв'язок шукаємо у вигляді

$$y_n(x) = x(1 - x)\sum_{k=1}^{n} a_k x^{k-1}.$$

Використовуючи рівняння (6.144), отримуємо в першому і другому наближенні:

$$y_1 = \frac{5}{18}x(1 - x), \quad y_2 = x(1 - x)\left(\frac{71}{369} + \frac{7}{41}x\right).$$

Таблиця 18

x	y_1	y_2	y_T
0,25	0,052	0,044	0,044
0,50	0,069	0,069	0,070
0,75	0,052	0,060	0,060

Точний розв'язок цієї задачі має вигляд

$$y = \frac{\sin(x)}{\sin(1)} - x.$$

Значення одержаних розв'язків в деяких точках наведено в табл. 18.

Приклад 2. Розглянемо крайову задачу для рівняння четвертого порядку:

$$[(x + 2l)y'']'' + qy - kx = 0 \quad (0 <= x <= l)$$

з граничними умовами

$$(x + 2l)y'' = 0, \quad [(x + 2l)y'']' = 0 \text{ при } x = 0;$$

$$y = 0, \ y' = 0 \text{ при } x = l.$$

Розв'язання. Знайдемо розв'язок задачі при $k = l = q = 1$. Базисні функції виберемо у вигляді

$$\psi_1(x) = (x - l)^2(x^2 + 2lx + 3l^2), \quad \psi_2(x) = (x - l)^3(3x^2 + 4lx + 3l^2).$$

При даних значеннях k, l, q знаходимо перше й друге наближення розв'язку:

$$y_1(x) = 0{,}011917(x - 1)^2(x^2 + 2x + 3);$$

$$y_2(x) = 0{,}013743(x - 1)^2(x^2 + 2x + 3) + 0{,}002279(x - 1)^3(3x^2 + 4x + 3).$$

У точці $x = 1/2$ маємо: $y_1(1/2) = 0{,}012662$; $y_2(1/2) = 0{,}012964$.

§ 6.6. МЕТОД СПЛАЙН-КОЛОКАЦІЇ

Метод колокації розв'язання крайових задач для звичайних диференціальних рівнянь базується на використанні апарату наближення многочленами. Але в силу складності реалізації та не зовсім задовільних апроксимаційних властивостей многочленів такий підхід не знайшов практичного застосування.

Метод сплайн-колокації на відміну від попереднього базується на апроксимації сплайнами, що дозволяє побудувати алгоритми, чисельне розв'язання яких не складніше за різницеві схеми. Принципова відзнака цього підходу від різницевих методів полягає в тому, що наближений розв'язок знаходиться у вигляді сплайна на всьому інтервалі інтегрування, а не на дискретній множині точок як в різницевих методах. Це дозволяє одержати більш повну інформацію про розв'язок задачі.

Викладемо метод сплайн-колокації на прикладі крайової задачі для диференціального рівняння другого порядку. Розглянемо крайову задачу для рівняння

$$L[y(x)] = y''(x) + p(x)y'(x) + q(x)y(x) = f(x) \quad (a \le x \le b) \quad (6.145)$$

з граничними умовами

$$\alpha_1 y(a) + \beta_1 y'(a) = \gamma_1, \quad \alpha_2 y(b) + \beta_2 y'(b) = \gamma_2. \quad (6.146)$$

Введемо на $[a, b]$ сітку Δ: $a = x_0 < x_1 < \ldots < x_N = b$. Шукаємо розв'язок крайової задачі (6.145), (6.146) у вигляді кубічного сплайна $S(x)$ на сітці Δ (гл. 1, § 1.6). Вимагатимемо, щоб сплайн $S(x)$ задовольняв рівняння (6.145) в точках $\xi_k \in [a, b]$, $k = 0, 1, \ldots, N$ (точки колокації) і граничні умови (6.146). Маємо

$$L[S(\xi_k)] = S''(\xi_k) + p(\xi_k)S'(\xi_k) + q(\xi_k)S(\xi_k) = f(\xi_k)$$
$$(k = 0, 1, \ldots, N); \qquad (6.147)$$

$$\alpha_1 S(a) + \beta_1 S'(a) = \gamma_1, \quad \alpha_2 S(b) + \beta_2 S'(b) = \gamma_2. \qquad (6.148)$$

Співвідношення (6.147), (6.148) утворюють систему алгебраїчних рівнянь відносно параметрів сплайна. Точки ξ_k називаються *вузлами колокації*, і їх число визначається розмірністю простору кубічних сплайнів $N + 3$. Оскільки $S(x)$ задовольняє двом граничним умовам, то число точок колокації дорівнює $N + 1$. Точки колокації можуть вибиратися довільним чином. На даному відрізку $[x_i, x_{i+1}]$ не повинно бути більше трьох вузлів колокації, бо в противному разі сплайн $S(x)$ визначався би незалежно від інших відрізків і в тому числі граничних умов. Вважаємо, що точки колокації упорядковані, тобто $\xi_0 < \xi_1 < \ldots < \xi_N$. Система рівнянь (6.147), (6.148) залежить від способу запису сплайна $S(x)$ і від вибору колокації.

Для реалізації методу сплайн-колокації доцільно застосувати базисні сплайни, які позначають як B-сплайни. У зв'язку з цим наведемо деякі відомості про B-сплайни.

Розширимо сітку Δ: $a = x_0 < x_1 < \ldots < x_N = b$ допоміжними точками $x_{-m} < \ldots < x_{-1} < a$, $b < x_{N+1} < \ldots < x_{N+m}$ і будемо розглядати сітку Δ_1: $x_{-m} < \ldots < x_{-1} < x_0 < x_1 < \ldots < x_N < x_{N+1} < \ldots < x_{N+m}$.

Розглянемо функцію $\varphi_m(x, t) = (-1)^{m+1}(m+1)(x-t)_+^m$ і побудуємо поділену різницю $(m+1)$-го порядку за значеннями аргументу $t = x_i, \ldots, x_{i+m+1}$. В результаті отримуємо функції змінної x_i

$$\tilde{B}_m^i = \varphi_m[x, x_i, \ldots, x_{i+m+1}] \quad (i = -m, \ldots, N-1). \qquad (6.149)$$

Ці функції називаються *B-сплайнами* степеня m і є сплайнами степеня m дефекту 1 на розширеній сітці Δ_1. З

$$(x-t)_+^m = (x-t)^m + (-1)^{m+1}(t-x)_+^m$$

можна одержати інший вигляд запису (6.149):

$$\tilde{B}_m^i(x) = (m+1)\sum_{p=1}^{i+m+1} \frac{(x_p - x)_+^m}{\omega'_{m+1,i}(x_p)} \quad (i = -m, \ldots, N-1), \qquad (6.150)$$

де

$$\omega_{m+1,i}(t) = \prod_{j=1}^{i+m+1} (t - x_j).$$

При практичних обчислюваннях зручно використовувати не самі B-сплайни, а нормалізовані B-сплайни, які мають вигляд

$$B_m^i(x) = \frac{x_{i+m+1} - x_i}{m+1} \tilde{B}_m^i(x). \qquad (6.151)$$

Для нормальних B-сплайнів існує рекурентна формула

$$B_m^l(x) = \frac{x - x_l}{x_{l+m} - x_l} B_{m-1}^l(x) + \frac{x_{l+m+1} - x}{x_{l+m+1} - x_{l+1}} B_{m-1}^{l+1}(x), \qquad (6.152)$$

яка може використовуватися як визначення B-сплайнів. При цьому

$$B_0^l(x) = \begin{cases} 1 & \text{для } x \in [x_l, x_{l+1}), \\ 0 & \text{для } x \notin [x_l, x_{l+1}). \end{cases}$$

Функції $B_m^l(x)$ є сплайнами степеня m дефекту 1 із скінченними носіями мінімальної довжини. Крім цього, система функцій $B_m^l(x)$ $(l = -m, \ldots, N-1)$ є лінійно незалежна і утворює базис у просторі сплайнів $S_m(\Delta)$. Це означає, що кожний сплайн $S_m(x) \in S_m(\Delta)$ може бути єдиним способом записаний у вигляді

$$S_m(x) = \sum_{l=-m}^{N-1} b_l B_m^l(x), \qquad (6.153)$$

де b_l — деякі сталі коефіцієнти.

Сплайни $B_m^l(x)$ мають такі властивості:

а) $B_m^l(x) = \begin{cases} > 0 & \text{для } x \in [x_l, x_{l+1}); \\ \equiv 0 & \text{для } x \notin [x_l, x_{l+1}); \end{cases}$

b) $\int_{-\infty}^{\infty} B_m^l(x)dx = \frac{x_{l+m+1} - x_l}{m+1}$.

Розглянемо розширену сітку Δ': $x_{-m} < \ldots < x_{-1} < x_0 < x_1 < \ldots < x_N < x_{N+1} < \ldots < x_{N+m}$ $(x_{k+1} - x_k = h = \text{const})$.

Побудуємо перші три B-сплайни непарного степеня. При цьому нумерувати їх будемо по середньому вузлу носіїв. B-сплайни непарного степеня позначимо через $B_m^l(x)$ замість $B_m^{l-\frac{m+1}{2}}(x)$, тобто нумерація сплайнів зсувається на $\frac{m+1}{2}$ одиниць вправо.

Таким чином, B-сплайни першого степеня мають вигляд

$$B_1^l(x) = \begin{cases} 0 & \text{при } -\infty \le x < x_{l-1}; \\ t & \text{при } x_{l-1} \le x < x_l; \\ 1 - t & \text{при } x_l \le x < x_{l+1}; \\ 0 & \text{при } x_l \le x < \infty; \end{cases} \qquad (6.154)$$

B-сплайни третього степеня:

$$B_3^i(x) = \frac{1}{6}\begin{cases} 0 & \text{при} \quad -\infty \le x < x_{i-2}; \\ t^3 & \text{при} \quad x_{i-2} \le x < x_{i-1}; \\ -3t^3 + 3t^2 + 3t^1 + 1 & \text{при} \quad x_{i-1} \le x < x_i; \\ 3t^3 - 6t^2 + 4 & \text{при} \quad x_i \le x < x_{i+1}; \\ (1-t)^3 & \text{при} \quad x_{i+1} \le x < x_{i+2}; \\ 0 & \text{при} \quad x_{i+2} \le x < \infty; \end{cases} \tag{6.155}$$

B-сплайни п'ятого степеня:

$$B_5^i(x) =$$

$$= \frac{1}{120}\begin{cases} 0 & \text{при} \quad -\infty \le x < x_{i-3}; \\ t^5 & \text{при} \quad x_{i-3} \le x < x_{i-2}; \\ -5t^5 + 5t^4 + 10t^3 + 10t^2 + 5t + 1 & \text{при} \quad x_{i-2} \le x < x_{i-1}; \\ 10t^5 - 20t^4 - 20t^3 + 20t^2 + 50t + 26 & \text{при} \quad x_{i-1} \le x < x_i; \\ -10t^5 + 30t^4 - 60t^2 + 66 & \text{при} \quad x_i \le x < x_{i+1}; \\ 5t^5 - 20t^4 + 20t^3 + 20t^2 - 50t + 26 & \text{при} \quad x_{i+1} \le x < x_{i+2} \\ (1-t)^5 & \text{при} \quad x_{i+2} \le x < x_{i+3}; \\ 0 & \text{при} \quad x_{i+3} \le x < \infty, \end{cases}$$

(6.156)

де $t = \dfrac{x - x_k}{h}$ на інтервалі $[x_k, x_{k+1}]$, $k = \overline{i - \dfrac{m+1}{2}, \; i + \dfrac{m+1}{2} - 1}$;

$i = \overline{-\dfrac{m+1}{2} + 1, \; N + \dfrac{m+1}{2} - 1}$, $m = 1, 3, 5$.

Значення сплайнів $B_3^i(x)$ і $B_5^i(x)$ та їхні похідні у вузлах, що належать до їх носіїв, наведені в табл. 19, 20.

Таблиця 19

x	$B_3^i(x)$	$B_3^{i\prime}(x)$	$B_3^{i\prime\prime}(x)$
x_{i-2}	0	0	0
x_{i-1}	$\frac{1}{6}$	$\frac{1}{2h}$	$\frac{1}{h^2}$
x_i	$\frac{4}{6}$	0	$-\frac{2}{h^2}$
x_{i+1}	$\frac{1}{6}$	$-\frac{1}{2h}$	$\frac{1}{h^2}$
x_{i+2}	0	0	0

Таблиця 20

x	$B_5^i(x)$	$B_5^{i\prime}(x)$	$B_5^{i\prime\prime}(x)$	$B_5^{i\prime\prime\prime}(x)$	$B_5^{IV}(x)$
x_{i-3}	0	0	0	0	0
x_{i-2}	$\frac{1}{120}$	$\frac{1}{24h}$	$\frac{1}{6h^2}$	$\frac{1}{2h^3}$	$\frac{1}{h^4}$
x_{i-1}	$\frac{26}{120}$	$\frac{10}{24h}$	$\frac{2}{6h^2}$	$-\frac{1}{h^3}$	$-\frac{4}{h^4}$
x_i	$\frac{66}{120}$	0	$-\frac{1}{h^2}$	0	$\frac{6}{h^4}$
x_{i+1}	$\frac{26}{120}$	$-\frac{10}{24h}$	$\frac{2}{6h^2}$	$\frac{1}{h^3}$	$-\frac{4}{h^4}$
x_{i+2}	$\frac{1}{120}$	$-\frac{1}{24h}$	$\frac{1}{6h^2}$	$-\frac{1}{2h^3}$	$\frac{1}{h^4}$
x_{i+3}	0	0	0	0	0

Розглянемо метод сплайн-колокації з використанням B-сплайнів. Запишемо кубічний сплайн у вигляді ряду по кубічних сплайнах:

$$S(x) = \sum_{i=-1}^{N+1} b_i B_i(x), \quad [x_0, x_N] \tag{6.157}$$

(для спрощення запису опущено значок $m = 3$, а i знаходиться в індексі).

Зупинимось на випадку, коли вузли колокації ξ_i збігаються з вузлами сплайна x_i. Підставляючи (6.157) у (6.147), дістанемо

$$b_{i-1}L[B_{i-1}(x_i)] + b_iL[B_i(x_i)] + b_{i+1}L[B_{i+1}(x_i)] = f_i$$

$$(i = 0, 1, 2, \ldots, N).$$

Після диференціювання ці рівняння можна записати у вигляді

$$b_{i-1}A_i + b_iC_i + b_{i+1}B_i = D_i \quad (i = 0, 1, 2, \ldots, N), \tag{6.158}$$

де

$$A_i = \frac{1}{x_{i+1} - x_{i-1}} \left(1 - \frac{1}{2} p_i h_i + \frac{1}{6} q_i h_i^2\right);$$

$$B_i = \frac{1}{x_{i+1} - x_{i-1}} \left(1 + \frac{1}{2} p_i h_{i-1} + \frac{1}{6} q_i h_{i-1}^2\right);$$

$$C_i = -A_i - B_i + \frac{1}{6} q_i (h_i + h_{i-1});$$

$$D_i = \frac{1}{6} f_i (h_i + h_{i-1}).$$

З рівнянь (6.146), покладаючи $h_{-j} = h_{j-1}$, $h_{N-1+j} = h_{N-j}$, $j = 1, 2, 3$, отримуємо

$$b_{-1}A_{-1} + b_0C_{-1} + b_1B_{-1} = D_{-1};$$

$$b_{N-1}A_{N+1} + b_NC_{N+1} + b_{N+1}B_{N+1} = D_{N+1}, \tag{6.159}$$

де

$$A_{-1} = \alpha_1 h_0 - 3\beta_1, \quad B_{-1} = \alpha_1 h_0 + 3\beta_1, \quad C_{-1} = 2\alpha_1(h_1 + h_0);$$

$$A_{N+1} = \alpha_2 h_{N-1} - 3\beta_2, \quad B_{N+1} = \alpha_2 h_{N-1} + 3\beta_2, \quad C_{N+1} = 2\alpha_2(h_{N-2} + h_{N-1});$$

$$D_{-1} = 2\gamma_1(h_1 + 2h_0), \quad D_{N-1} = 2\gamma_2(h_{N-2} + 2h_{N-1}).$$

Рівняння (6.158), (6.159) утворюють систему $N + 3$ рівнянь відносно $N + 3$ невідомих b_i. Виключаючи за допомогою рівнянь (6.159) невідомі b_{-1} і b_{N-1}, отримаємо систему з тридіагональною матрицею:

$$b_0\bar{C}_0 + b_1\bar{B}_0 = \bar{D}_0;$$

$$b_{i-1}A_{i-1} + b_iC_i + b_{i+1}B_i = D_i, \quad i = 1, 2, \ldots, N-1;$$

$$\tilde{b}_{N-1}\tilde{A}_N + b_N\tilde{C}_N = \tilde{D}_N, \tag{6.160}$$

де

$$\tilde{C}_0 = C_0 - \frac{C_{-1}A_0}{A_{-1}}, \quad \tilde{B}_0 = B_0 - \frac{B_{-1}A_0}{A_{-1}}, \quad \tilde{D}_0 = D_0 - \frac{D_{-1}A_0}{A_{-1}};$$

$$\tilde{A}_N = A_N - \frac{A_{N+1}B_N}{B_{N+1}}, \quad \tilde{C}_N = C_N - \frac{C_{N+1}B_N}{B_{N+1}}, \quad \tilde{D}_N = D_N - \frac{D_{N+1}B_N}{B_{N+1}}.$$

Таким чином, з системи рівнянь (6.160) знаходимо $b_0, b_1, \ldots, b_N$, а потім з рівнянь (6.159) — b_{-1} і b_{N-1}. Підставивши ці коефіцієнти у вираз (6.157), знаходимо розв'язок крайової задачі (6.145), (6.146).

Аналогічно можна за допомогою методу сплайн-колокації розв'язати крайову задачу для диференціального рівняння четвертого порядку, використовуючи при цьому B-сплайни п'ятого порядку.

Як приклад, наведемо результати розв'язання крайової задачі про згин пластини сталої товщини, дві протилежні сторони якої шарнірно оперті, а дві інші ($x = \text{const}$, $y = \text{const}$) шарнірно закріплені під дією нормального навантаження $q = q_1(x) \sin \pi y$ $(0 \le y \le b,\ 0 \le x \le a)$.

Задача описується рівнянням

$$w^{(IV)} - 2(\pi/6)^2 w'' + (\pi/6)^4 w = q/D_m$$

$$D_m = \frac{Eh^3}{12(1-\nu^2)}$$

Таблиця 21

i	w			w''		
	Розв'язання в сплайнах		Точний розв'язок	Розв'язання в сплайнах		Точний розв'язок
	$N = 16$	$N = 32$		$N = 16$	$N = 32$	
0	0	0	0	810,3	806,1	806,2
1	1,339	1,333	1,333	462,8	460,5	460,8
2	4,516	4,493	4,495	205,9	204,9	205,1
3	8,519	8,477	8,481	18,28	18,19	18,39
4	12,61	12,55	12,56	−116,2	−115,5	−115,5
5	16,26	16,18	16,19	−209,5	−208,4	−208,7
6	19,10	19,01	19,02	−270,4	−269,1	−269,3
7	20,90	20,80	20,81	−304,7	−303,5	−303,5
8	21,51	21,41	21,42	−315,8	−314,2	−314,5

з граничними умовами $w = w' = 0$ при $x = 0; a$.

Розв'язання задачі одержано методом сплайн-колокації з використанням B-сплайнів п'ятого порядку (6.156).

Розв'язок задачі отримано при таких даних: $a = b = 1$; $h = 0{,}1$; $E = 1$; $\nu = 0{,}3$; число точок колокації $N = 16; 32$; $t = 16x$; $q_1 = \pi/4$.

Результати розв'язання задачі для w і w' наведено в табл. 21 (в силу симетрії результати дано для $0 \leq x \leq 0{,}5$).

ГЛАВА 7

МЕТОДИ РОЗВ'ЯЗАННЯ НЕЛІНІЙНИХ КРАЙОВИХ ЗАДАЧ ДЛЯ ЗВИЧАЙНИХ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ

§ 7.1. ВСТУПНІ ЗАУВАЖЕННЯ. ПОСТАНОВКА ЗАДАЧІ

Багато математичних моделей, які описують фізичні або технічні задачі, зводяться до нелінійних диференціальних рівнянь, і лише за рахунок нехтування деякими факторами їх можна лінеаризувати. При певних обмеженнях деякі з цих задач можна описати безпосередньо нелінійними звичайними диференціальними рівняннями з відповідними граничними умовами. До нелінійних звичайних диференціальних рівнянь можуть бути зведені тим чи іншим способом зниження розмірності й більш складні задачі, які описуються нелінійними диференціальними рівняннями в частинних похідних.

Розглянемо деякі методи розв'язання нелінійних крайових задач для систем звичайних диференціальних рівнянь у загальному випадку з нелінійними граничними умовами, які ілюструються прикладами розв'язання конкретних задач.

У загальному випадку нелінійну крайову задачу для системи звичайних диференціальних рівнянь можна записати у вигляді

$$\frac{d\bar{y}}{dt} = \bar{f}(t, \bar{y}), \quad t_0 \le t \le t_l; \tag{7.1}$$

$$\bar{g}(\bar{y}_0, \bar{y}_l) = 0, \tag{7.2}$$

де вектор-функція $\bar{f}(t, \bar{y})$ і вектор $\bar{g}(\bar{y}_0, \bar{y}_l)$ нелінійні. Припускаємо, що функції $\bar{f}(t, \bar{y})$ і $\bar{g}(\bar{y}_0, \bar{y}_l)$ достатньо гладкі за всіма аргументами.

Скорочені співвідношення (7.1) і (7.2) мають такий вираз:

$$\frac{dy_1}{dt} = f_1(t, y_1, y_2, \dots, y_n);$$

$$\frac{dy_2}{dt} = f_2(t, y_1, y_2, \dots, y_n);$$

$$\dots\dots\dots\dots\dots\dots$$

$$\frac{dy_n}{dt} = f(t, y_1, y_2, \dots, y_n); \tag{7.1'}$$

$$g_1(y_{01}, y_{02}, \dots, y_{0n}, y_{l1}, y_{l2}, \dots, y_{ln}) = 0;$$

$$g_2(y_{01}, y_{02}, \ldots, y_{0n}, y_{l1}, y_{l2}, \ldots, y_{ln}) = 0;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$g_n(y_{01}, y_{02}, \ldots, y_{0n}, y_{l1}, y_{l2}, \ldots, y_{ln}) = 0. \qquad (7.2')$$

У простіших випадках граничні умови (7.2′) можуть бути сформульовані окремо в точках t_0 і t_l:

$$\bar{g}_1(\bar{y}_0) = 0; \quad \bar{g}_2(\bar{y}_l) = 0. \qquad (7.3)$$

Якщо умови (7.2′) набувають вигляду

$$\bar{y}_0 = \bar{d}_0, \qquad (7.4)$$

то крайова задача (7.1), (7.2) зводиться до задачі Коші (тут $\bar{d}_0$ — заданий вектор розмірності n).

§ 7.2. МЕТОД ЗВЕДЕННЯ НЕЛІНІЙНОЇ КРАЙОВОЇ ЗАДАЧІ ДО СИСТЕМИ НЕЛІНІЙНИХ РІВНЯНЬ І ЗАДАЧІ КОШІ

Розглянемо нелінійну крайову задачу (7.1), (7.2). Оскільки розв'язок нелінійної задачі Коші існує не завжди, припустимо існування та єдиність розв'язку задачі Коші для системи рівнянь (7.1) при всіх значеннях початкового вектора y_0, які використовуються при розв'язанні задачі.

Перед розглядом методу у загальному вигляді викладемо його основні положення на прикладі крайової задачі для системи двох рівнянь

$$\frac{dy_1}{dt} = f_1(t, y_1, y_2);$$

$$\frac{dy_2}{dt} = f_2(t, y_1, y_2), \quad t_0 \le t \le t_l \qquad (7.5)$$

з граничними умовами

$$g_1(y_{01}, y_{02}) = 0; \qquad (7.6)$$

$$g_2(y_{l1}, y_{l2}) = 0, \qquad (7.7)$$

де f_1, f_2, g_1, g_2 — нелінійні функції своїх аргументів.

На лівому кінці інтервалу при $t = t_0$ задано лише одну умову (7.6), тому задамо ще одну умову:

$$y_1(t_0) = y_0. \qquad (7.8)$$

З (7.6) знаходимо

$$y_2(t_0) = y_2(t_0, y_0). \qquad (7.9)$$

Таким чином, якщо задано y_0, то співвідношення (7.5), (7.8) і (7.9) визначають задачу Коші, розв'язок якої залежить від y_0 як від параметра.

Підставимо розв'язок визначеної вище задачі Коші в граничну умову (7.7) і дістанемо

$$g_2(y_1(t_l, y_0), y_2(t_l, y_0)) = 0. \tag{7.10}$$

Скорочено це рівняння можна подати у вигляді

$$q(y_0) = 0. \tag{7.11}$$

При розв'язанні рівняння (7.11) методом Ньютона дістанемо

$$y_0^{(k+1)} = y_0^{(k)} - \frac{q(y_0^{(k)})}{\partial q(y_0^{(k)})/\partial y_0}, \quad y_0^{(0)}, \quad k = 0, 1, 2, \ldots \tag{7.12}$$

Продиференціюємо тепер рівняння системи (7.5) по параметру y_0. Дістанемо

$$\begin{aligned} \frac{d}{dt}\left(\frac{dy_1}{dy_0}\right) &= \frac{\partial f_1}{\partial y_1}\frac{\partial y_1}{\partial y_0} + \frac{\partial f_1}{\partial y_2}\frac{\partial y_2}{\partial y_0}; \\ \frac{d}{dt}\left(\frac{dy_2}{dy_0}\right) &= \frac{\partial f_2}{\partial y_1}\frac{\partial y_1}{\partial y_0} + \frac{\partial f_2}{\partial y_2}\frac{\partial y_2}{\partial y_0}. \end{aligned} \tag{7.13}$$

Відповідні початкові умови знайдемо аналогічно (7.8) і (7.9):

$$\frac{\partial y_1(t_0)}{\partial y_0} = 1; \quad \frac{\partial y_2(t_0)}{\partial y_0} = -\frac{\partial g_1/\partial y_0}{\partial g_1/\partial y_2}. \tag{7.14}$$

Таким чином, задаючи $y_0^{(0)}$, розв'язуємо дві задачі Коші (7.5), (7.8); (7.9) і (7.13), (7.14). Підставивши одержані розв'язки в (7.12), знайдемо $y_0^{(1)}$ і т.д.

Розглянемо в загальному випадку застосування методу зведення до розв'язання нелінійної крайової задачі для системи звичайних диференціальних рівнянь (7.1), (7.2). Задамо початковий вектор $\bar{y}(t_0) = \bar{y}_0$. При заданому векторі $\bar{y}_0$ і фіксованому значенні t розв'язок задачі Коші для системи рівнянь (7.1) подаємо у вигляді

$$\bar{y}(t) = \bar{z}(t, \bar{y}_0). \tag{7.15}$$

Вектор-функція $\bar{z}(t, \bar{y}_0)$ — однозначна функція n-змінних компонент вектора $\bar{y}_0$ і нелінійно залежить від $\bar{y}_0$. Після підставлення виразу (7.15) в граничні умови (7.2) дістаємо

$$g[\bar{y}_0, \bar{z}(t_l, \bar{y}_0)] = 0. \tag{7.16}$$

Позначаючи $g[\bar{y}_0, \bar{z}(t_l, \bar{y}_0)]$ через $\bar{q}(\bar{y}_0)$, запишемо рівність (7.16) у вигляді

$$\bar{q}(\bar{y}_0) = 0. \tag{7.17}$$

У випадку граничних умов (7.3) маємо

$$\bar{q}_1(\bar{y}_0) = 0, \quad \bar{g}_2[\bar{z}(t_l, \bar{y}_0)] = 0. \tag{7.18}$$

Система (7.17) або (7.18) являє собою систему нелінійних алгебраїчних або трансцендентних рівнянь відносно вектора $\bar{y}_0$.

Отже, розв'язання нелінійної крайової задачі (7.1), (7.2) еквівалентне розв'язанню системи нелінійних рівнянь (7.17) відносно $\bar{y}_0$ і задачі Коші для системи (7.1) при початковій умові $\bar{y}(t_0) = \bar{y}_0$. При цьому слід зазначити, що система (7.17) у явному вигляді не задана й визначена лише алгоритмічно, тобто можна сформулювати алгоритм, за яким обчислюється вектор $\bar{q}(\bar{y})$ за заданим $\bar{y}$.

Використовуючи зазначену еквівалентність, вкажемо чисельний метод розв'язання нелінійної крайової задачі (7.1), (7.2). Оскільки неможливо точно знайти похідні від функції, що задана неявно, для розв'язання системи нелінійних рівнянь (7.17) застосуємо дискретний варіант методу Ньютона. Ітераційний процес Ньютона запишемо у вигляді

$$\Gamma_{h_k}(\bar{y}_0^{(k)})\Delta\bar{y}_0^{(k)} = -\bar{q}(\bar{y}_0^{(k)});$$

$$\bar{y}_0^{(k+1)} = \bar{y}_0^{(k)} + \Delta\bar{y}_0^{(k)};$$

$$\bar{y}_0^{(0)}; \quad k = 0, 1, 2, \ldots, \tag{7.19}$$

де матриця Γ_{h_k}

$$\Gamma_{h_k} = \begin{bmatrix} \dfrac{q_1(\bar{y}_0^{(k)} + h_k\bar{e}_1) - q_1(\bar{y}_0^{(k)})}{h_k} & \cdots & \dfrac{q_1(\bar{y}_0^{(k)} + h_k\bar{e}_n) - q_1(\bar{y}_0^{(k)})}{h_k} \\ \dfrac{q_2(\bar{y}_0^{(k)} + h_k\bar{e}_1) - q_2(\bar{y}_0^{(k)})}{h_k} & \cdots & \dfrac{q_2(\bar{y}_0^{(k)} + h_k\bar{e}_n) - q_2(\bar{y}_0^{(k)})}{h_k} \\ \cdots\cdots & \cdots & \cdots\cdots \\ \dfrac{q_n(\bar{y}_0^{(k)} + h_k\bar{e}_1) - q_n(\bar{y}_0^{(k)})}{h_k} & \cdots & \dfrac{q_n(\bar{y}_0^{(k)} + h_k\bar{e}_n) - q_n(\bar{y}_0^{(k)})}{h_k} \end{bmatrix}.$$

Скорочено позначимо

$$\Gamma_{h_k}(\bar{y}_0^{(k)}) = \left[\frac{\bar{q}(y_0^{(k)} + h_k e_1) - \bar{q}(y_0^{(k)})}{h_k}, \ldots, \frac{\bar{q}(y_0^{(k)} + h_k e_n) - \bar{q}(y_0^{(k)})}{h_k}\right].$$

Для забезпечення квадратичної збіжності ітераційного процесу (7.19) повинна виконуватись нерівність

$$h_k \le \|g[\bar{y}_0^{(k)}, \bar{z}(t_l, \bar{y}_0^{(k)})]\|, \tag{7.20}$$

де норму зручно визначити таким чином:

$$\|g[\bar{y}_0^{(k)}, \bar{z}(t_l, \bar{y}_0^{(k)})]\| =$$

$$= \max_i |g_i[\bar{y}_0^{(k)}, \bar{z}(t_l, \bar{y}_0^{(k)})]|. \tag{7.21}$$

Побудуємо алгоритм для обчислення елементів матриці $\Gamma_{h_k}(\bar{y}_0^{(k)})$. Обчислимо вектори $\bar{q}(\bar{y}_0^{(k)} + h_k e_i)$ $(i = 0, 1, \ldots, n)$. Покладемо $\bar{e}_0 = 0$, а $\bar{e}_i$ $(i = 1, \ldots, n)$ — вектори розмірності n, у яких i-та компонента дорівнює 1, а всі інші — 0. Маємо

$$\bar{q}(\bar{y}_0^{(k)} + h_k\bar{e}_i) = \bar{g}[\bar{y}_0^{(k)} + h_k\bar{e}_i, \bar{z}(t_i, \bar{y}_0^{(k)} + h_k\bar{e}_i)]. \qquad (7.22)$$

Вектор-функція $\bar{z}(t_i, \bar{y}_0^{(k)} + h_k\bar{e}_i)$ — розв'язок задачі Коші для системи (7.1) у точці $t = t_i$ при початковому значенні $\bar{y}_0^{(k)} + h_k\bar{e}_i$. Позначимо розв'язок цієї задачі Коші через $\bar{y}_0^{(i,k)}$. Тоді задачу Коші запишемо у вигляді

$$\frac{d\bar{y}_0^{(i,k)}}{dt} = \bar{f}(t, \bar{y}_0^{(i,k)});$$

$$\bar{y}_0^{(i,k)}(t_0) = \bar{y}_0^{(k)} + h_k\bar{e}_i, \quad i = 0, 1, 2, \ldots, n. \qquad (7.23)$$

Вираз (7.22) тепер можна подати як

$$\bar{q}(\bar{y}_0^{(k)} + h_k\bar{e}_i) = \bar{g}[\bar{y}_0^{(k)} + h_k\bar{e}_i, \bar{y}_{1,0}^{(i,k)}] = \bar{g}[\bar{y}_{0,0}^{(i,k)}, \bar{y}_{1,0}^{(i,k)}]. \qquad (7.24)$$

Із виразів (7.24) і (7.23) випливає, що для побудови матриці $\Gamma_{h_k}(\bar{y}_0^{(k)})$ потрібно розв'язати в загальному випадку $(n + 1)$ задачу Коші вигляду (7.23).

Враховуючи сказане, запишемо обчислювальну схему методу розв'язання нелінійної крайової задачі шляхом зведення до системи нелінійних рівнянь і задачі Коші:

задамо довільно вектор $\bar{y}_0^{(0)}$;

розв'яжемо послідовність задач Коші

$$\frac{d\bar{y}^{(i,k)}}{dt} = \bar{f}(t, \bar{y}_0^{(i,k)}); \quad \bar{y}_{0,0}^{(i,k)} = \bar{y}_0^{(k)} + h_k\bar{e}_i, \quad i = 0, 1, 2, \ldots, n;$$

$$h_k \le \max_i |\bar{g}_i(\bar{y}_{0,0}^{(0,k)}, \bar{y}_{1,0}^{(0,k)})|;$$

побудуємо матрицю

$$\Gamma_{h_k}(\bar{y}_0^{(k)}) = [\bar{g}(\bar{y}_{0,0}^{(1,k)}, \bar{y}_{1,0}^{(1,k)}) - \bar{g}(\bar{y}_{0,0}^{(0,k)}, \bar{y}_{1,0}^{(0,k)}), \ldots,$$
$$\bar{g}(\bar{y}_{0,0}^{(n,k)}, \bar{y}_{1,0}^{(n,k)}) - \bar{g}(\bar{y}_{0,0}^{(0,k)}, \bar{y}_{1,0}^{(0,k)})];$$

розв'яжемо систему рівнянь

$$\Gamma_{h_k}(\bar{y}_0^{(k)})\overline{\Delta(y^{(k)})} = -\bar{g}(\bar{y}_{0,0}^{(0,k)}, \bar{y}_{1,0}^{(0,k)});$$

$$\bar{y}_0^{(k+1)} = \bar{y}_0^{(k)} + \overline{\Delta y^{(k)}}; \quad k = 0, 1, 2, \ldots. \qquad (7.25)$$

Можливі простіші випадки, коли в граничних умовах частина компонент векторів $y_{0,0}$, $y_{1,0}$ задана. Тоді число підлеглих розв'язанню задач Коші зменшується.

Розглянутий ітераційний процес Ньютона розв'язання системи нелінійних алгебраїчних або трансцендентних рівнянь суттєво залежить від вдалого вибору початкового наближення. Якщо початковий вектор розв'язання вибрано не досить близько до шуканого розв'язку системи, ітераційний процес може збігатись дуже повільно або взагалі не збігається. Але оскільки розв'язок системи невідомий і в більшості випадків неможливо вказати обмежену область, в якій знаходиться розв'язок, проблема вибору початкового наближення, що забезпечує збіжність ітераційного процесу, у зазначених випадках ускладнює застосування методу Ньютона. Для подолання труднощів застосовують метод продовження розв'язання по параметру.

Як приклад розглянемо клас задач про напружено-деформований стан круглих пластин змінної у радіальному напрямі товщини під дією нормальних осесиметричних навантажень. Розв'язувана система нелінійних звичайних диференціальних рівнянь має вигляд

$$\frac{dN_r}{dr} = -\frac{1-\nu}{r} N_r + \frac{(1-\nu^2)D_N}{r^2} u;$$

$$\frac{du}{dr} = \frac{1}{D_N} N_r - \frac{\nu}{r} u - \frac{1}{2}\theta_r^2;$$

$$\frac{dQ}{dr} = -\frac{1}{r} Q_r - \frac{(1-\nu^2)D_N}{r^2} u\theta_r - \frac{1}{D_M} N_r M_r + q_\nu;$$

$$\frac{dM_r}{dr} = -\frac{1-\nu}{r} M_r + \frac{(1-\nu)D_M}{r^2}\theta_r - Q_r;$$

$$\frac{d\omega}{dr} = -\theta_r;$$

$$\frac{d\theta_r}{dr} = \frac{1}{D_M} M_r - \frac{\nu}{r}\theta_r, \quad b \le r \le a, \tag{7.26}$$

де N_r, Q_r, M_r — зусилля і момент; u, ω, θ_r — переміщення й кут повороту.

Введемо такі безрозмірні величини:

$$x = r/a; \quad h^* = h/h_0; \quad N_r^* = \frac{3(1-\nu^2)a^2}{Eh_0^3} N_r; \quad u^* = 3(1-\nu^2)au/h_0^2;$$

$$\omega^* = \sqrt{3(1-\nu^2)}\,\omega/h_0; \quad M_r^* = \frac{\sqrt{[3(1-\nu^2)]^3}}{Eh_0^4} M_r;$$

$$Q_r^* = \frac{\sqrt{[3(1-\nu^2)]^3}\,a^3}{Eh_0^4} Q_r; \quad \theta_r^* = \sqrt{3(1-\nu^2)}\,\frac{a}{h_0}\theta_r;$$

$$q_\nu^* = \frac{\sqrt{[3(1-\nu^2)]^3}\,a^4}{Eh_0^4} q_\nu; \quad x_0 = b/a. \tag{7.27}$$

Тоді розв'язувана система рівнянь (7.26) запишеться у вигляді

$$\frac{dN_r}{dx} = -\frac{1-\nu}{x} N_r + \frac{h}{x^2} u;$$

$$\frac{du}{dx} = \frac{1-\nu^2}{h} N_r - \frac{\nu}{x} u - \frac{1}{2} \theta_r^2;$$

$$\frac{dQ_r}{dx} = -\frac{1}{x} Q_r - \frac{h}{x^2} u\theta_r - \frac{4}{h^3} N_r M_r + q_\gamma;$$

$$\frac{dM_r}{dx} = -\frac{1-\nu}{x} M_r + \frac{(1-\nu^2)h^3}{4x^2} \theta_r - Q_r;$$

$$\frac{d\omega}{dx} = -\theta_r;$$

$$\frac{d\theta_r}{dx} = \frac{4}{h^3} M_r - \frac{\nu}{x} \theta_r, \quad x_0 \le x \le 1. \tag{7.28}$$

У цих рівняннях у безрозмірних величинах для простоти опущено індекс *.

Систему рівнянь (7.18) запишемо у векторній формі

$$\frac{d\overline{N}}{dx} = \overline{f}(x, \overline{N}), \tag{7.29}$$

де $N = \{N_i\}^T = \{N_r, u, Q_r, M_r, \omega, \theta_r\}^T$ — шестивимірний вектор.

Граничні умови в загальному випадку мають вигляд

$$B_1\overline{N}(x_0) = \overline{b}_1; \quad B_2\overline{N}(1) = \overline{b}_2, \tag{7.30}$$

де B_1 і B_2 — задані прямокутні матриці відповідно порядків $k \times 6$ і $(6-k) \times 6$; $(k < 6)$; $\overline{b}_1$, $\overline{b}_2$ — задані вектори.

Наведемо результати розв'язання задачі за допомогою даного методу.

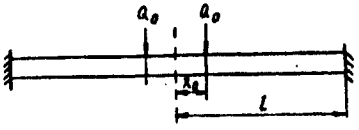


Рис. 11

Розглянемо геометрично нелінійну деформацію жорстко закріпленої на зовнішньому контурі $x = 1$ кільцевої пластини сталої товщини h під дією прикладеного на внутрішньому контурі $x = x_0$ поперечного зусилля Q_0 (рис. 11). Граничні умови набувають вигляду:

при $x = x_0$

$$u = \theta_r = 0; \quad Q_r = Q_0;$$

при $x = 1$

$$u = \omega = \theta_r = 0.$$

При розв'язанні задачі приймаємо: $h_0 = 1$; $x_0 = 0{,}2$; $q_\gamma = 0$; $\nu = 0{,}3$.

Результати розв'язання задачі в точці $x = x_0$ при $Q_0 = 20$ наведені в табл. 22.

Таблиця 22

Наближення	ω^*	N_r^*	M_r^*
0	1,0360	0,0000	4,7060
1	0,8732	1,0482	4,1419
2	0,8983	1,4484	4,1516
3	0,8979	1,4463	4,1509
4	0,8979	1,4463	4,1509

За початкове наближення використано розв'язання лінійної задачі. Як видно з табл. 22, значення функцій у третьому і четвертому наближеннях збігаються до четвертої значущої цифри. Але зі збільшенням прикладеного зусилля задача ускладнюється щодо вибору початкового наближення та швидкості збіжності.

§ 7.3. МЕТОД ЛІНЕАРИЗАЦІЇ

Метод зведення нелінійної крайової задачі до системи нелінійних алгебраїчних або трансцендентних рівнянь і задачі Коші для початкового вектора (див. § 7.2.) в деяких задачах може виявитись нестійким, оскільки задачі Коші для рівнянь, які необхідно розв'язувати в процесі застосування методу, стають жорсткими. Для їх розв'язання необхідно застосувати спеціальні методи, що пов'язано з додатковими обсчислювальними труднощами.

У цих випадках для розв'язання нелінійних крайових задач ефективніше застосовувати такі ітераційні процеси, на кожному кроці яких розв'язується лінійна крайова задача, що дозволяє використати для розв'язання будь-який стійкий чисельний метод, коли при його реалізації розв'язання задачі Коші відбувається на невеликих інтервалах інтегрування.

До зазначених методів розв'язання нелінійних крайових задач належить метод лінеаризації, що базується на побудові ітераційного процесу, на кожному кроці якого розв'язується лінійна крайова задача для наступного наближення, яка використовує інформацію попереднього.

Розглянемо окремий випадок нелінійної крайової задачі для звичайного диференціального рівняння другого порядку

$$x'' = f(t, x, x'), \quad 0 \le t \le 1 \tag{7.31}$$

з нелінійними граничними умовами

$$g_1(x_0, x') = 0, \quad g_2(x_l, x_l') = 0. \tag{7.32}$$

Побудуємо ітераційний процес на основі лінеаризації крайової задачі (7.31), (7.32). Нехай відомо якесь наближення x_k розв'язку задачі. У цьому випадку маємо

$$x^* = x_k + \Delta x_k^*, \tag{7.33}$$

де x^* — точний розв'язок,

Підставляючи розв'язок (7.33) в (7.31) і (7.32), знаходимо

$$(x_k + \Delta x_k^*)'' = f[t, x_k + \Delta x_k^*, (x_k + \Delta x_k^*)'\,]. \tag{7.34}$$

На основі формули Лагранжа

$$(x_k + \Delta x_k^*)'' = f(t, x_k, x'_x) + f'_x(t, \xi_k, \xi'_k)(x^* - x_k) +$$

$$+ f'_{x'}(t, \xi_k, \xi'_k)[(x_k^*)' - (x_k)'\,]. \tag{7.35}$$

Оскільки точка ξ_k невідома, то покладаємо $\xi_k \approx x_k$, $x^* \approx x_{k+1} = x_k + \Delta x_k$. Тоді з (7.35) дістаємо

$$x''_{k+1} = f(t, x_k, x'_k) + f'_x(t, x_k, x'_k)(x_{k+1} - x_k) +$$

$$+ f'_{x'}(t, x_k, x'_k)(x'_{k+1} - x'_k)$$

або

$$x''_{k+1} - f'_{x'}(t, x_k, x'_k)x'_{k+1} - f'_x(t, x_k, x'_k)x_{k+1} = f(t, x_k, x'_k) -$$

$$- f'_{x'}(t, x_k, x'_k)x'_k - f'_x(t, x_k, x'_k)x_k, \; x_0, \; x'_0, \quad k = 1, 2, \ldots \tag{7.36}$$

Для граничних умов (7.32) отримуємо

$$g_1[x_k(0), x'_k(0)\,] + g'_{1,x}[x_k(0), x'_k(0)\,][x_{k+1}(0) - x_k(0)\,] +$$

$$+ g'_{1,x'}[x_k(0), x'_k(0)\,][x'_{k+1}(0) - x'_k(0)\,] = 0;$$

$$g_2[x_k(1), x'_k(l)\,] + g'_{2,x}[x_k(l), x'_k(l)\,][x_{k+1}(l) - x_k(l)\,] +$$

$$+ g'_{2,x'}[x_k(l), x'_k(l)\,][x'_{k+1}(l) - x'_k(l)\,] = 0$$

або

$$g'_{1,x'}[x_k(0), x'_k(0)\,]x'_{k+1}(0) + g'_{1,x}[x_k(0), x'_k(0)\,]x_{k+1}(0) =$$

$$= - g_1[x_k(0), x'_k(0)\,] + g'_{1,x}[x_k(0), x'_k(0)\,]x_k(0) +$$

$$+ g'_{1,x'}[x_k(0), x'_k(0)\,]x'_k(0);$$

$$g'_{2,x'}[x_k(l), x_k'(l)\,]x'_{k+1}(l) + g'_{2,x}[x_k(l), x'_k(l)\,]x_{k+1}(l) =$$

$$= - g_2[x_k(l), x'_k(l)\,] + g'_{2,x}[x_k(l), x'_k(l)\,]x_k(l) +$$

$$+ g'_{2,x'}[x_k(l), x'_k(l)\,]x'_k(l);$$

$$x_0(0), x'_0(0), x_0(l), x'_0(l), \quad k = 0, 1, 2, \ldots \tag{7.37}$$

Таким чином, для кожного k-го наближення приходимо до лінійної крайової задачі (7.36), (7.37). Перебираючи k, отримаємо ітераційний

процес. При деяких обмеженнях на функції f, g_1, g_2 та їхні похідні для невеликих значень l можна довести єдність і квадратичну збіжність процесу.

Приклад. Задача про скінченну деформацію пружної струни під дією поперечного навантаження описується рівнянням

$$-x'' = a^2(x')^2 + 1, \quad 0 \le t \le l \tag{7.38}$$

з граничними умовами

$$x(0) = 0, \quad x(l) = 0. \tag{7.39}$$

Розв'язання. Точне аналітичне розв'язання має вигляд

$$x(t) = \frac{1}{a^2} \ln \left[\frac{\cos a(t - 1/2)}{\cos a/2} \right]. \tag{7.40}$$

За допомогою методу лінеаризації будуємо ітераційну схему

$$x''_{k+1} + 2a^2 x'_k x''_{k+1} = -1 + a^2 (x'_k)^2; \tag{7.41}$$

$$x_{k+1}(0) = 0; \quad x_{k+1}(l) = 0;$$

$$a^2 = 0{,}49; \quad k = 0, 1, 2, \ldots \tag{7.42}$$

Результати чисельного розв'язання крайової задачі (7.38), (7.39) за ітераційною схемою (7.41), (7.42) наведено в табл. 23.

Таблиця 23

t	$x_0(t)$	$x_1(t)$	$x_2(t)$	Точний розв'язок
0,0	0,00000	0,00000	0,000000	0,000000
0,1	0,00000	0,04500	0,046570	0,046571
0,2	0,00000	0,08000	0,082302	0,082304
0,3	0,00000	0,10500	0,107571	0,107573
0,4	0,00000	0,12000	0,122632	0,122635
0,5	0,00000	0,12500	0,127636	0,127639
0,6	0,00000	0,12000	0,122632	0,122635
0,7	0,00000	0,10500	0,107571	0,107573
0,8	0,00000	0,08000	0,082302	0,082304
0,9	0,00000	0,04500	0,046570	0,046571
1,0	0,00000	0,00000	0,000000	0,000000

Як видно з таблиці, друге наближення збігається з точним розв'язком до п'яти значущих цифр.

Розглянемо застосування методу лінеаризації до розв'язання нелінійних крайових задач для системи звичайних диференціальних рівнянь

$$\frac{d\bar{x}}{dt} = \bar{f}(t, \bar{x}), \quad 0 \le t \le 1 \tag{7.43}$$

з граничними умовами

$$\bar{g}(\bar{x}_0, \bar{x}_l) = 0. \tag{7.44}$$

Нехай визначено наближений розв'язок крайової задачі (7.43), (7.44), який позначимо через $\bar{x}_k(t)$. Тоді точний розв'язок $\bar{x}^*(t)$ можна записати у вигляді

$$\bar{x}^*(t) = \bar{x}_k(t) + \Delta\bar{x}_k^*(t). \tag{7.45}$$

Підставляючи розв'язок (7.45) в (7.43), (7.44), знаходимо

$$\frac{d}{dt}(\bar{x}_k + \overline{\Delta x_k^*}) = \bar{f}(t, \bar{x}_k + \overline{\Delta x_k^*}). \tag{7.46}$$

За допомогою формули Лагранжа маємо

$$\frac{d}{dt}(\bar{x}_k + \overline{\Delta x_k^*}) = \bar{f}(t, \bar{x}_k) + \Gamma(t, \bar{\xi}_k)\overline{\Delta x_k^*}, \tag{7.47}$$

де $\|\bar{\xi}_k(t) - \bar{x}_k(t)\| \le \|\overline{\Delta x_k^*}\|$.

Граничні умови (7.43) перетворюються в

$$\bar{g}(\bar{x}_{0,k} + \overline{\Delta x_{0,k}^*}, \bar{x}_{l,k} + \overline{\Delta x_{l,k}^*}) = \bar{g}(\bar{x}_{0,k}, \bar{x}_{l,k}) + \\ + \Gamma_0(\bar{\xi}_{0,k}, \bar{\xi}_{l,k})\overline{\Delta x_{0,k}^*} + \Gamma_l(\bar{\xi}_{0,k}, \bar{\xi}_{l,k})\overline{\Delta x_{l,k}^*}, \tag{7.48}$$

де $\|\bar{\xi}_{0,k} - \bar{x}_{0,k}\| \le \|\overline{\Delta x_{0,k}^*}\|$; $\|\bar{\xi}_{l,k} - \bar{x}_{l,k}\| \le \|\overline{\Delta x_{l,k}^*}\|$.

Оскільки вектори $\bar{\xi}_k$, $\bar{\xi}_{0,k}$, $\bar{\xi}_{l,k}$ невідомі, то замінюємо їх відповідно векторами $\bar{x}_k$, $\bar{x}_{0,k}$, $\bar{x}_{l,k}$ і покладемо $\bar{x}^* \approx \bar{x}_{k+1}$; $\bar{x}_{0,k}^* \approx \bar{x}_{0,k+1}$; $\bar{x}_{l,k}^* \approx \bar{x}_{l,k+1}$. Тоді з (7.47), (7.48) дістаємо ітераційний процес

$$\frac{d\bar{x}_{k+1}}{dt} = \Gamma(t, \bar{x}_k)\bar{x}_{k+1} + \bar{f}(t, \bar{x}_k) - \Gamma(t, \bar{x}_k)\bar{x}_k; \tag{7.49}$$

$$\Gamma_0(\bar{x}_{0,k}, \bar{x}_{l,k})\bar{x}_{0,k+1} + \Gamma_l(\bar{x}_{0,k}, \bar{x}_{l,k})\bar{x}_{l,k+1} = \bar{g}(\bar{x}_{0,k}, \bar{x}_{l,k}) + \\ + \Gamma_0(\bar{x}_{0,k}, \bar{x}_{l,k})\bar{x}_{0,k} + \Gamma_l(\bar{x}_{0,k}, \bar{x}_{l,k})\bar{x}_{l,k}, \bar{x}_0(t) \quad k = 1, 2, \ldots \tag{7.50}$$

На k-му наближенні розв'язується лінійна крайова задача для $\bar{x}_{k+1}$, для чого застосовується стійкий чисельний метод (зокрема, метод дискретної ортогоналізації).

На конкретних прикладах зупинимось на аспектах застосування методу лінеаризації до розв'язання задач.

Розглянемо напружено-деформований стан геометрично нелінійних круглих пластин зі змінною в радіальному напрямі товщиною під дією осесиметричних навантажень.

На ряді прикладів проаналізуємо розв'язання нелінійних задач теорії круглих пластин і за допомогою індуктивних засобів одержимо деякі оцінки точності їх розв'язків.

Система нелінійних рівнянь (7.26), що описує даний клас задач, має вигляд

$$\frac{d\overline{N}}{dx} = \overline{f}(x, \overline{N}), \quad x_0 \le x \le 1, \tag{7.51}$$

де $\overline{N} = \{N_r, u, Q_r, M_r, \omega, \theta_r\}^T$.

Розглянемо лінійні граничні умови

$$B_1\overline{N}(x_0) = \overline{b}_1; \quad B_2\overline{N}(1) = \overline{b}_2. \tag{7.52}$$

Для розв'язання нелінійної крайової задачі (7.51), (7.52) скористаємось методом лінеаризації. Відповідно до нього побудуємо ітераційну схему

$$\frac{d\overline{N}^{(k)}}{dx} = \overline{f}(x, \overline{N}^{(k)}) + \Gamma(\overline{N}^{(k)})(\overline{N}^{(k+1)} - \overline{N}^{(k)}); \tag{7.53}$$

$$B_1\overline{N}^{(k+1)}(x_0) = \overline{b}_1; \quad B_2\overline{N}^{(k+1)}(1) = \overline{b}_2, \ \overline{N}^{(0)}, \ k = 0, 1, 2, \dots \tag{7.54}$$

Для крайової задачі (7.51), (7.52) ітераційна схема (7.53) набуває вигляду (в безрозмірних величинах)

$$\frac{dN_r^{(k+1)}}{dx} = -\frac{1-\nu}{x} N_r^{(k+1)} + \frac{h}{x^2} u^{(k+1)};$$

$$\frac{du^{(k+1)}}{dx} = \frac{1-\nu^2}{x} N_r^{(k+1)} - \frac{\nu}{x} u^{(k+1)} - \theta_r^{(k)}\theta_r^{(k+1)} + \frac{1}{2}(\theta_r^{(k)})^2;$$

$$\frac{dQ_r^{(k+1)}}{dx} = -\frac{4}{h^3} M_r^{(k)} N_r^{(k+1)} - \frac{h}{x^2} u^{(k+1)}\theta_r^{(k)} - \frac{1}{x} Q_r^{(k+1)} -$$

$$- \frac{4}{h^3} N_r^{(k)} M_r^{(k+1)} - \frac{h}{x^2} \theta_r^{(k+1)} u^{(k)} + \frac{4}{h^3} M_r^{(k)} N_r^{(k)} + \frac{h}{x^2} \theta_r^{(k)} u^{(k)} + q_r;$$

$$\frac{dM_r^{(k+1)}}{dx} = -Q_r^{(k+1)} + \frac{h^3(1-\nu^2)}{4x^2} \theta_r^{(k+1)} - \frac{1-\nu}{x} M_r^{(k+1)};$$

$$\frac{d\omega^{(k+1)}}{dx} = -\theta_r^{(k+1)};$$

$$\frac{d\theta_r^{(k+1)}}{dx} = \frac{4}{h^3} M_r^{(k+1)} - \frac{\nu}{x} \theta_r^{(k+1)}, \quad \overline{N}^{(0)}, \quad k = 0, 1, 2, \dots \tag{7.55}$$

Таким чином, вихідна крайова задача для нелінійної системи звичайних диференціальних рівнянь зводиться до послідовності лінійних крайових задач. Але при цьому виникає ряд питань, пов'язаних з чисельною реалізацією ітераційного процесу. Лінійну крайову задачу для k-го наближення розв'яжемо методом дискретної ортогоналізації (див. гл. 6).

Одержана числова інформація повинна бути врахована в коефіцієнтах і правих частинах наступного $(k+1)$-го наближення. При цьому необхідно зберігати в пам'яті ЕОМ великі масиви значень функцій $N_i^{(k)}$, крім того, об'єм необхідної інформації з кожним наступним наближенням збільшується вдвічі, що пов'язане з розв'язком задач Коші методом Рунге—Кутта. У деяких задачах можна застосувати лінійну чи квадратичну інтерполяцію за дискретними значеннями функцій $N_i^{(k)}$.

Збіжність ітераційного процесу (7.55) для розглядуваного класу задач перевіряється за допомогою різних індуктивних засобів. На конкретних прикладах проілюструємо аспекти обчислювальної реалізації і збіжності ітераційного процесу розв'язання даного класу задач і проаналізуємо напружено-деформований стан деяких гнучких кільцевих пластин.

З метою одержання деяких оцінок збіжності ітераційного процесу розглянемо деформацію жорстко закріпленої на зовнішньому контурі кільцевої пластини сталої товщини h_0 під дією прикладеного на внутрішньому контурі поперечного зусилля Q_0 (див. § 7.2.). Граничні умови: при $x = x_0$

$$U = 0, \quad \theta_r = 0, \quad Q_r = 0;$$

при $x = 1$

$$u = 0, \quad \omega = 0, \quad \theta_r = 0. \tag{7.56}$$

При розв'язанні задачі приймалось: $Q_0 = P/x_0$; $P = 4$; $h = 1$; $x_0 = 0{,}2$; $n = 8$.

Результати розв'язання задачі в точці x_0 наведено для варіанта 1 в табл. 24. Як видно, вже в четвертому наближенні для всіх функцій має місце збіг до шести значущих цифр. При порівнянні кінцевих результатів з розв'язком задачі за допомогою методу збурення виявлено їх добрий збіг для $P = 2, 4, 6, 8$ і розходження в третій значущій цифрі для $P \geq 10$. Це, мабуть, можна пояснити тим, що в методі збурення враховується тільки два члени.

Для оцінки збіжності процесу застосований індуктивний спосіб, при якому лінійна крайова задача розв'язувалась в еквівалентному, але іншому математичному формулюванні. Замість умов (7.56) у точці x_0 задавались такі: при $x = x_0$ $u = 0$; $Q_r = Q_0$; $M_r = M_0$, де за M_0 приймалось значення M_r, одержане в попередньому розв'язку, тобто $M_0 = 4{,}150904$.

Результат розв'язання задачі в такому формулюванні наведено для варіанта II в табл.24. Порівнюючи результати варіантів I і II, можна зроби-

Таблиця 24

Варіант	Наближення	u^*	N_r^*	Q_r^*	M_r^*	w^*	v_r^*
I	0	0	0,000000	20	4,705999	1,056890	0
	1	0	1,343001	20	4,101359	0,883097	0
	2	0	1,446459	20	4,151989	0,898048	0
	3	0	1,446308	20	4,150904	0,897877	0
	4	0	1,446308	20	4,150904	0,897877	0
II	0	0	0,000000	20	4,150904	1,189961	0,566903
	1	0	1,313352	20	4,150904	0,908850	0,017061
	2	0	1,446234	20	4,150904	0,835576	0,000093
	3	0	1,446308	20	4,150904	0,897877	0,000000
	4	0	1,446308	20	4,150904	0,897877	0,000000

ти висновок, що в четвертому наближенні вони збігаються. На підставі того, що застосовувались цілком різні математичні формулювання однієї й тієї ж задачі, а отже, й обчислювальні схеми, одержаний збіг результатів свідчить про їхню високу точність. Для великих значень P ($P \geq 20$) така точність досягається при шести наближеннях.

§ 7.4. МЕТОД ПРОДОВЖЕННЯ РОЗВ'ЯЗАННЯ ПО ПАРАМЕТРУ

При розв'язанні нелінійних крайових задач для систем звичайних диференціальних рівнянь за допомогою методу зведення до системи алгебраїчних або трансцендентних рівнянь і задачі Коші (§ 7.2), і методу лінеаризації (§ 7.3) виникають труднощі, пов'язані з вибором початкового наближення, що забезпечує збіжність ітераційного процесу в цих випадках. У зв'язку з цим, як і при чисельному розв'язанні недиференціальних систем, одним з шляхів подолання їх є застосування методу продовження розв'язку по параметру безпосередньо до диференціальних систем.

Таким чином, за вихідну будемо розглядати нелінійну крайову задачу вигляду

$$\frac{d\bar{x}}{dt} = \bar{f}(t, \bar{x}), \quad 0 \leq t \leq 1; \tag{7.57}$$

$$\bar{g}(\bar{x}_0, \bar{x}_l) = 0. \tag{7.58}$$

При застосуванні методу продовження розв'язання по параметру замість крайової задачі (7.57), (7.58) розглянемо задачу з параметром λ

$$\frac{d\bar{x}}{dt} = \bar{\varphi}(t, \bar{x}, \lambda), \quad 0 \leq \lambda \leq 1; \tag{7.59}$$

$$\bar{g}_1(\bar{x}_0, \bar{x}_l, \lambda) = 0. \tag{7.60}$$

При цьому покладемо, що вектор-функції $\bar{\varphi}(t, \bar{x}, \lambda)$ і $\bar{g}_1(\bar{x}_0, \bar{x}_l, \lambda)$ неперервні, достатнє число раз диференційовні по λ, і при всіх значеннях λ крайова задача (7.59), (7.60) має розв'язок.

Параметр λ в крайову задачу (7.59), (7.60) введемо таким чином, щоб при $\lambda = 0$ розв'язок крайової задачі

$$\frac{d\bar{x}}{dt} = \bar{\varphi}(t, \bar{x}, 0); \tag{7.61}$$

$$\bar{g}_1(\bar{x}_0, \bar{x}_l, 0) = 0 \tag{7.62}$$

можна було знайти порівняно просто, а при $\lambda = 1$ виконувались рівності

$$\bar{\varphi}(t, \bar{x}, 1) = \bar{f}(t, \bar{x}); \quad \bar{g}_1(\bar{x}_0, \bar{x}_l, 1) = \bar{g}(\bar{x}_0, \bar{x}_l), \tag{7.63}$$

тобто крайова задача (7.59), (7.60) збігалася з вихідною крайовою задачею (7.57), (7.58). Тоді шукану вектор-функцію $\bar{x}$ можна розглядати як функцію від λ, тобто $\bar{x} = \bar{x}(\lambda)$, де $\bar{x}(0) = \bar{x}^{(0)}$ — відомий розв'язок крайової задачі (7.61), (7.62); $\bar{x}(1)$ — шуканий розв'язок крайової задачі (7.57), (7.58).

Відповідно до вказаних умов розглянемо деякі способи введення параметра й одержання крайової задачі (7.59), (7.60).

Перший спосіб. Задамо довільно вектор $\bar{x}_0$ і сформулюємо задачу Коші

$$\frac{d\bar{x}}{dt} = \bar{f}(t, \bar{x}^{(0)}); \quad \bar{x}^{(0)}(0) = \bar{x}_0. \tag{7.64}$$

Знаходимо будь-яким чисельним методом (наприклад, Рунге—Кутта) розв'язок задачі Коші (7.64) $\bar{x}^{(0)}(t)$. Використовуючи одержаний розв'язок $\bar{x}^{(0)}(t)$, визначаємо

$$\bar{g}(\bar{x}_0, \bar{x}_l^{(0)}) = \bar{d}_0. \tag{7.65}$$

Сформулюємо крайову задачу (7.59), (7.60), яка містить параметр, таким чином:

$$\frac{d\bar{x}}{dt} = \bar{f}(t, \bar{x}); \tag{7.66}$$

$$\bar{g}(\bar{x}_0, \bar{x}_l) - (1 - \lambda)\bar{d}_0 = 0. \tag{7.67}$$

Перевіримо виконання вказаних умов:
при $\lambda = 0$ дістанемо крайову задачу

$$\frac{d\bar{x}}{dt} = \bar{f}(t, \bar{x}); \quad \bar{g}(\bar{x}_0, \bar{x}_l) = \bar{d}_0, \tag{7.68}$$

розв'язок якої в силу (7.65) збігається з розв'язком задачі Коші (7.64), тобто $\bar{x} = \bar{x}^{(0)}(t)$;

при $\lambda = 1$ з (7.66), (7.67) приходимо до крайової задачі

$$\frac{d\bar{x}}{dt} = \bar{f}(t, \bar{x}); \quad \bar{g}(\bar{x}_0, \bar{x}_l) = 0, \tag{7.69}$$

яка ідентична вихідній крайовій задачі (7.57), (7.58).

Побудуємо для функції $\bar{x}(\lambda)$ математичну модель. Диференціюючи співвідношення (7.66), (7.67) по λ і враховуючи при цьому, що $\bar{x}(t)$, $\bar{x}_0$, $\bar{x}_l$ є функції λ, знаходимо

$$\frac{d}{dt}\left(\frac{\partial x}{\partial \lambda}\right) = \sum_{i=1}^{n} \frac{\partial \bar{f}}{\partial x_i} \frac{\partial x_i}{\partial \lambda};$$

$$\sum_{i=1}^{n} \frac{\partial \bar{g}}{\partial x_{0i}} \frac{\partial x_{0i}}{\partial \lambda} + \sum_{i=0}^{n} \frac{\partial \bar{g}}{\partial x_{li}} \frac{\partial x_{li}}{\partial \lambda} = -\bar{d}_0. \tag{7.70}$$

Введемо позначення:

$$\bar{u}(t, \lambda) = \frac{\partial \bar{x}}{\partial \lambda}; \quad \Gamma_f = \Gamma_f(t, \bar{x}) = \left(\frac{\partial \bar{f}}{\partial x_1}, \frac{\partial \bar{f}}{\partial x_2}, \ldots, \frac{\partial \bar{f}}{\partial x_n}\right);$$

$$\Gamma_0 = \Gamma_0(\bar{x}_0, \bar{x}_l) = \left(\frac{\partial \bar{g}}{\partial x_{01}}, \frac{\partial \bar{g}}{\partial x_{02}}, \ldots, \frac{\partial \bar{g}}{\partial x_{0n}}\right);$$

$$\Gamma_l = \Gamma_l(\bar{x}_0, \bar{x}_l) = \left(\frac{\partial \bar{g}}{\partial x_{l1}}, \frac{\partial \bar{g}}{\partial x_{l2}}, \ldots, \frac{\partial \bar{g}}{\partial x_{ln}}\right). \tag{7.71}$$

З урахуванням (7.71) співвідношення (7.70) запишемо у вигляді

$$\frac{d\bar{u}}{dt} = \Gamma_f(t, \bar{x})\bar{u}; \quad \Gamma_0(\bar{x}_0, \bar{x}_l)\bar{u}_0 + \Gamma_l(\bar{x}_0, \bar{x}_l)\bar{u}_l = -\bar{d}_0, \tag{7.72}$$

де $\bar{u}_0 = \bar{u}(0, \lambda)$; $\bar{u}_l = \bar{u}(1, \lambda)$.

Таким чином, співвідношеннями (7.72) формулюється лінійна крайова задача для вектор-функції $\bar{u}(t, \lambda)$, до якої слід додати ще задачу Коші

$$\frac{\partial \bar{x}}{\partial \lambda} = \bar{u}(t, \lambda); \quad \bar{x}(t, 0) = \bar{x}^{(0)}(t). \tag{7.73}$$

При $\lambda = 1$ розв'язок одержаної задачі є розв'язком вихідної нелінійної крайової задачі (7.57), (7.60).

Для знаходження функції $\bar{x} = \bar{x}(t)$, що є розв'язком крайової задачі (7.66), (7.67), яка містить параметр, необхідно разом розв'язувати задачі (7.71), (7.73). При цьому змінну t в задачі (7.73) слід розглядати як параметр.

Другий спосіб. Запишемо крайову задачу з параметром λ з (7.59), (7.60):

$$\frac{d\bar{x}}{dt} = A(t)\bar{x} + \bar{b}(t) + \lambda[\bar{f}(t, \bar{x}) - A(t)\bar{x} - \bar{b}(t)];$$

$$B\bar{x}_0 + C\bar{x}_l + \lambda[\bar{g}(\bar{x}_0, \bar{x}_l) - B\bar{x}_0 - C\bar{x}_l + \bar{d}_0] = \bar{d}_0. \qquad (7.74)$$

При $\lambda = 0$

$$\frac{d\bar{x}}{dt} = A(t)\bar{x} + \bar{b}(t); \quad B\bar{x}_0 + C\bar{x}_l = \bar{d}_0. \qquad (7.75)$$

Розв'язок лінійної крайової задачі (7.75) при заданих матрицях $A(t)$, B, C і векторах $\bar{b}(t)$, $\bar{d}_0$ можна знайти чисельним методом, наприклад, дискретної ортогоналізації. При $\lambda = 1$ приходимо до вихідної крайової задачі (7.57), (7.60).

Матриці $A(t)$, B, C і вектори $\bar{b}(t)$, $\bar{d}_0$ вибираються таким чином, щоб вектор-функції $A(t)\bar{x} + \bar{b}(t)$ і $B\bar{x}_0 + C\bar{x} = \bar{d}_0$ апроксимували в якомусь відношенні відповідно вектор-функції $\bar{f}(t, x)$ і $\bar{g}(\bar{x}_0, \bar{x}_l)$ в околі розв'язку крайової задачі (7.57), (7.58) ($0 \le t \le 1$).

Побудова зазначених матриць і векторів є складною задачею, оскільки розв'язок вихідної задачі невідомий. Тому зручно застосувати такий підхід до введення параметра в задачах, які близькі до лінійних. Наприклад, шукана вектор-функція може мати головну лінійну частину і малий нелінійний додаток

$$\bar{f}(t, \bar{x}) = A(t)\bar{x} + \bar{b}(t) + \varepsilon_1\bar{f}_1(t, \bar{x}), \qquad (7.76)$$

де ε_1 — мала величина; вектор-функція $\bar{g}(\bar{x}_0, \bar{x}_l)$ має подібну структуру, тобто

$$\bar{g}(\bar{x}_0, \bar{x}_l) = B\bar{x}_0 + C\bar{x}_l + \bar{d}_0 + \varepsilon_2\bar{g}_1(\bar{x}_0, \bar{x}_l), \qquad (7.77)$$

де ε_2 — мала величина.

При умові визначення невеликої області, у якій міститься розв'язок вихідної крайової задачі (7.57), (7.58) при кожному фіксованому t, матриці $A(t)$, B, C і вектори $\bar{b}(t)$, $\bar{d}_0$ можна побудувати за допомогою методу найменших квадратів.

Побудуємо рівняння для знаходження функції $\bar{x}(\lambda)$. Продиференціюємо співвідношення (7.74), розглядаючи $\bar{x}$, $\bar{x}_0$, $\bar{x}_l$ як функції λ. Приймаючи раніше прийняті позначення, дістанемо

$$\frac{d\bar{x}}{dt} = \bar{u}(t, \lambda);$$

$$\frac{d\bar{u}}{dt} = [(1 - \lambda)A(t) - \lambda\Gamma_f(t, \bar{x})]\bar{u} + \bar{f}(t, \bar{x}) - A(t)\bar{x} - \bar{b}(t);$$

$$[(1 - \lambda)B + \lambda\Gamma_0(\bar{x}_0, \bar{x}_l)]\bar{u}_0 + [(1 - \lambda)C + \lambda\Gamma_l(\bar{x}_0, \bar{x}_l)]\bar{u}_l =$$

$$= B\bar{x}_0 + C\bar{x}_l + \bar{d}_0 - \bar{g}(\bar{x}_0, \bar{x}_l). \qquad (7.78)$$

Таким чином, необхідно проінтегрувати систему звичайних диференціальних рівнянь, права частина яких визначається з розв'язку деякої лінійної крайової задачі.

Розглянемо метод розв'язання одержаних задач. Для розв'язання задачі (7.72), (7.73) застосуємо метод Ейлера, обчислювальну схему якого запишемо у вигляді

$$\bar{x}^{(0)}(t) \text{ — розв'язок задачі Коші (7.61);}$$

$$\frac{d\bar{u}^{(k)}}{dt} = \Gamma_f(t, \bar{x}^{(k)})\bar{u};$$

$$\Gamma_0(\bar{x}_0^{(k)}, \bar{x}_l^{(k)})\bar{u}_0^{(k)} + \Gamma_l(\bar{x}_0^{(k)}, \bar{x}_l^{(k)})\bar{u}_l^{(k)} = -\bar{d}_0;$$

$$\bar{x}^{(k+1)}(t) = \bar{x}^{(k)}(t) + (\lambda_{k+1} - \lambda_k)\bar{u}^{(k)}(t), \quad k = 0, 1, 2, \ldots \quad (7.79)$$

Лінійну крайову задачу, що входить до складу обчислювальної схеми (7.79), можна розв'язати методом дискретної ортогоналізації.

При нестійкості обчислень за схемою (7.79) можна застосувати інший метод, який не потребує інтегрування системи диференціальних рівнянь. Розглянемо його. Інтервал зміни параметра $\lambda \in [0, 1]$ розбиваємо точками $0 = \lambda_0 < \lambda_1 < \lambda_2 < \ldots < \lambda_m = 1$ на m достатньо малих частин. Нехай $\bar{x}^{(k)}(t)$ — розв'язок крайової задачі (7.66), (7.67) при $\lambda = \lambda_k$. При $\lambda = \lambda_0 = 0$ розв'язок крайової задачі відомий і дорівнює $\bar{x}(t)$, як розв'язок задачі (7.64).

При $\lambda = \lambda_1$ знаходимо $\bar{x}^{(1)}(t)$ як розв'язок крайової задачі

$$\frac{d\bar{x}^{(1)}}{dt} = \bar{f}(t, \bar{x}^{(1)}); \quad \bar{g}(\bar{x}_0, \bar{x}_l) = (1 - \lambda_1)\bar{d}_0. \quad (7.80)$$

Розв'язок цієї задачі можна знайти методом лінеаризації (див. § 7.3.), де за початкове наближення ітераційного процесу береться розв'язок крайової задачі при $\lambda = \lambda_0$, тобто $x^{(0)}(t)$. Таким же чином знаходимо розв'язок для всіх наступних значень λ_i ($i = 2, 3, \ldots, k-1$).

Якщо при $\lambda = \lambda_{k-1}$ знайдено розв'язок крайової задачі $x^{(k-1)}(t)$, то $x^{(k)}(t)$ при $\lambda = \lambda_k$ знаходимо з розв'язання крайової задачі

$$\frac{d\bar{x}^{(k)}}{dt} = \bar{f}(t, \bar{x}^{(k)}); \quad \bar{g}(\bar{x}_0, \bar{x}_l) = (1 - \lambda_k)\bar{d}_0, \quad (7.81)$$

де за початкове наближення приймається $x^{(k-1)}(t)$.

Продовжуючи цей процес до $k = m$ при $\lambda = \lambda_m = 1$, отримуємо розв'язок $x^{(m)}(t)$, що є розв'язком вихідної крайової задачі (7.57), (7.58). При такій реалізації методу продовження розв'язання по параметру обчислення буде стійким. При цьому слід зазначити, що такий підхід до розв'язання задачі Коші для функції $\bar{x} = \bar{x}(\lambda)$ потребує більшого обсягу обчислень

у порівнянні з методом Ейлера, але зате не виникає питань, пов'язаних із стійкістю обчислень.

Далі розглянемо підходи до розв'язання задачі для другого способу введення параметра. Застосуємо метод Ейлера для розв'язання задачі (7.78). Обчислювальна схема має вигляд

$$\frac{d\bar{x}^{(0)}}{dt} = A(t)\bar{x}^{(0)} + \bar{b}(t); \quad B\bar{x}_0^{(0)} + C\bar{x}_l^{(0)} = -\bar{d}_0;$$

$$\frac{d\bar{u}^{(k)}}{dt} = [(1-\lambda_k)A(t) - \lambda_k\Gamma_f(t, \bar{x}^{(k)})]\bar{u}^{(k)} + f(t, \bar{x}^{(k)}) - A(t)\bar{x}^{(k)} - \bar{b}(t);$$

$$[(1-\lambda_k)B + \lambda_k\Gamma_0(\bar{x}_0^{(k)}, \bar{x}_l^{(k)})]\bar{u}_0^{(k)} + [(1-\lambda_k)C + \lambda_k\Gamma_l(\bar{x}_0^{(k)}, \bar{x}_l^{(k)})]\bar{u}_l =$$

$$= B\bar{x}_0^{(k)} + C\bar{x}_l^{(k)} + \bar{d}_0 - \bar{g}(\bar{x}_0^{(k)}, \bar{x}_l^{(k)});$$

$$\bar{x}^{(k+1)}(t) = \bar{x}^{(k)}(t) + (\lambda_{k+1} - \lambda_k)\bar{u}^{(k)}(t), \quad k = 0, 1, 2, \ldots \qquad (7.82)$$

При цьому на кожному кроці ітераційного процесу (7.82) необхідно розв'язувати одну лінійну крайову задачу. Метод Ейлера застосуємо при стійких обчисленнях. В іншому випадку слід використати інший метод, що базується на поділі всього інтервалу для λ на малі відрізки й послідовному розв'язанні в окремих точках λ_k $(k = 0, 1, 2, \ldots, m)$ нелінійної крайової задачі

$$\frac{d\bar{x}^{(k)}}{dt} = A(t)\bar{x}^{(k)} + \bar{b}(t) + \lambda_k[f(t, \bar{x}^{(k)}) - A(t)\bar{x}^{(k)} - \bar{b}(t)];$$

$$B\bar{x}_0^{(k)} + C\bar{x}_l^{(k)} + \lambda_k[-B\bar{x}_0^{(k)} - C\bar{x}_l^{(k)} + \bar{d}_0 + \bar{g}(\bar{x}_0^{(k)}, \bar{x}_l^{(k)})] = d_0;$$

$$k = 0, 1, 2, \ldots, m, \qquad (7.83)$$

де за початкове наближення для $\bar{x}^{(k)}$ обирається $\bar{x}^{(k-1)}$; $\bar{x}^{(0)}$ вважається відомим при $\lambda = \lambda_0 = 0$. Такий процес завжди збігається.

Як приклад розглянемо деформацію жорстко закріпленої на зовнішньому контурі кільцевої пластини сталої товщини h під дією прикладеного на внутрішньому контурі поперечного зусилля Q_0 (див. § 7.2).

Граничні умови: при $x = x_0$

$$u = \theta_r = 0; \quad Q_r = Q_0;$$

при $x = x_1$

$$u = \omega = \theta_r = 0.$$

Використаємо перший спосіб введення параметра. За початковий вектор при розв'язанні задачі Коші обираємо вектор $\{0; 0; 0; 0; 20; 0\}^T$. Вихідні дані при розв'язанні задачі: $h = 1$; $x_0 = 0{,}2$; $x_1 = 1$; $Q_0 = 20$. Значення

прогину ω^*, зусилля N_r^*, моменту M_r^* в залежності від зміни параметра λ наведено в табл. 25.

Таблиця 25

λ	Наближення	ω^*	N_r^*	M_r^*	Q_r^*
0,0	—	0,0000	0,0000	0,0000	20
	0	0,0000	0,0000	0,0000	20
	1	− 0,3174	− 1,3022	−0,1358	20
0,2	2	− 0,1726	− 1,2711	0,1295	20
	3	− 0,1793	− 1,2929	0,1132	20
	4	− 0,1793	− 1,2928	0,1132	20
	0	− 0,1793	− 1,2928	0,1132	20
	1	− 0,3272	− 2,5005	0,3110	20
0,4	2	− 0,2022	− 2,6862	0,5743	20
	3	− 0,1977	− 2,7089	0,5862	20
	4	− 0,1977	− 2,7090	0,5863	20
	0	− 0,1977	− 2,7090	0,5863	20
	1	− 0,0296	− 3,8415	1,5542	20
0,6	2	0,2012	− 4,2117	2,2000	20
	3	0,2469	− 4,2286	2,3359	20
	4	0,2477	− 4,2271	2,3379	20
	0	0,2477	− 4,2271	2,3379	20
	1	1,7881	− 3,7655	7,0018	20
0,8	2	1,0643	− 3,9235	5,0280	20
	3	1,0963	− 2,0893	4,6454	20
	4	1,0985	− 1,9745	4,6701	20
	5	1,0980	− 1,9749	4,6689	20
	0	1,0980	− 1,9749	4,6689	20
	1	0,9179	1,0957	4,1562	20
1,0	2	0,8982	1,4446	4,1518	20
	3	0,8979	1,4463	4,1509	20
	4	0,8979	1,4463	4,1509	20

Одержані значення функцій при $\lambda = 1$ для четвертого наближення цілком збігаються з розв'язком задачі методом лінеаризації.

Як видно з таблиці, для проміжних значень λ (0,2; 0,4; 0,6; 0,8) розв'язок суттєво відрізняється від точного, але це не впливає на збіжність ітераційного процесу. Чисельні експерименти за допомогою методу продовження по параметру показують, що, змінюючи крок $\Delta\lambda$ і початкове наближення N_0 при деяких витратах машинного часу, можна завжди одержати розв'язок задачі.

§ 7.5. МЕТОД СКІНЧЕННИХ РІЗНИЦЬ

В основу цього методу покладена заміна всіх диференціальних співвідношень скінченнорізницевими.

Нехай потрібно знайти розв'язок $\bar{y}(t)$ крайової задачі (7.1), (7.2) або (7.1'), (7.2').

Інтервал $[0, l]$, в якому потрібно знайти розв'язок крайової задачі, розділимо на n однакових частин довжиною h, де $h = l/n$ (h називають кроком). Для спрощення вважатимемо, що крайова задача описується одним диференціальним рівнянням, до якого можна звести систему рівнянь (7.1) з відповідними граничними умовами (7.9), тобто:

$$F(t, y(t), y'(t), y''(t), \ldots, y^{(n)}(t)) = 0; \tag{7.84}$$

$$\Phi_s = (y_0, y_0', \ldots, y_0^{(n-1)}, y_l, y_l', \ldots, y_l^{(n-1)}) = 0 \quad (s = 1, \ldots, n), \tag{7.85}$$

де Φ_s і F — нелінійні вирази, а шукану функцію позначимо через $y(t)$. Точки ділення інтервалу $[0, l]$ позначимо через $t_i = ih$ $(i = 0, 1, 2, \ldots, n)$. Значення шуканої функції $y(t)$ в точках t_i позначимо через $y(t_i)$, а її наближені значення в цих точках — через y_i.

Складемо систему рівнянь для визначення наближених значень розв'язку y_i. Для цього розглядаємо диференціальне рівняння (7.84) в точці $t = t_i$ і замінюємо у ньому певним способом усі диференціальні співвідношення різницевими, які є лінійними виразами від y_i.

Таким чином, похідну $y'(t_i)$, яка при малому кроці h наближено дорівнює першому різницевому співвідношенню $(y(t_i + h) - y(t_i - h))/2h$, замінюємо через

$$(y_{i+1} - y_{i-1})/2h; \tag{7.86}$$

другу похідну $y''(t)$ — через

$$(y_{i+1} - 2y_i + y_{i-1})/h^2. \tag{7.87}$$

У загальному випадку вираз для заміни k-ї похідної можна одержати за допомогою різницевих операторів (гл. 1): похідну $y^{(k)}(t_i)$ парного порядку k можна замінити співвідношенням

$$\frac{1}{h^k}\Delta^k y_{i-k/2} \tag{7.88}$$

і похідну $y^{(p)}(t_i)$ непарного порядку p — середнім арифметичним двох p-х різницевих співвідношень

$$\frac{1}{2}\frac{1}{h^p}\left(\Delta^p y_{i-\frac{p+1}{2}} + \Delta^p y_{i-\frac{p-1}{2}}\right). \tag{7.89}$$

Легко бачити, що вирази (7.86) і (7.87) є окремими випадками відповідно виразів (7.89) і (7.88). Таким чином, диференціальне рівняння можна замінити в точці t_i скінченнорізницевим рівнянням, тобто алгебраїчним рівнянням для наближених значень розв'язку y_i.

Граничні умови (7.85) також можна замінити скінченнорізницевими рівняннями. Наприклад, для крайової задачі парного порядку k з k граничними умовами, які задані на кінцях інтервалу $t = 0$ і $t = l$, можна скласти скінченнорізницеве рівняння.

Таким чином, отримуємо $n + k + 1$ рівнянь для такої ж кількості невідомих $y_{-k/2}, y_{-k/2+1}, \ldots, y_{n+k/2}$.

Ці рівняння нелінійні, що відповідає вихідним диференціальним рівнянням (7.84) і (7.85). Методи розв'язання таких рівнянь викладені в гл. 3.

Зазначимо, що крім цього розглядуване диференціальне співвідношення можна по-різному заміняти скінченнорізницевими виразами і отримувати різні системи рівнянь для y. Так, у виразі $(fy')'$ можна спочатку виконати диференціювання і застосувати спосіб для заміни кожної похідної:

$$f_i' \frac{y_{i+1} - y_{i-1}}{2h} + f_i \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}.$$

Але можна, не виконуючи попередньо диференціювання у виразі $(fy')'$, замінити його p скінченнорізницевим співвідношенням відповідно з (7.86):

$$\frac{1}{2h}\left(f_{i+1} \frac{y_{i+2} - y_i}{2h} - f_{i-1} \frac{y_i - y_{i-2}}{2h}\right).$$

Для складання останнього виразу також можна застосувати вдвічі менший крок сітки і за наближений прийняти такий вираз:

$$\frac{f_{i+1/2}(y_{i+1} - y_i) - f_{i-1/2}(y_i - y_{i-1})}{h^2} = \frac{1}{h^2} \Delta(f_{i-1/2} \Delta y_{i-1}), \qquad (7.90)$$

де $f_{i-1/2} = f[a + (i - 1/2)h]$.

Таким чином, скінченнорізницеві рівняння, одержані різними способами, можуть відрізнятися одне від одного і приводити до різних результатів. Для апроксимації диференціальних рівнянь застосовуються і більш точні скінченнорізницеві методи вищого порядку.

Як приклад розглянемо таку нелінійну крайову задачу:

$$y'' = 3/2y^2, \quad y(0) = 4; \quad y(1) = 1. \qquad (7.91)$$

Крайова задача (7.91) має два розв'язки, один з яких виражається через елементарні функції

$$y = \frac{4}{(1+t)^2}, \tag{7.92}$$

а другий розв'язок — через еліптичні.

Розглянемо розв'язання крайової задачі (7.91) за допомогою скінченнорізницевого методу. Спочатку отримаємо наближений розв'язок цієї задачі при всяких значеннях кроку h. Так, при $h = 1/2$ одержуємо одне рівняння відносно невідомої y_1, яка є наближеним значенням для $y(1/2)$:

$$\frac{4 - 2y_1 + 1}{h^2} = \frac{3}{2} y_1^2,$$

звідки маємо

$$y_1 = 1/3\,(-8 \pm \sqrt{184}) = \begin{cases} 1{,}8549 \ (\varepsilon = 4{,}3\%), \\ -7{,}188. \end{cases}$$

При $h = 1/3$ для обох невідомих y_1 і y_2 дістанемо два нелінійних рівняння:

$$9(4 - 2y_1 + y_2) = 3/2y_1^2;$$

$$9(y_1 - 2y_2 + 1) = 3/2y_2^2.$$

У площині змінних y_1, y_2 цим рівнянням відповідають дві параболи (рис. 12), точки перетину яких і дають шукані наближені значення невідомих:

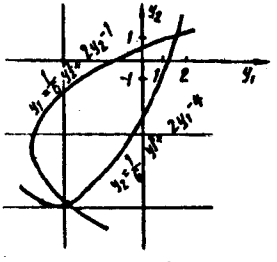


Рис. 12

$$y_1 = 2{,}2950 \ (\varepsilon = 2{,}0\%), \quad y_1 = -4{,}70;$$

$$y_2 = 1{,}4680 \ (\varepsilon = 1{,}9\%), \quad y_2 = -9{,}72.$$

При меншому значенні кроку h і відповідно більшому числі невідомих розв'язання системи рівнянь потребує більших зусиль. Для знаходження розв'язків з малим кроком h можна скористатися наближеними розв'язками для більшого значення h. Так, при $h = 1/5$ будемо виходити із значень $y_1 = 3$ і $y_2 = 2{,}8$ та з різницевого рівняння

$$y_{i+1} = 0{,}06y_i^2 + 2y_i - y_{i-1} \quad (i = 1, 2, 3, 4). \tag{7.93}$$

Одержані значення y_i наведені в перших двох стовпчиках табл. 26. Із значень на кінцях інтервалу 3,509 і 1,0577 лінійною інтерполяцією знаходимо значення y_1:

$$y_1 = 2{,}8 - \frac{2{,}8 - 3}{3{,}509 - 1{,}0577}\,0{,}0577 = 2{,}7953.$$

З цим значенням y_1 повторюємо обчислення (третій стовпчик табл. 26) і за допомогою повторної інтерполяції знаходимо значення $y_1 = 2{,}79464$. Результати обчислення для цього значення y_1 наведені в четвертому стовпчику таблиці, в якому для оцінки похибки дано також значення y_{-1}, яке визначається з (7.93) при $i = 0$. Аналогічний процес обчислення для другого розв'язку крайової задачі приводить до результатів, які наведені в останньому стовпчику таблиці. Наведені значення дозволяють судити про точність одержаних результатів розв'язку нелінійної крайової задачі.

Таблиця 26

i	y_i				
((-1) 0)	4	4	4	(6,16536) 4	4
1	3	2,8	2,7953	2,79464	−2,5138
2	2,54	2,0704	2,0594	2,05787	−8,6484
3	2,467	1,5980	1,5780	1,57519	−10,2953
4	2,759	1,2788	1,2460	1,24138	−5,5826
5	3,509	1,0577	1,0071	1,00003	1

§ 7.6. МЕТОД РІТЦА

Метод Рітца, який відноситься до варіаційних, де крайовій задачі для звичайного диференціального рівняння ставиться у відповідність задача варіаційного числення — про мінімум (максимум) функціонала, який звичайно виражається через інтеграл від шуканої функції і в деяких випадках від інших операцій над цією функцією, тобто нелінійній крайовій задачі (7.84), (7.85):

$$I[\varphi] = \int_0^l F(t, \varphi, \varphi', \ldots, \varphi^{(m)})dt, \tag{7.94}$$

і серед всіх неперервно диференційовних функцій φ, які задовольняють граничні умови (7.85), потрібно знайти таку, для якої інтеграл (7.94) набуває найменшого значення. Більш детально ці питання для лінійної крайової задачі розглянуті в гл. 6.

Метод Рітца полягає в тому, що з якихось міркувань вибирається сім'я функцій φ, які залежать від кількох параметрів $a_1, a_2, \ldots, a_p$ у вигляді

$$\varphi = \varphi(t, a_1, a_2, \ldots, a_p), \tag{7.95}$$

де взагалі вираз (7.95) є нелінійна функція від параметрів a_i $(i = 1, \ldots, p)$. При цьому вважається, що функції φ задовольняють граничну умову (7.85).

Якщо функцію φ підставити в інтеграл (7.94), то функціонал $I[\varphi]$ перетворюється в функцію параметрів $a_1, a_2, \ldots, a_p$. Ці параметри під-

бираються так, щоб функція досягала екстремуму, тобто щоб функціонал $I[\varphi]$, який розглядається тепер як функція параметрів $a_1, a_2, \ldots, a_p$, набував мінімального значення, тобто від необхідності знаходження мінімуму функціоналу переходимо до знаходження мінімуму функції p змінних. Звідси випливає, що параметри $a_1, a_2, \ldots, a_p$ повинні бути визначені із системи нелінійних рівнянь

$$\frac{\partial I[\varphi(a_1, a_2, \ldots, a_p)]}{\partial a_i} = 0 \quad (i = 1, \ldots, p). \tag{7.96}$$

Цю систему рівнянь можна розв'язати за допомогою методів, викладених в гл. 3. Знайшовши значення невідомих параметрів $a_1, a_2, \ldots, a_p$ і підставивши їх у вираз (7.95), отримуємо розв'язок розглядуваної задачі.

Як приклад розглянемо нелінійну крайову задачу, що викладена в § 7.5:

$$y'' = 3/2\, y^2, \quad y(0) = 4; \quad y(1) = 1. \tag{7.97}$$

Цій крайовій задачі, згідно з (7.94), відповідає варіаційна задача для функціоналу

$$I[\varphi] = \int_0^1 [(\varphi')^2 + \varphi^3]dt. \tag{7.98}$$

Невідому функцію φ шукаємо у вигляді

$$\varphi = 4 - 3t + a_1(t - t^2) + a_2(t - t^3), \tag{7.99}$$

яка для довільних a_1 і a_2 задовольняє граничну умову. З виразу (7.98) отримуємо рівняння (7.96)

$$\frac{\partial I}{\partial a_1} = \int_0^1 [2\varphi'(1 - 2t) + 3\varphi^2(t - t^2)]dt = 0;$$

$$\frac{\partial I}{\partial a_2} = \int_0^1 [2\varphi'(1 - 3t^2) + 3\varphi^2(t - t^3)]dt = 0. \tag{7.100}$$

Після підстановки в рівняння (7.100) функції φ з виразу (7.99) та інтегрування дістанемо систему нелінійних алгебраїчних рівнянь:

$$1407 + 490a_1 + 726a_2 + 9a_1^2 + 27a_1a_2 + 41/2a_2^2 = 0;$$

$$1302 + 484a_1 + 750a_2 + 9a_1^2 + 82/3a_1a_2 + 21a_2^2 = 0. \tag{7.101}$$

Перетворимо її:

$$a_1 = \frac{-105 + 24a_2 + 1/2a_2^2}{6 - 1/3a_2};$$

$$a_2 = \frac{2856 + 368a_1 + 9/2a_1^2}{129 - 20/3a_1}. \qquad (7.102)$$

Розв'язуючи систему (7.102) графічно, в площині змінних a_1a_2 (рис. 13) маємо дві гіперболи, які перетинаються в двох точках.

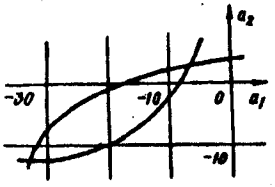


Рис. 13

Таблиця 27

t	$y^{(1)}$	$y^{(2)}$
0,25	2,56197	− 5,75
0,50	1,76270	− 10,28
0,75	1,31701	− 8,43

Розв'язки системи набувають значень

$$a_1^{(1)} = -7{,}07004;\ a_2^{(1)} = 2{,}72044;$$

$$a_1^{(2)} = -32{,}20;\ a_2^{(2)} = -12{,}60.$$

Значення шуканої функції $y(t)$ наведені в табл. 27. При цьому похибка не перевищує 1%. А у випадку одночленної апроксимації функції φ ($a_2 = 0$) похибка розв'язку досягає 10%.

ГЛАВА 8

МЕТОДИ РОЗВ'ЯЗАННЯ КРАЙОВИХ ЗАДАЧ ДЛЯ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ У ЧАСТИННИХ ПОХІДНИХ

§ 8.1. ВСТУПНІ ЗАУВАЖЕННЯ. ПОСТАНОВКА ОСНОВНИХ ЗАДАЧ

Диференціальні рівняння в частинних похідних та інтегральні рівняння використовуються в різноманітних галузях природознавства. Отримувати їх розв'язок у явному вигляді вдається лише в найпростіших випадках. У зв'язку з цим особливого значення набувають наближені методи розв'язання різних задач для диференціальних рівнянь в частинних похідних, систем диференціальних рівнянь в частинних похідних та інтегральних рівнянь (тобто задач математичної фізики). Нами розглядаються деякі, найбільш поширені методи чисельного розв'язання задач математичної фізики — в основному методи чисельного розв'язування задач лінійних диференціальних рівнянь в частинних похідних другого порядку з двома незалежними змінними. Викладання методів у випадку довільної кількості змінних вимагає громіздких записів, в той час як головні ідеї методів добре зрозумілі й у простіших випадках.

Наближені методи чисельного розв'язання різних задач для диференціальних рівнянь в частинних похідних можна розбити на три групи:

1) методи, у яких наближений розв'язок отримується в аналітичній формі, наприклад, у вигляді відрізку деякого ряду;

2) методи, за допомогою яких можна дістати таблицю наближених значень шуканого розв'язку в деяких точках розглядуваної області;

3) методи, що базуються на поєднанні аналітичних перетворень та чисельних методів.

До першої групи методів відноситься, перш за все, метод Фур'є розв'язання крайових задач для диференціальних рівнянь у частинних похідних, точний розв'язок за яким отримується у вигляді деякого ряду; при цьому за наближений може вважатися сума деякої кількості його перших членів. До цієї групи відносяться й варіаційні методи розв'язання крайових задач для диференціальних рівнянь в частинних похідних, в основу яких покладено заміну крайової задачі для диференціального рівняння еквівалентною їй варіаційною. Наближене розв'язання крайової задачі зводиться до побудови наближеного розв'язку відповідної варіаційної задачі.

До другої групи методів відносяться скінченнорізницеві методи чисельного розв'язання задач для диференціальних рівнянь в частинних похідних. Відповідно до них диференціальна задача замінюється скінченнорізницевою і різницевий розв'язок визначається на сітці.

Методи прямих, сплайн-апроксимацій, граничних елементів та інші в загальному випадку становлять третю групу методів, у яких на першому етапі відбуваються різні аналітичні перетворення, що перетворюють вихідну крайову задачу в одновимірну або в систему лінійних алгебраїчних рівнянь, а на другому етапі останні реалізують чисельно. У методі прямих шукається наближений розв'язок диференціального рівняння в частинних похідних вздовж деякої сім'ї прямих. При цьому замість диференціального рівняння в частинних похідних виникає система звичайних диференціальних рівнянь, яка розв'язується чисельними методами. Застосовуючи метод сплайн-апроксимацій, наближений розв'язок вихідного диференціального рівняння в частинних похідних записується у вигляді деякого ряду і наближений розв'язок шукається на всій області визначення розв'язання задачі. У методі граничних елементів головним є залежність між значеннями шуканих функцій в середині розглядуваної області та їх значеннями на границі. Встановлюється ця залежність переходом від диференціальних рівнянь до інтегральних співвідношень, що реалізуються чисельно.

Розглянемо постановку задач, які описуються диференціальними рівняннями в частинних похідних. Математичні моделі суцільних тіл призводять до рівнянь в частинних похідних. Наприклад, зміна температури у нерухомому тілі описується рівнянням теплопровідності

$$c(u, \bar{r}, t)\,\frac{\partial u}{\partial t} = \operatorname{div}\,[k(u, \bar{r}, t)\,\operatorname{grad} u] + q(u, \bar{r}, t), \tag{8.1}$$

де u — температура; c — теплоємність; k — коефіцієнт теплоємності; q — густина джерел тепла. Незалежними змінними у фізичних задачах звичайно є час t та координата $\bar{r}$; можуть бути й інші змінні, наприклад, швидкість частинок $\bar{v}$ у задачах переносу. Розв'язок потрібно шукати в деякій області зміни незалежних змінних $G(t, \bar{r}, \bar{v}, \ldots)$. Повна математична постановка задачі містить диференціальне рівняння, а також додаткові умови, що дозволяють з семи розв'язків диференціального рівняння обрати один. Додаткові умови, як правило, задаються на границі області G.

Якщо однією з незалежних змін є t, то найчастіше розглядають область вигляду

$$G(t, \bar{r}, \ldots) = g(\bar{r}, \ldots) \times [t_0, T],$$

тобто розв'язок шукають у деякій просторовій області $g(\bar{r}, \ldots)$ на відрізку часу $t_0 \le t \le T$. У цьому випадку додаткові умови, що задано при $t = t_0$, називаються початковими, а додаткові умови, що задано на границі $\Gamma(\bar{r})$ області $g(\bar{r})$, — *граничними* або *крайовими*.

Задача, яка має лише початкові умови, називається *задачею Коші*. Наприклад, для рівняння теплопровідності (8.1) у необмеженому просторі можна поставити задачу з початковими умовами

$$u(t_0, \bar{r}) = \mu(\bar{r}). \tag{8.2}$$

Якщо $\mu(\bar{r})$ — кусково-неперервна обмежена функція, то розв'язок задачі (8.1), (8.2) є єдиним у класі обмежених функцій.

Задача з початковими і граничними умовами називається *мішаною крайовою задачею* або *нестаціонарною крайовою задачею*. Для рівняння (8.1) додаткові умови такої задачі можуть мати, наприклад, вигляд

$$u(t_0, \bar{r}) = \mu(\bar{r}),\ \bar{r} \in g(\bar{r}),\ u(t, \bar{r})_\Gamma = \mu_1(t, \bar{r}),\ t_0 \le t \le T. \qquad (8.3)$$

Для цього рівняння припустимі й інші граничні умови, наприклад, які містять нормальну похідну розв'язку.

При дослідженні усталених станів або стаціонарних (що не залежать від часу) процесів у суцільному середовищі формулюються математичні задачі, що не залежать від часу. Їх розв'язок шукається в області $g(\bar{r})$, а додаткові умови граничні. Такі задачі називаються *крайовими*.

Ми обмежимося розглядом коректно поставлених задач, коли для деякого класу початкових і граничних даних розв'язок (у заданому класі функцій) існує, однозначно й неперервно залежить від цих даних. Вважатимемо, що розв'язок неперервно залежить від коефіцієнтів у рівнянні.

Наведемо класифікацію розглядуваних рівнянь другого порядку, що мають вигляд

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + F = 0. \qquad (8.4)$$

Коефіцієнти рівняння (8.4) залежать взагалі від u, x, y. Якщо коефіцієнти не залежать від змінних, то це буде лініине рівняння зі сталими коефіцієнтами; якщо F лінійно залежить від u, а інші коефіцієнти від u не залежать, — лінійне рівняння зі змінними коефіцієнтами. Якщо коефіцієнти залежать від u, то рівняння (8.4) називається *квазілінійним*.

При $A \equiv B \equiv C \equiv 0$, але $D \neq 0$ та $E \neq 0$, то рівняння (8.4) має перший порядок і називається рівнянням переносу,

Рівняння другого порядку класифікуються за знаком дискримінанту $B^2 - AC$: в гіперболічних рівняннях дискримінант додатний, в параболічних — дорівнює нулю, в еліптичних — від'ємний.

Фізичні процеси, які описуються вищезгаданими типами рівнянь, істотно різняться одні від інших. Відповідно повні постановки задач для цих типів рівнянь мають свої особливості, докладно розглянуті в праці (5). Коротко про це буде йти мова у відповідних параграфах цієї глави.

§ 8.2. РІВНЯННЯ ПАРАБОЛІЧНОГО ТИПУ. ЯВНІ ТА НЕЯВНІ СКІНЧЕННОРІЗНИЦЕВІ МЕТОДИ

Розглянемо параболічне рівняння

$$\frac{\partial u}{\partial t} = C\frac{\partial^2 u}{\partial t^2}, \qquad (8.5)$$

прототипом якого є рівняння теплопровідності.

Покажемо, яким чином рівняння (8.5) виступає як математична модель поширення тепла. Розглянемо довгий стержень, який витягнуто вздовж осі x (рис. 14). Вважатимемо, що стержень повністю теплоізольований, за винятком, можливо, кінців, і потік тепла може поширюватися лише в напрямі осі x. Нехай $u(t, x)$ — температура стержня (в кельвінах) у точці x у момент часу t та a — площа поперечного перерізу стержня. З елементарної фізики відомо, що кількість тепла, що надходить за одиницю часу крізь переріз, перпендикулярний осі стержня, є — kau_x, де $k > 0$ — коефіцієнт теплопровідності. Таким чином, якщо градієнт температури u_x у даному перерізі від'ємний, тобто температура зліва вища, ніж справа, то тепло крізь цей переріз буде проходити зліва направо. Отже, якщо ми розглянемо елемент стержня довжини x, то за одиницю часу до цього елемента через переріз x надходить кількість тепла, що дорівнює $(-kau_x)|_x$ та виходить крізь переріз $x + \Delta x$ кількість тепла, що дорівнює $(-kau_x)|_{x+\Delta x}$, тобто кількість тепла в елементі змінюється на

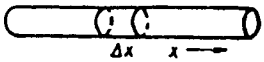


Рис. 14

$$(-kau_x)|_x - (-kau_x)|_{x+\Delta x}. \tag{8.6}$$

З іншого боку, як відомо, кількість тепла, яку має елемент, пропорційна масі елемента та його температурі; точніше, воно дорівнює $sa\Delta x\rho u_t$, де s — питома теплоємність матеріалу, а ρ — його густина (маса на одиницю об'єму).

З цього випливає, що похідна за часом від кількості тепла в елементі буде дорівнювати кількості тепла за формулою (8.6), тобто

$$sa\Delta x\rho u_t = (kau_x)|_{x+\Delta x} - (kau_x)|_x.$$

При переході до границі по $x \to \infty$ дістанемо

$$u_t = \frac{1}{s\rho}(ku_x)|_x. \tag{8.7}$$

Якщо k не залежить від x, то це рівняння перетворюється на (8.5), де $c = k/s\rho$. Отже, похідна від температури за часом у такому тонкому стержні пропорційна другій похідній від температури за просторовою змінною.

Як і у випадку звичайних диференціальних рівнянь, для виділення розв'язку рівняння в частинних похідних необхідно задати деякі початкові та граничні умови. Нехай у розглядуваній задачі про тонкий стержень у початковий момент часу $t = 0$ нам відомий розподіл $g(x)$ температури у стержні. Вважаємо, що довжина стержня дорівнює 1, його лівий кінець знаходиться в точці $x = 0$. Припускаючи також, що кінці стержня утримуються при постійній температурі, тобто

$$u(t, 0) = \alpha, \quad u(t, 1) = \beta, \tag{8.8}$$

де α і β — задані сталі. Співвідношення (8.8) є граничними умовами за просторовою змінною x. Як початкову умову використовуємо заданий початковий розподіл температури

$$u(0, x) = g(x), \quad 0 \le x \le 1. \tag{8.9}$$

Таким чином, можна сформулювати задачу з фізичної точки зору. Маємо стержень, кінці якого утримуються при фіксованих температурах α і β. Вважаючи початковий розподіл температури вздовж стержня відомим, слід знайти температуру будь-якої точки стержня x у довільний момент часу $t > 0$. Математичною моделлю цієї проблеми є диференціальне рівняння в частинних похідних

$$\frac{\partial u}{\partial t} = c \frac{\partial^2 u}{\partial x^2}, \quad c = k/(s\rho), \quad 0 \le x \le 1,$$

із граничними (8.8) та початковими умовами (8.9).

Інший варіант формулювання задачі пов'язаний з врахуванням неоднорідності матеріалу стержня. Якщо, наприклад, стержень має сплав, склад якого повільна змінна функція від x, тоді густина, теплопровідність і теплоємність стержня також, взагалі кажучи, будуть функціями від x. Отже, задача тепер буде описуватися рівнянням (8.7), де $\rho = \rho(x)$, $s = s(x)$ і $k = k(x)$. Це вже рівняння не з постійними, а зі змінними коефіцієнтами. Можна взяти до уваги, що в загальному випадку коефіцієнт теплопровідності залежить не тільки від матеріалу, але й від температури та рівняння (8.7) стає нелінійним.

Зазначимо, що рівняння теплопровідності є математичною моделлю і для цілого ряду інших фізичних явищ, як от, наприклад, дифузія газу.

У цій главі починається вивчення скінченнорізницевих методів розв'язку рівнянь у частинних похідних. Розглянемо мішану задачу для рівняння теплопровідності зі сталими коефіцієнтами. В області $\{0 < x < 1, \ 0 < t \le T\}$ потрібно знайти розв'язок рівняння

$$\frac{\partial u}{\partial t} = c \frac{\partial^2 u}{\partial x^2} + f(t, x), \tag{8.10}$$

що задовольняє початкову умову

$$u(0, x) = g(x), \quad 0 \le x \le 1 \tag{8.11}$$

і граничну умову

$$u(t, o) = \alpha(t), \quad u(t, 1) = \beta(t), \tag{8.12}$$

де $g(x)$, $\alpha(t)$, $\beta(t)$ — задані функції. Як відомо, при означених припущеннях щодо гладкості розв'язок задачі (8.10) — (8.12) існує та єдиний. При вивченні апроксимації різницевими схемами припустимо, що розв'язок $u(t)$ має необхідну кількість похідних за t і x. Розв'язок задачі (8.10) — (8.12) неперервно залежить від початкових та граничних даних.

При побудові різницевої схеми введемо сітку в області змінних і задамо шаблон, тобто множину точок сітки, що бере участь в апроксимації диференціального виразу. Сітка вводиться за змінною x із кроком $h = \Delta x$ таким чином:

$$\omega_h = \{x_j = j\Delta x = jh, \quad j = 0, 1, \ldots, N, \quad hN = 1\},$$

а за змінною t із кроком $\tau = \Delta t$ (позначимо її $\omega_\tau = \{t_m = m\Delta t = m\tau, \quad m = 0, 1, \ldots, K, \quad K\tau = T\}$).

Точки (t_m, x_j) $j = 0, 1, \ldots, N$, $m = 0, 1, \ldots, K$ утворюють вузли просторово-часової сітки $\omega_{\tau,h} = \omega_\tau \times \omega_h$ (див. рис. 15). Вузли (t_m, x_j), що належать відрізкам $I_0 = \{0 \le x \le 1, \ t = 0\}$, $I_1 = \{x = 0, \ 0 \le t \le T\}$ $I_2 = \{x = 1, \ 0 \le t \le T\}$, відносяться до граничних вузлів сітки $\omega_{\tau,h}$, а всі інші вузли — внутрішні. На рис. 15 *1* — граничні вузли, *2* — внутрішні. Варіанти шаблонів наведено на рис. 16 (*а* — явна схема; *б* — суто неявна схема, *в* — симетрична схема; *г* — тришарова схема).

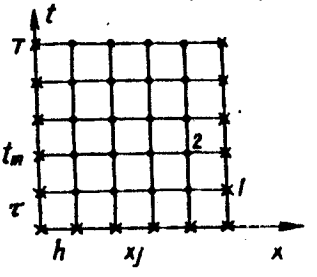


Рис. 15

Шаром називається множина всіх вузлів сітки $\omega_{\tau,h}$, що мають одну й ту ж часову координату. Таким чином, m-м шаром буде множина вузлів (t_m, x_0), (t_m, x_1), ..., (t_m, x_N). Для функції $u(t, x)$, визначеної на сітці $\omega_{\tau,h}$, введемо позначення:

$$u_j^m = u(t_m, x_j);$$

$$\frac{\partial u}{\partial t} = \frac{u_j^{m+1} - u_j^m}{\Delta t}; \tag{8.13}$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{j+1}^m - 2u_j^m + u_{j-1}^m}{(\Delta x^2)}. \tag{8.14}$$

Для апроксимації рівняння (8.10) у точці (t_m, x_j) введемо шаблон (рис. 16,*а*), що складається з чотирьох вузлів $(t_m, x_{j\pm1})$, (t_m, x_j), (t_{m+1}, x_j). Похідну $\partial u/\partial t$ замінимо в точці (t_m, x_j) різницевим співвідношенням (8.13), а $\partial^2 u/\partial t^2$ — різницевою похідною (8.14). Праву частину $f(x, t)$ заміним мо наближеною сітчастою функцією φ_j^m, якою може бути один із виразів:

$$f(t_m, x_j), \quad \frac{1}{\Delta x}\int_{x_j - 1/2}^{x_j + 1/2} f(t_m, x)dx;$$

$$\frac{1}{\Delta x \Delta t} \int_{t_m}^{t_{m+1}} dt \int_{x_j - 1/2}^{x_j + 1/2} f(t, x) dx.$$

Результатом буде різницеве рівняння

$$\frac{u_j^{m+1} - u_j^m}{\Delta t} = c \frac{u_{j+1}^m - 2u_j^m + u_{j-1}^m}{(\Delta x)^2} + \varphi_j^m, \qquad (8.15)$$

що апроксимує початкове диференціальне рівняння в точці (t_m, x_j) за першим порядком по τ і другим порядком по h при умові, що різниця $\varphi_j^m - f(t_m, x_j)$ має той самий порядок мализни.

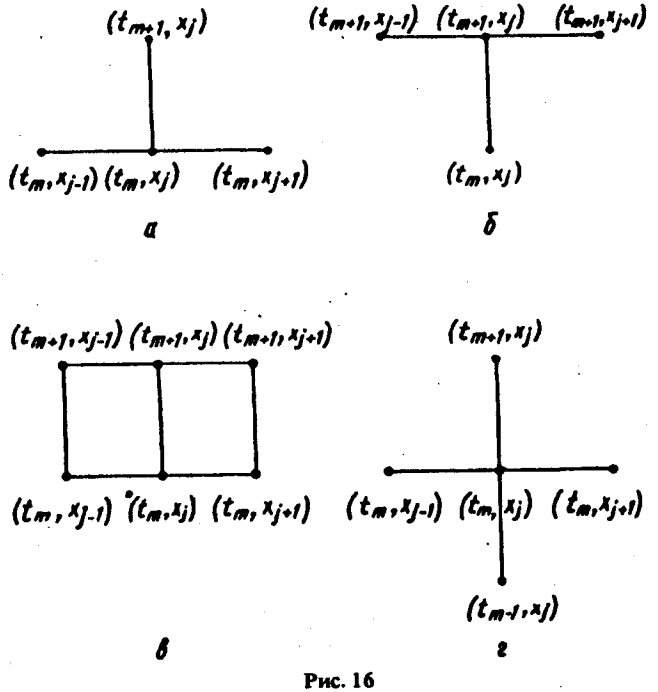


Рис. 16

Різницевою схемою є сукупність різницевих рівнянь, що апроксимують первісне диференціальне рівняння в усіх внутрішніх вузлах сітки, та додаткові (початкові і граничні) умови — в граничних вузлах сітки. Різницеву схему за аналогією з диференціальною задачею назвемо *різницевою задачею.* У даному випадку різницева схема має вигляд

$$\frac{u_j^{m+1} - u_j^m}{\Delta t} = c \frac{u_{j+1}^m - 2u_j^m + u_{j-1}^m}{(\Delta x)^2} + \varphi_j^m; \qquad (8.16)$$

$$j = 1, 2, \ldots, N-1, \quad m = 0, 1, \ldots, k-1, \quad hN = 1, \quad k\tau = T;$$

$$u_0^m = \mu_1(t_m), \quad u_N^m = \mu_2(t_m), \quad m = 0, 1, \ldots, K,$$

$$u_j^0 = g(x_j), \quad j = 1, 2, \ldots, N-1. \tag{8.17}$$

Ця схема є системою лінійних алгебраїчних рівнянь, кількість яких дорівнює кількості невідомих. Розв'язувати таку систему слід за шарами. Розв'язок на нульовому шарі заданий початковими умовами $u_j^0 = g(x_j)$, $j = 1, 2, \ldots, N-1$. Якщо розв'язок u_j^m, $j = 0, 1, \ldots, N$ на шарі m вже знайдений, то розв'язок u_j^{m+1} на $m+1$ шарі відшукується за явною формулою

$$u_j^{m+1} = u_j^m + \mu(u_{j+1}^m - 2u_j^m + u_{j-1}^m) + \varphi_j^m, \quad j = 1, 2, \ldots, N-1; \tag{8.18}$$

$$\mu = c\Delta t/(\Delta x)^2. \tag{8.19}$$

Значення $u_0^{m+1} = \mu_1(t_{m+1})$, $u_N^{m+1} = \mu_2(t_{m+1})$ остаточно можна визначити з граничних умов. Тому схема (8.18) називається *явною різницевою схемою.*

Похибка різницевої схеми (8.18) визначається як різниця між розв'язками вихідної задачі (8.10) — (8.12) та задачі (8.16). Нехай $u(t, x)$ — точний розв'язок рівняння (8.10) із початковою умовою (8.11) та граничними умовами (8.12). Якщо підставити цей точний розв'язок у різницеве рівняння (8.16), то воно задовольниться не повністю, а з деякою похибкою, яка називається *локальною похибкою дискретизації.* Таким чином, локальна похибка дискретизації у точці (t, x) дорівнює

$$e = \frac{u(t+\Delta t, x) - (t, x)}{\Delta t} - \frac{c}{(\Delta x)^2}[u(t, x+\Delta x) - 2u(t, x) + u(t, x-\Delta x)]. \tag{8.20}$$

Припустимо, що нам відомо значення точного розв'язку $u(t, x)$ для деякого t і всіх $x \in [0, 1]$ і ми бажаємо використати (8.16) при $\varphi_j^m = 0$ для отримання наближеного розв'язку в момент $t + \Delta t$. Якщо позначити цей наближений розв'язок як $\hat{u}(t + \Delta t, x)$, тоді

$$o = \frac{\hat{u}(t+\Delta t, x) - u(t, x)}{\Delta t} - \frac{c}{(\Delta x)^2}[u(t, x+\Delta x) - 2u(t, x) + u(t, x-\Delta x)].$$

Віднімаючи цю рівність від (8.20), дістанемо

$$\hat{u}(t + \Delta t, x) - u(t + \Delta t, x) = e\Delta t. \tag{8.21}$$

Таким чином, похибка, яку припустили на одному кроці за часом по різницевій схемі (8,16), дорівнює локальній похибці дискретизації, помноженій на Δt. Значення e у (8.20) легко оцінити за допомогою Δt і Δx.

Дійсно, якщо вважати u функцією тільки від t, вважаючи x фіксованим, то із розвинення в ряд Тейлора

$$u(t + \Delta t, x) = u(t, x) + u_t'(t, x)\Delta t + 0[(\Delta t)^2]$$

отримаємо

$$\frac{u(t + \Delta t, x) - u(t, x)}{\Delta t} = u_t'(t, x) + 0(\Delta t). \qquad (8.22)$$

Розглянемо u як функцію від x, вважаючи t фіксованим. Розвиваючи в ряд Тейлора $u(t, x + \Delta x)$, $u(t, x - \Delta x)$, матимемо

$$u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x) = u_x''(t, x)(\Delta x)^2 + 0[(\Delta x)^4],$$

отже, й

$$\frac{u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)}{(\Delta x)^2} = u_x''(t, x) + 0((\Delta x)^2). \qquad (8.23)$$

Якщо тепер підставити (8.22) і (8.23) у (8.20) і скористатися тим, що $\partial u/\partial t = c\,\partial^2 u/\partial x^2$ (оскільки u — точний розв'язок диференціального рівняння), отримаємо

$$e = 0(\Delta t) + 0[(\Delta x)^2]. \qquad (8.24)$$

Це говорить про те, що скінченнорізницевий метод (8.16) має перший порядок точності за часом і другий порядок точності за просторовою змінною. Вираз (8.24) дає нам похибку наближеного розв'язку на одному кроці за часом. Доведення того, що похибка дискретизації наближається до нуля на всьому відрізку $[0, t]$, достатньо складне і в загальному випадку вимагає додаткових даних про характер наближення до нуля Δt і Δx. Ствердження про те, що локальна похибка дискретизації (8.24) при $\Delta t \to 0$, $\Delta x \to 0$ наближається до нуля є, по суті, необхідною умовою наближення до нуля глобальної похибки дискретизації і називається умовою узгодженості різницевої схеми. Те, що із узгодженості різницевого методу не обов'язково випливає збіжність наближеного розв'язку до точного, пов'язане з проблемою стійкості різницевої схеми. Деякі аспекти цієї проблеми розглянемо на прикладі явної різницевої схеми (8.18). Отримаємо точний розв'язок різницевого рівняння (8.18), що задовольняє граничним і початковим умовам

$$u_0^m = u_n^m, \quad m = 0, 1, \ldots; \quad u_j^0 \text{ — задане}, \; j = 0, 1, \ldots, N - 1. \qquad (8.25)$$

Розв'яжемо за допомогою методу відокремлення змінних і зображення розв'язку у вигляді розвинення в ряди Фур'є. Припустимо, що розв'язок u_j^m може бути зображений у вигляді

$$u_j^m = v_m w_j, \quad j = 1, 2, \ldots, N - 1, \quad m = 0, 1, \ldots \qquad (8.26)$$

Ця формула є прикладом відокремлення змінних для різницевих рівнянь. Підставляючи (8.26) у (8.18) і групуючи члени, дістанемо

$$\frac{v_{m+1} - v_m}{\mu v_m} = \frac{w_{j+1} - 2w_j + w_{j-1}}{w_j}, \quad j = 1, 2, \ldots, N-1, \quad m = 0, 1, \ldots$$

У цьому рівнянні ліва частина не залежить від j, а права — від m; отже, обидві частини повинні дорівнювати деякій сталій, нехай $-\lambda$. Таким чином,

$$v_{m+1} - v_m = -\lambda\mu v_m, \quad m = 0, 1, \ldots; \tag{8.27}$$

$$w_{j+1} - 2w_j + w_{j-1} = -\lambda w_j, \quad j = 1, 2, \ldots, N-1, \tag{8.28}$$

де з граничних умов (8.25) $w_0 = w_N = 0$. Рівняння (8.28) являє собою задачу на власні значення для тридіагональної матриці

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & & & \\ . & . & & . & . \\ & & & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

Для того, щоб матриця мала саме такий вигляд, була обрана стала λ із знаком мінус. Ця матриця має власні значення

$$\lambda_k = 2 - 2\cos\frac{k\pi}{N}, \quad k = 1, \ldots, N-1, \tag{8.29}$$

і відповідні власні вектори

$$w_k = [\sin(k\pi\Delta x), \sin(2k\pi\Delta x), \ldots, \sin((N-1)k\pi\Delta x)],$$
$$k = 1, \ldots, N-1, \tag{8.30}$$

де $\Delta x = 1/N$. Таким чином, при кожному $\lambda = \lambda_k$ значення

$$w_j = \sin(jk\pi\Delta x), \quad j = 0, 1, \ldots, N$$

задовольняють (8.28). У той же час (8.28) при будь-якому λ має розв'язок $v_m = (1 - \lambda\mu)^m v_0$, $m = 0, 1, \ldots$.

Таким чином, співвідношення

$$v_m w_j = (1 - \lambda_k\mu)^m \sin(jk\pi\Delta x), \quad m = 0, 1, \ldots, \quad j = 0, 1, \ldots, N$$

при кожному k визначають розв'язок (8.18). Як і у випадку диференціальних рівнянь, лінійна комбінація цих розв'язків також є розв'язком. З цього випливає, що

$$u_j^m = \sum_{k=1}^{N-1} a_k(1 - \lambda_k\mu)^m \sin(jk\pi l\Delta x),$$
$$m = 0, 1, \ldots, \quad j = 0, 1, \ldots, N, \tag{8.31}$$

є розв'язком (8.18) при будь-якому значенні a_k. Якщо a_k обчислити за формулами

$$a_k = \sum_{l=1}^{N-1} g(x_l) \sin(k\pi l \Delta x), \quad k = 1, \ldots, N-1, \tag{8.32}$$

то розв'язок u_j^m буде задовольняти початкову умову

$$u_j^0 = g(x_j), \quad j = 1, 2, \ldots, N-1. \tag{8.33}$$

Розглянемо (8.31) з іншого боку. Рівняння $\partial u/\partial t = c\partial^2 u/\partial x^2$ з граничними умовами $u(t, 0) = u(t, 1) = 0$ є математичною моделлю задачі про розподіл температури в тонкому ізольованому стержні, кінці якого утримуються при нульовій температурі. За відсутністю джерел тепла очікується, що температура усіх точок стержня буде наближатися до нуля, тобто що $u(t, x) \to 0$ при $t \to \infty$. Природно вимагати, щоб скінченнорізницеві апроксимації u_j^m наближалися до нуля при $m \to \infty$. Із (8.31) випливає, що при довільних початкових умовах це буде мати місце в тому і лише в тому випадку, якщо

$$|1 - \mu\lambda_k| < 1, \quad k = 1, \ldots, N-1. \tag{8.34}$$

Оскільки μ і всі λ_k додатні, (8.34) буде виконуватися в тому і лише в тому випадку, коли $-(1 - \mu\lambda_k) < 1$, $k = 1, \ldots, N-1$, або, якщо

$$\mu \le \min_k \frac{2}{\lambda_k} = \left(1 - \cos\frac{\pi(N-1)}{N}\right)^{-1} = \left(1 + \cos\frac{\pi}{N}\right)^{-1}. \tag{8.35}$$

Використано те, що найбільшим із λ_k є λ_{N-1}. Оскільки $\mu = c\Delta t/(\Delta x)^2$, із (8.34) дістанемо

$$\Delta t < (\Delta x)^2 / \left[c\left(1 + \cos\frac{\pi}{N}\right)\right]. \tag{8.36}$$

Умова (8.36) дає обмеження на співвідношенні кроків Δt і Δx. Якщо цю умову не буде виконано, то у загальному випадку отриманий за різницевою схемою (8.18) наближений розв'язок u_j^m буде розбігатися при $m \to \infty$ і, що легко побачити, буде все гірше апроксимувати збіжний до нуля розв'язок диференціального рівняння.

Можна замінити (8.36) більш сильнішою умовою

$$\Delta t \le (\Delta x)^2/(2c), \tag{8.37}$$

із якої завжди випливає (8.36). Умова (8.37) називається *умовою стійкості різницевої схеми* (8.18). Різницеві схеми, стійкі лише при деякому обмеженні на співвідношення кроків за простором і часом, називаються

умовно стійкими. Отже, схема (8.18) умовно стійка; при $c = 1$ умова стійкості має вигляд $\tau/h^2 \le 0{,}5$. Найзручніший вигляд схема (8.18) має при $\mu = 1/2$

$$u_j^{m+1} = (u_{j+1}^m + u_{j-1}^m)/2 + \Delta t \varphi_j^m; \tag{8.38}$$

при $\mu = 1/6$

$$u_j^{m+1} = \frac{1}{6}(u_{j+1}^m + 4u_j^m + u_{j-1}^m) + \Delta t \varphi_j^m. \tag{8.39}$$

Оцінки похибок наближених розв'язків, отриманих із рівняння (8.38), (8.39) для задачі (8.10)—(8.12), мають відповідно вигляд

$$|u - \hat{u}| \le \frac{T}{4}\left(M_2 + \frac{1}{3}M_4\right)(\Delta x)^2;$$

$$|u - \hat{u}| \le \frac{T}{72}\left(\frac{1}{3}M_3 + \frac{1}{5}M_5\right)(\Delta x)^4,$$

де u, $\hat{u}$ — точний і наближений розв'язок;

$$M_2 = \max\left|\frac{\partial^2 u}{\partial t^2}\right|; \quad M_3 = \max\left|\frac{\partial^3 u}{\partial t^3}\right|;$$

$$M_4 = \max\left|\frac{\partial^4 u}{\partial t^4}\right|; \quad M_5 = \max\left|\frac{\partial^6 u}{\partial t^6}\right|.$$

Умова (8.37) стосується також і питання про збіжність до нуля похибки дискретизації при наближенні до нуля Δt і Δx. Можна показати, що якщо Δt і Δx наближатимуться до нуля так, що буде виконано (8.37), то відповідний наближений розв'язок збігатиметься до точного. Це — окремий випадок більш загального принципу, відомого як теорема еквівалентності Лакса, який для широкого класу диференціальних рівнянь і узгоджених різницевих схем стверджує, що глобальна похибка дискретизації буде наближатися до нуля в тому і лише в тому випадку, якщо використаний різницевий метод стійкий.

Таблиця 28

Δx	Δt
0,1	$0{,}5 \cdot 10^{-2}$
0,01	$0{,}5 \cdot 10^{-4}$
0,001	$0{,}5 \cdot 10^{-5}$

При зменшенні кроку за простором Δx умова (8.37) накладає все жорсткіші обмеження на значення кроку за часом (див. табл. 28, де $c = 1$). Це може призвести до того, що доведеться рухатися за часом із значно меншим кроком, аніж визначено залежністю від часу розв'язку самого диференціального рівняння. Хоча аналіз обмежувався найпростішим диференціальним рівнянням і найпростішою різницевою схемою, вимоги мализни кроку за часом є загальною проблемою, яка виникає при розв'язанні параболічних і подібних їм рівнянь за допомогою явних скінченнорізницевих методів.

Ця вимога є основним спонукальним моментом для використання так званих неявних методів.

Скінченнорізницевий метод (8.18) називається *явним*, бо значення u_j^{m+1} на наступному часовому шарі обчислюються за явними формулами через значення на попередньому шарі. Всупереч цьому розглянемо для рівняння теплопровідності

$$\frac{\partial u}{\partial t} = c\frac{\partial^2 u}{\partial x^2}, \quad 0 \le x \le 1, \quad t \ge 0 \tag{8.40}$$

різницеву схему

$$\frac{u_j^{m+1} - u_j^m}{\Delta t} = c\frac{u_{j+1}^{m+1} - 2u_j^{m+1} + u_{j-1}^{m+1}}{(\Delta x)^2}. \tag{8.41}$$

$$j = 1, \ldots, N-1.$$

За формулою ця схема подібна (8.15), але має істотну різницю, яка полягає в тому, що значення u_j у правій частині обчислюються на $(m+1)$-му часовому шарі, а не на m-му. Якщо при відомому значенні u_j^m $(j = 1, \ldots, N-1)$, потрібно обчислити u_j^{m+1} $(j = 1, \ldots, N-1)$, то, зрозуміло, що значення всіх змінних u_j у правій частині (8.41) нам невідомі. Отже, враховуючи шаблон для суто неявних схем (рис. 16,*б*), розглядатимемо (8.41) як систему рівнянь, що визначають значення u_j^{m+1} $(j = 1, \ldots, N-1)$. У цьому й полягає одна з основних різниць між явними та неявними методами: в явних методах ми маємо явні формули (як-то (8.18)), які дають змогу виразити u_j^{m+1} через значення u_j у попередні моменти часу, а в неявних методах для переходу на наступний часовий шар нам доводиться розв'язувати систему рівнянь. Проаналізуємо систему детальніше. Якщо встановити $\mu = c\Delta t/(\Delta x)^2$, то (8.41) можна переписати як

$$(1+2\mu)u_j^{m+1} - \mu(u_{j+1}^{m+1} + u_{j-1}^{m+1}) = u_j^m, \quad j = 1, 2, \ldots, N-1, \tag{8.42}$$

або в матрично-векторному вигляді

$$\begin{bmatrix} 1+2\mu & -\mu & & & \\ -\mu & 1+2\mu & -\mu & & \\ & \cdot & \cdot & \cdot & \\ & & -\mu & 1+2\mu & \end{bmatrix} \begin{bmatrix} u_1^{m+1} \\ u_2^{m+1} \\ \cdots \\ u_{N-1}^{m+1} \end{bmatrix} = \begin{bmatrix} u_1^m + \mu\alpha \\ u_2^m \\ \cdots \\ u_{N-2}^m \\ u_{N-1}^m + \mu\beta \end{bmatrix}. \tag{8.43}$$

Для виведення (8.43) ми використовували крайові (8.12) та початкові (8.11) умови; тому, як і раніше, $u_0^m = \alpha$, $u_N^m = \beta$ для $m = 0, 1, \ldots$; ці значення увійшли в праві частини першого та останнього рівнянь (8.43). Початкові умови мають вигляд $u_j^0 = g(x_j)$, $j = 1, \ldots, N-1$. Неявний метод (8.42) складається тепер із визначення u_j^{m+1} з u_j^m за допомогою розв'язання на кожному кроці за часом системи лінійних рівнянь (8.43). Матриця систе-

ми (8.43) тридіагональна і, через те, що $c > 0$, а тому й $\mu > 0$, є діагонально-домінуючою. Як видно з § 2.4, таку систему рівнянь можна ефективно розв'язувати методом прогонки.

Хоча система (8.43) може бути розв'язана досить ефективно, цей метод вимагає більших витрат на один крок за часом, ніж явний метод (8.18). Проте, як компенсація за ці додаткові витрати, ми отримуємо істотний виграш у стійкості методу, що в багатьох випадках дозволяє використовувати значно більший крок за часом, ніж в явному методі, і таким чином значно скоротити необхідний машинний час.

Накреслимо схему аналізу стійкості неявного методу за зразком явної схеми.

Припустимо, як і раніш, що $\alpha = \beta = 0$. Тоді розвиненню (8.31) відповідає зображення

$$u_j^m = \sum_{k=1}^{N-1} a_k \gamma_k^m \sin(jk\pi\Delta x). \tag{8.44}$$

Якщо

$$\gamma_k = \left[1 + 2\mu\left(1 + \cos\frac{k\pi}{N}\right)\right]^{-1}, \quad k = 1, \ldots, N-1, \tag{8.45}$$

то (8.44) тотожно задовольняє різницеві рівняння (8.42) для будь-яких значень a_k. Якщо a_k обчислюється за формулами

$$a_k = \sum_{l=1}^{N-1} g(x_l) \sin(k\pi j\Delta x), \tag{8.46}$$

тоді значення u_j^0 задовольнятимуть початкові умови.

Як і раніше, можна зробити висновок, що наближений розв'язок відбиває поведінку розв'язання диференціального рівняння лише в тому випадку, коли $u_j^m \to 0$ при $m \to \infty$. Із (8.44) видно, що в загальному випадку це матиме місце тоді й лише тоді, коли

$$|\gamma_k| < 1, \quad k = 1, \ldots, N-1. \tag{8.47}$$

Але оскільки $\mu > 0$, із (8.45) дістанемо

$$0 < \gamma_k < 1, \quad k = 1, \ldots, N-1; \tag{8.48}$$

тому (8.47) дійсно виконується. Найістотнішим є те, що (8.48) виконується для будь-якого $\mu > 0$. Оскільки $\mu = c\Delta t/(\Delta x)^2$, звідси випливає, що (8.48) виконується для кожного співвідношення Δt і Δx. У такому випадку кажуть, що метод є безумовно стійким, розуміючи під цим стійкість методу при будь-якому відношенні Δt і Δx.

Той факт, що метод (8.42) безумовно стійкий, не означає, що ми отримуватимемо непоганий наближений розв'язок при будь-якому виборі

Δt і Δx. Як і завжди, вони повинні бути достатньо малими, щоб забезпечити мализну похибки дискретизації. Можна показати, що метод (8.42), як і відповідний явний метод (8.18), мають перший порядок мализни за часом та другий порядок мализни за просторовою змінною, тобто похибку дискретизації можна відобразити у вигляді

$$e(\Delta t, \Delta x) = 0[\Delta t + (\Delta x)^2].$$

Припустимо, що

$$e(\Delta t, \Delta x) \approx c_1 \Delta t + c_2 (\Delta x)^2.$$

Тоді для сумірності внеску в повну похибку від дискретизації за часом та дискретизації за простором ми повинні поставити вимогу, щоб $\Delta t \approx$ $\approx c_3 (\Delta x)^2$.

Таким чином, хоча умова стійкості неявного методу не накладає жодних обмежень на співвідношення кроків Δt та Δx, вимоги до стійкості можуть привести до виникнення подібних обмежень. Оцінка похибки наближеного розв'язку (8.42) при розв'язанні задачі (8.10)—(8.12) має вигляд

$$|u - \hat{u}| \le T\left(\frac{\Delta t}{2} + \frac{(\Delta x)^2}{12}\right),$$

де u і $\hat{u}$ — точний та наближений розв'язки. При чисельному розв'язанні практичних задач, які описуються параболічними рівняннями в частинних похідних, звичайно використовуються неявні методи. Справа в тому, що виграш, пов'язаний із гарною стійкістю неявних різницевих схем, значно перевищує ту додаткову роботу, що необхідна при реалізації кроку за часом. Більшість фактичних методів мають складнішу структуру, ніж схема (8.42), та принципи залишаються тими ж. Часто використовують метод Кранка—Ніколсона, що є усередненням явного (8.18) та неявного (8.42) методів. Дійсно якщо взяти середнє правих частин (8.18) і (8.42), ми дістанемо (використовуючи шаблон *б* (див. рис. 16))

$$u_j^{m+1} - u_j^m = \frac{c\Delta t}{2(\Delta x)^2}\left(u_{j+1}^{m+1} - 2u_j^{m+1} + u_{j-1}^{m+1} + u_{j+1}^m - 2u_j^m + u_{j-1}^m\right). \quad (8.49)$$

Якщо встановити $\mu = c\Delta t/(2(\Delta x)^2)$, при запису (8.49) у матрично-векторній формі, ліва частина буде ідентична (8.43), а права матиме вигляд

$$\mu u_2^m - (2\mu - 1)u_1^m + 2\mu\alpha;$$

$$\mu u_{j+1}^m - (2\mu - 1)u_j^m + \mu u_{j-1}^m, \quad j = 2, \ldots, N-2;$$

$$2\mu\beta - (2\mu - 1)u_{N-1}^m + \mu u_{N-2}^m. \quad (8.50)$$

Отже, метод (8.49) зводиться до розв'язання на кожному кроці за часом тридіагональної системи рівнянь. Матриця коефіцієнтів цієї системи збігається з матрицею системи (8.43), а вектор правих частин, що

задається формулами (8.50), визначається трохи складніше, ніж у (8.43). Таким чином, метод (8.49) вимагає на кожному кроці за часом трохи більшого обсягу обчислень. Перевага методу (8.49) полягає, проте, в тому, що він є не лише безумовно стійким, як і (8.42), але й має другий порядок точності як за просторовою змінною, так і за часовою.

Приклад. Використовуючи метод сіток, розв'язати рівняння теплопровідності $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$ для заданих початкових $u(0, x) = \cos 2x$, $0 \le x \le 0{,}6$ і граничних умов $u(t, 0{,}6) = 0{,}3624$.

Розрахунки виконуються за розрахунковою схемою (8.39), вважаючи, що $x_j = 0{,}6 - j\Delta x$, $j = 0, 1, \ldots, 8$, $\Delta x = 0{,}6/8$, $t = m\Delta t$, $m = 0, \ldots, 64$; $\Delta t = 0{,}1/64$. Результати розрахунків наведені в табл. 29 для деяких значень m.

Таблиця 29.

x/t	0	0,075	0,150	0,225	0,300	0,375	0,450	0,525	0,600
64	0,400	0,469	0,509	0,524	0,517	0,493	0,457	0,412	0,362
63	0,409	0,478	0,516	0,529	0,521	0,497	0,459	0,413	0,362
62	0,419	0,486	0,523	0,535	0,526	0,500	0,461	0,414	0,362
61	0,428	0,494	0,530	0,541	0,531	0,504	0,464	0,415	0,362
60	0,438	0,502	0,537	0,547	0,536	0,507	0,466	0,416	0,362
59	0,447	0,510	0,544	0,553	0,540	0,511	0,468	0,417	0,362
58	0,456	0,519	0,551	0,559	0,545	0,514	0,471	0,419	0,362
57	0,466	0,527	0,558	0,565	0,550	0,518	0,473	0,420	0,362
56	0,475	0,535	0,566	0,571	0,554	0,521	0,475	0,421	0,362
55	0,484	0,543	0,573	0,576	0,559	0,525	0,478	0,422	0,362
...									
10	0,906	0,913	0,891	0,844	0,775	0,689	0,588	0,478	0,362
9	0,916	0,921	0,898	0,849	0,780	0,693	0,591	0,479	0,362
8	0,925	0,929	0,905	0,855	0,785	0,697	0,594	0,481	0,362
7	0,934	0,937	0,911	0,861	0,790	0,701	0,597	0,483	0,362
6	0,944	0,945	0,918	0,867	0,795	0,705	0,600	0,484	0,362
5	0,953	0,953	0,925	0,872	0,800	0,709	0,603	0,486	0,362
4	0,963	0,961	0,931	0,878	0,805	0,714	0,607	0,488	0,362
3	0,972	0,968	0,937	0,884	0,810	0,718	0,610	0,490	0,362
2	0,981	0,976	0,943	0,889	0,815	0,723	0,614	0,492	0,362
1	0,991	0,983	0,949	0,895	0,820	0,727	0,618	0,494	0,362
0	1,000	0,989	0,955	0,900	0,825	0,732	0,622	0,498	0,362

§ 8.3. МЕТОД ПРОГОНКИ ДЛЯ РІВНЯННЯ ТЕПЛОПРОВІДНОСТІ

Метод прогонки розглянемо на прикладі розв'язання крайових задач для рівняння теплопровідності.

Нехай потрібно у смузі $0 \le x \le a$, $0 \le t \le T$ розв'язати рівняння

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \tag{8.51}$$

що задовольняє умови

$$u(0, x) = f(x), \quad u(t, 0) = \varphi(t), \quad u(t, a) = \psi(t). \tag{8.51'}$$

Обираємо кроки $h = \Delta x$, $\tau = \Delta t$ за аргументами x і t відповідно; у кожному внутрішньому вузлі замінюємо похідні скінченнорізницевими виразами

$$\frac{\partial u}{\partial t} \approx \frac{u_j^m - u_j^{m-1}}{\tau}, \quad \frac{\partial^2 u}{\partial x^2} \approx \frac{u_{j+1}^m - 2u_j^m + u_{j-1}^m}{h^2},$$

обчислюємо значення $f(x)$, $\varphi(t)$, $\psi(t)$ у граничних вузлах. Позначивши $\mu = h^2/\tau$, отримуємо систему

$$u_{j-1}^{m+1} - (2 + \mu)u_j^{m+1} + u_{j+1}^{m+1} + \mu u_j^m = 0,$$

$$j = 1, 2, \ldots, N-1; \quad m = 0, 1, 2, \ldots; \tag{8.52}$$

$$u_j^0 = f(x_j); \tag{8.53}$$

$$u_0^m = \varphi(t_m); \tag{8.54}$$

$$u_N^m = \psi(t_m). \tag{8.55}$$

Метод прогонки розв'язання системи (8.52)—(8.55) полягає в тому, що рівняння (8.52) зводиться до вигляду

$$u_j^{m+1} = a_j^{m+1}(b_j^{m+1} + u_{j+1}^{m+1}). \tag{8.56}$$

На прямому ході, задовольняючи граничні умови (8.54), обчислимо величини a_j^{m+1}, b_j^{m+1} за формулами:

$$a_1^{m+1} = \frac{1}{2 + \mu}, \quad b_1^{m+1} = \varphi(t_{m+1}) + \mu u_1^m; \tag{8.57}$$

$$a_j^{m+1} = \frac{1}{2 + \mu - a_{j-1}^{m+1}}, \quad b_j^{m+1} = a_{j-1}^{m+1} b_{j-1}^{m+1} + \mu u_j^m, \quad j = 2, 3, \ldots, N-1. \tag{8.58}$$

На першому кроці зворотного ходу з правих граничних умов (8.55) знаходимо

$$u_N^{m+1} = \psi(t_{m+1}). \tag{8.59}$$

і послідовно, виконуючи зворотний хід, визначаємо значення u_j^{m+1} ($j = N - 1, \ldots, 1$) за формулами (8.56)

$$u_{N-1}^{m+1} = (u_N^{m+1} + b_{N-1}^{m+1})a_{N-1}^{m+1};$$

$$u_{N-2}^{m+1} = (u_{N-1}^{m+1} + b_{N-2}^{m+1})a_{N-2}^{m+1};$$

. . . . . .

$$u_1^{m+1} = (u_2^{m+1} + b_1^{m+1})a_1^{m+1}. \qquad (8.60)$$

Таким чином, метод прогонки дозволяє визначити значення функції $u(t, x)$ на шарі $t = t_{m+1}$, якщо відомі її значення на шарі $t = t_m$.

Приклад. Методом прогонки розв'язати рівняння теплопровідності, наведене у попередньому параграфі.

Користуючись розрахунковими формулами (вважаючи, що $\Delta x = 0{,}6$; $\Delta t = 0{,}01$), дістанемо розподіл температури $u(t, x)$ у деяких точках за x і за t, наведений у табл. 30.

Таблиця 30

x/t	0,00	0,12	0,24	0,36	0,48	0,60
0,10	0,400	0,497	0,524	0,499	0,436	0,362
0,09	0,460	0,545	0,560	0,523	0,452	0,362
0,08	0,520	0,593	0,596	0,548	0,464	0,362
0,07	0,580	0,641	0,633	0,572	0,476	0,362
0,06	0,640	0,689	0,669	0,597	0,488	0,362
0,05	0,700	0,738	0,706	0,621	0,501	0,362
0,04	0,760	0,786	0,742	0,646	0,514	0,362
0,03	0,820	0,834	0,779	0,671	0,527	0,362
0,02	0,880	0,881	0,815	0,697	0,540	0,362
0,01	0,940	0,928	0,851	0,724	0,556	0,362
0,00	1,000	0,971	0,887	0,752	0,574	0,362

§ 8.4. РІВНЯННЯ ГІПЕРБОЛІЧНОГО ТИПУ. МЕТОД СІТОК

Звернемося до хвильового рівняння

$$\frac{\partial^2 u}{\partial t^2} = c\,\frac{\partial^2 u}{\partial x^2}, \qquad (8.61)$$

що відноситься до класу гіперболічних. Це рівняння, а також його узагальнення, моделює широке коло питань, пов'язаних із поширенням хвиль, наприклад, задачі акустики. Класична задача, яка моделюється цим рівнянням, — задача про коливання струни.

Розглянемо пружну струну, натягнену вздовж осі x та закріплену в точках $x = 0$, $x = L$ (рис. 17)). Якщо струну відтягнути і відпустити, то вона почне коливатися, як зображено на рисунку. Вважатимемо струну «ідеальною», тобто припустимо, що струна абсолютно гнучка і що натяг

T сталий за всією довжиною струни, не залежить від t і великий порівняно з вагою струни. Позначимо відхилення струни в точці x і в момент часу t через $u(t, x)$ і вважатимемо, що відхилення u малі порівняно з довжиною струни L. До того ж вважатимемо, що у будь-якій точці тангенс кута нахилу вигнутої струни набагато менший за одиницю і що струна коливається в одній площині; нехтуватимемо горизонтальним зсувом точок струни порівняно з вертикальним.

Рис. 17

За другим законом Ньютона сила дорівнює добутку маси та прискорення. За допомогою досить елементарних міркувань можна показати, що відхилення струни задовольняє рівняння (8.61), де $c = T/m$, m — питома маса матеріалу (маса одиниці довжини). Для формулювання задачі необхідно задати початкові та граничні умови. Оскільки кінці струни закріплені, у точках $x = 0$ і $x = L$, граничні умови мають вигляд

$$u(t, 0) = 0, \quad u(t, L) = 0. \tag{8.62}$$

За початкову умову слід задати початкове відхилення і початкову швидкість точок струни. У даному випадку

$$u(0, x) = f(x), \quad u_t(0, x) = 0. \tag{8.63}$$

Таким чином, диференціальне рівняння (8.61), граничні (8.62) і початкові (8.63) умови повністю визначають математичну модель задачі.

Розглянемо побудову різницевої схеми для рівняння коливань (8.61) на відрізку $0 \le x \le 1$, $0 \le t \le T$ із початковими і граничними умовами:

$$u(0, x) = f(x), \quad u_t(0, x) = g(x); \tag{8.64}$$

$$u(t, 0) = \alpha(t), \quad u(t, 1) = \beta(t). \tag{8.65}$$

Вважаємо, що ця задача поставлена коректно, тобто її розв'язок існує, він єдиний і неперервно залежить від початкових і граничних умов. Використовуватимемо ту ж саму сітку ω_{th}, що й у § 8.2, тобто $\omega_{th} = \omega_\tau \times \omega_h$

$$\omega_\tau = \{t_m = m\Delta t = m\tau, \ m = 0, 1, \dots, k, \ k\tau = T\},$$

$$\omega_h = \{x_j = j\Delta x = jh, \ j = 0, 1, \dots, N, \ hN = 1\}.$$

Зрозуміло, що мінімальний шаблон, на якому можна апроксимувати рівняння (8.61), це п'ятиточковий шаблон, зображений на рис. 16,*г*. Таким чином, на відміну від схем рівняння теплопровідності, де використовувалися лише два часових шара (шари m і $m + 1$), тут потрібно використовувати три часових шара: $m - 1$, m, $m + 1$. Такі схеми називаються тришаровими. При їх застосуванні вважається, що при відшуканні значень u_j^{m+1} на верхньому шарі значення на попередніх шарах u_j^{m-1}, u_j^m, $j = 0, 1, \dots, N$ зберігаються у пам'яті ЕОМ. При цьому, у противагу до методу для рівнянь параболічного типу, доводиться вилучати додатковий обсяг

пам'яті для зберігання u_j^{m-1}; усе це випливає із того, що диференціальне рівняння (8.61) містить у собі другу похідну за часом.

Найпростішою різницевою апроксимацією рівняння (8.61) і граничних умов (8.65) є система рівнянь

$$\frac{u_j^{m+1} - 2u_j^m + u_j^{m-1}}{(\Delta t)^2} = \frac{c}{(\Delta x)^2}(u_{j+1}^m - 2u_j^m + u_{j-1}^m), \tag{8.66}$$

$$j = 1, 2, \ldots, N-1, \quad m = 1, 2, \ldots, k-1;$$

$$u_0^{m+1} = \alpha(t_{m+1}), \quad u_N^{m+1} = \beta(t_{m+1}), \quad m = 0, 1, \ldots, k-1. \tag{8.67}$$

Для переходу до обчислень необхідно задати значення u_j^0 і u_j^1, які можна отримати із початкових умов (8.64)

$$u_j^0 = f(x_j), \quad u_j^1 = f(x_j) + \Delta t g(x_j), \quad j = 1, \ldots, N-1. \tag{8.68}$$

Друга формула виникає як апроксимація початкової умови $u_t(0, x) = g(x)$ різницевого відношення $[u(\Delta t, x) - u(0, x)]/\Delta t = g(x)$. Із граничних умов маємо

$$u_0^m = \alpha(t_m), \quad u_N^m = \beta(t_m), \quad m = 0, 1, \ldots. \tag{8.69}$$

Оцінка похибки апроксимації u_j^1 у цьому випадку виглядає як

$$|\widetilde{u_1^1} - u_j^1| \le \frac{\alpha h}{2} M_2, \quad M_2 = \max\left\{\left|\frac{\partial^2 u}{\partial t^2}\right|, \left|\frac{\partial^2 u}{\partial x^2}\right|\right\}, \tag{8.70}$$

де $\widetilde{u_j^1}$ — точне значення.

Таким чином, значення наближеного розв'язку на $(m+1)$ часовому шарі можна обчислити за формулою

$$u_j^{m+1} = 2u_j^m - u_j^{m+1} + \mu(u_{j+1}^m - 2u_j^m + u_{j-1}^m), \tag{8.71}$$

де $\mu = c(\Delta t)^2/(\Delta x)^2$.

Якщо $\mu = 1$, то формула має найпростіший вигляд:

$$u_j^{m+1} = u_{j+1}^m + u_{j-1}^m - u_j^{m-1}. \tag{8.72}$$

Оцінка похибки наближеного розв'язку (8.71) у смузі $0 \le x \le S$, $0 \le t \le T$:

$$|\tilde{u} - u| \le \frac{h^2}{12}[(M_4 h + 2M_3)T + T^2 M_4],$$

де $\tilde{u}$ — точний розв'язок; $M_k = \max\left\{\left|\frac{\partial^k u}{\partial t^k}\right|, \left|\frac{\partial^k u}{\partial x^k}\right|\right\}$, $k = 3, 4$.

Найпростіша заміна другого із початкових умов (8.64) другим рівнянням (8.68) має лише перший порядок апроксимації за τ. Оскільки рівняння (8.66) апроксимує вихідне рівняння (8.61) із другим порядком, бажано, щоб і різницева початкова умова також мала другий порядок апроксимації. Якщо функція $f(x)$ має скінченну другу похідну, то значення u_j^1 можна визначити за допомогою формули Тейлора

$$u_j^1 = u_j^0 + \tau \frac{\partial u_j^0}{\partial t} + \frac{\tau^2}{2} \frac{\partial^2 u_j^0}{\partial t^2}. \tag{8.73}$$

Використовуючи рівняння (8.61) для $c = 1$ і початкові умови (8.64), можна записати

$$u_j^0 = f_j, \quad \frac{\partial u_j^0}{\partial t} = g_j, \quad \frac{\partial^2 u_j^0}{\partial t^2} = \frac{\partial^2 u_j^0}{\partial x^2} f_j''. \tag{8.74}$$

Тоді за формулою (8.73) матимемо

$$u_j^1 = f_j + \tau g_j + \frac{\tau^2}{2} f_j''. \tag{8.75}$$

Отже, різницеве рівняння (8.75) апроксимує другу умову (8.64) із другим порядком за τ і за h.

Сукупність рівнянь (8.66), (8.67), першого із рівнянь (8.68) і (8.75) становить різницеву схему, яка апроксимує вихідну задачу (8.61)—(8.63).

Покажемо ще один спосіб побудови різницевої початкової умови із другим порядком апроксимації. Замінимо похідну $u_t(0, x)$ різницевим відношенням $(u_j^1 - u_j^{-1})/(2\Delta t)$, що має другий порядок. Тоді з початкових умов матимемо

$$u_j^0 = f_j, \quad \frac{u_j^1 - u_j^{-1}}{2\Delta t} = g_j. \tag{8.76}$$

Напишемо різницеве рівняння (8.72) для шару $m = 0$

$$u_j^1 = u_{j+1}^0 + u_{j-1}^0 - u_j^{-1}. \tag{8.77}$$

Вилучивши з рівняння (8.76), (8.77) значення u_j^{-1}, дістанемо

$$u_j^0 = f_j, \quad u_j^1 = \frac{1}{2}(f_{j+1} - f_{j-1}) + \Delta t g_j. \tag{8.78}$$

Оцінка похибки значень u_j^1 має вигляд

$$|\tilde{u}_j^1 - u_j^1| \le \frac{h^2}{12} M_4 + \frac{h^3}{6} M_3, \tag{8.79}$$

де

$$M_k = \max\left\{\left|\frac{\partial^k u}{\partial t^k}\right|, \left|\frac{\partial^k u}{\partial x^k}\right|\right\}, \quad k = 3, 4. \tag{8.80}$$

Легко показати, що локальна похибка дискретизації у схемі (8.66) дорівнює $0[(\Delta t)^2 + (\Delta x)^2]$, тобто різницева схема має другий порядок точності як за просторовою змінною, так і за часом.

Для аналізу стійкості можна знов скористатися методом відокремлення змінних, припускаючи, що значення α і β в граничних умовах (8.65) дорівнюють нулю. Нехай $u_j^m = v_m w_j$. Підставляючи це зображення у (8.66), отримуємо

$$v_{m+1} - 2v_m + v_{m-1} = -\lambda\mu v_m, \quad m = 1, 2, \ldots; \tag{8.81}$$

$$w_{j+1} - 2w_j + w_{j-1} = -\lambda w_j, \quad j = 1, \ldots, N-1. \tag{8.82}$$

Рівняння (8.82) збігається з (8.28), і їх розв'язок визначається виразами (8.29), (8.30). Рівняння (8.81) зовнішньо мають ту ж форму, що і (8.82), проте являють собою різницевий аналог задачі Коші, де значення v_0 і v_1 вважаються заданими. Розв'язок (8.82) можна записати у вигляді

$$v_m = \gamma_1\eta_+^m + \gamma_2\eta_-^m, \quad m = 0, 1, \ldots,$$

де $\eta_\pm$ — корені характеристичного рівняння $\eta^2 + (\lambda\mu - 2)\eta + 1 = 0$, тобто

$$\eta_\pm = \frac{1}{2}\left(2 - \lambda\mu \pm \sqrt{\lambda^2\mu^2 - 4\lambda\mu}\right),$$

а коефіцієнти γ_i можна визначити за заданими початковими умовами v_0 і v_1. Таким чином, розв'язок (8.66) може бути записаний у вигляді

$$u_j^m = \sum_{k=1}^{n} a_k(\gamma_{k,1}\eta_{k,+}^m + \gamma_{k,2}\eta_{k,-}^m)\sin(jk\pi\Delta x), \tag{8.83}$$

де індекс k вказує, що відповідні γ і η обчислюються для $\lambda = \lambda_k$. Щоб значення u_j^m залишалися обмеженими при довільних a_k і будь-яких початкових умовах, необхідно й достатньо, щоб $|\eta_{k,i}| \le 1$. Легко перевірити, що якщо

$$\lambda_k\mu - 4 \le 0, \tag{8.84}$$

то $|\eta_{k,i}| = 1$, а якщо $\lambda_k\mu - 4 > 0$, то $\eta_{k,-} < -1$. Отже, (8.84) є необхідною й достатньою умовою стійкості. Оскільки власні значення λ_k містяться в інтервалі $0 < \lambda_k < 4$, то для виконання (8.84) достатньо, щоб $\mu \le 1$, або

$$\Delta t \le \Delta x/\sqrt{c}. \tag{8.85}$$

Ця умова є, по суті, й необхідною. Справа в тому, що $\lambda_n \to 4$ при $n \to \infty$; так що будь-яка, слабша за (8.85), умова призведе при достатньо великих n до порушення умови (8.84). Умова стійкості (8.85) накладає на Δt значно менш жорсткі обмеження, ніж умови (8.37) у випадку рівняння тепло-

провідності. Дійсно, з (8.85) видно, що Δt потрібно зменшувати пропорційно Δx, а не квадрату Δx, як у випадку (8.37).

Приклад. Скласти розв'язок задачі для рівняння коливання струни $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$ із початковими умовами $u(0, x) = f(x)$, $u_t'(t, 0) = F(x)$ $(0 \le x \le 1)$ і крайовими умовами $u(t, 0) = \varphi(t)$, $u(t, 1) = \psi(t)$, $0 \le t \le 0{,}5$, де

$$f(x) = x^2 \cos \pi x; \quad F(x) = x^2(x + 1);$$

$$\varphi(t) = 0{,}5t; \quad \psi(t) = t - 1.$$

Для розв'язання використаємо співвідношення

$$u_j^{m+1} = u_{j+1}^m + u_{j-1}^m - u_j^{m-1}, \quad j, m = 1, 2, \ldots .$$

При цьому $u_j^0 = f_j$, а для визначення u_j^1 можна використати один із засобів, наприклад, $u_j^1 = \frac{1}{2}(f_{j+1} + f_{j-1}) + \Delta t F_j$.

При розрахунках вважаємо $x_j = jh$, $h = 0{,}1$, $j = 0, 1, \ldots, 10$, $t_m = m\Delta t$, $\Delta t = 0{,}1$, $m = 0, 1, \ldots, 5$. Результати розрахунку наведені в табл. 31.

Таблиця 31

t/x	0	0,1	0,2	0,3	0,4	0,5
0,0	0,000000	0,050000	0,100000	0,150000	0,200000	0,250000
0,1	0,009511	0,017284	0,076512	0,120286	0,146032	0,157271
0,2	0,032368	0,036023	0,037570	0,072544	0,077556	0,059746
0,3	0,052935	0,052654	0,032055	−0,005159	−0,013742	−0,042330
0,4	0,049540	0,048967	0,009925	−0,054231	−0,125046	−0,139754
0,5	0,000199	0,006810	−0,037319	−0,109961	−0,180243	−0,212499
0,6	−0,110919	−0,086087	−0,113076	−0,163330	−0,197415	−0,310246
0,7	−0,287573	−0,230805	−0,212099	−0,200530	−0,293334	−0,392670
0,8	−0,517291	−0,413584	−0,318259	−0,342102	−0,395785	−0,449746
0,9	−0,769996	−0,604745	−0,543588	−0,513514	−0,498514	−0,482271
1,0	−1,000000	−0,900000	−0,800000	−0,700000	−0,600000	−0,500000

§ 8.5. СКІНЧЕННОРІЗНИЦЕВИЙ МЕТОД РОЗВ'ЯЗАННЯ РІВНЯНЬ ЕЛІПТИЧНОГО ТИПУ

Рівняння з частинними похідними, які залежать від просторових координат, мають назву *стаціонарних рівнянь*. Такі рівняння можуть мати

природне походження і описувати процеси різної природи, які не змінюються у часі. Прикладами таких рівнянь можуть бути задачі про визначення прогину навантаженої балки, тиск газу в неоднорідному силовому полі, стаціонарний розподіл тепла в тілі та ін. Усі ці задачі мають загальну властивість: припускається, що зовнішній вплив не залежить від часу, а початкові умови задані достатньо давно, й тому фізична система встигла вийти на стаціонарний розв'язок $u(\overline{r})$, який не залежить від часу.

Прикладом повної математичної постановки задачі для рівняння еліптичного типу є задача з крайовими умовами першого роду, яка має назву *задачі Діріхле для рівняння Пуассона*

$$a(x, y)\frac{\partial^2 u}{\partial x^2} + b(x, y)\frac{\partial^2 u}{\partial y^2} + c(x, y)\frac{\partial u}{\partial x} + d(x, y)\frac{\partial u}{\partial y} + g(x, y) = f(x, y), \tag{8.86}$$

яке задано в однозв'язній області G з границею Г. Коефіцієнти в рівнянні, права частина та границя Г є достатньо гладкими: $a(x, y) > 0$, $b(x, y) > 0$, $g(x, y) \le 0$ в G. Постановка задачі: знайти функцію $u(x, y)$, яка задовольняє всередині деякої області G рівняння (8.86), а на границі Г умові

$$u(x, y)|_{\Gamma} = \varphi(x, y), \tag{8.87}$$

де $\varphi(x, y)$ — задана неперервна функція. Припускаємо, що $f(x, y)$, $\varphi(x, y)$ є такі, що розв'язок задачі (8.86), (8.87) існує, є єдиним і є достатньо гладкою функцією. При $f \equiv 0$ отримуємо задачу Діріхле для рівняння Лапласа, одною з важливих властивостей якої є виконання принципу максимуму: неперервний в G і відмінний від константи розв'язок $u(x, y)$ може досягати свого максимального за модулем значення тільки на границі Г. Звідси випливає, що є справедливою оцінка

$$\max_{x,y \in G} |u(x, y)| \le \max_{x,y \in \Gamma} \varphi(x, y)|, \tag{8.88}$$

яка означає стійкість задачі за граничними даними. Вкриємо область G площини (x, y) сіткою паралельних прямих (рис. 18)

$$x_i = x_0 + ih, \quad y_k = y_0 + kh,$$

$$i, k = 0, \pm 1, \pm 2, \ldots .$$

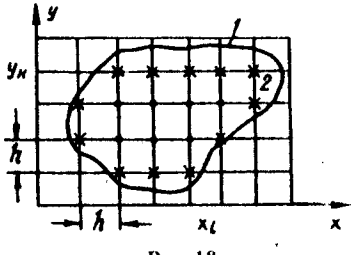


Рис. 18

Точки перетину цих прямих мають назву *вузлів*. Розглянемо тільки вузли, розташовані всередині області G. Два вузли (x_i, y_k) називають *сусідніми*, якщо відстань між ними за віссю x або за віссю y дорівнює h. Вузли, які мають чотирьох сусідів, розташованих всередині області G, мають назву *внутрішніх*. Множина всіх внутрішніх вузлів називається

сітчастою областю G_h (рис. 18,*1*). Вузли, у яких хоча б один сусідній вузол не належить G_h, називаються *граничними*. Множина всіх граничних вузлів називається *границею сітчастої області* Γ_h (рис. 18, *2*).

Як і в попередніх параграфах цієї глави, замінимо похідні внутрішніх вузлів в (8.86) різницевими відношеннями другого порядку точності апроксимації за формулами

$$\frac{\partial u}{\partial x}(x_i, y_k) = \frac{u(x_{i+1}, y_k) - (x_{i-1}, y_k)}{2h} + O(h)^2;$$

$$\frac{\partial u}{\partial y}(x_i, y_k) = \frac{u(x_i, y_{k+1}) - (x_i, y_{k-1})}{2h} + O(h)^2;$$

$$\frac{\partial^2 u}{\partial x^2}(x_i, y_k) = \frac{u(x_{i+1}, y_k) - 2u(x_i, y_k) + (x_{i-1}, y_k)}{h^2} + O(h)^2;$$

$$\frac{\partial^2 u}{\partial y^2}(x_i, y_k) = \frac{u(x_i, y_{k+1}) - 2u(x_i, y_k) + (x_i, y_{k-1})}{h^2} + O(h)^2.$$

Підставляючи ці співвідношення в (8.86), відкинувши похибку апроксимації похідних, отримаємо різницеві рівняння для невідомих $u_{i,k}$:

$$a_{i,k}\frac{u_{i+1,k} - 2u_{i,k} + u_{i-1,k}}{h^2} + b_{i,k}\frac{u_{i,k+1} - 2u_{i,k} + u_{i,k-1}}{h^2} +$$

$$+ c_{i,k}\frac{u_{i+1,k} - u_{i-1,k}}{2h} + d_{i,k}\frac{u_{i,k+1} - u_{i,k-1}}{2h} + g_{i,k}u_{i,k} = f_{i,k}, \qquad (8.89)$$

де введені позначки значень коефіцієнтів і правої частини у вузлі (x_i, y_k): $a_{i,k}$, $b_{i,k}$, $c_{i,k}$, $d_{i,k}$, $g_{i,k}$, $f_{i,k}$, наприклад $f_{i,k} = f(x_i, y_k)$, $(x_i, y_k) \in G_h$.

Співвідношення (8.89) містять, крім невідомих $u_{i,k}$ у внутрішніх вузлах, ще й невідомі $u_{i,k}$ на границі сітчастої області. Для граничних вузлів запишемо співвідношення

$$u(x_i, y_k) = \frac{\theta u(x_{i+1}, y_k) + \varphi(x_i \pm \theta h, y_k)}{\theta + 1} + O(h)^2,$$

або

$$u(x_i, y_k) = \frac{\theta u(x_i, y_{k+1}) + \varphi(x_i, y_k \pm \theta h)}{\theta + 1} + O(h)^2,$$

в залежності від того, яка точка $(x_i \pm \theta h, y_k)$ або $(x_i, y_k \pm \theta h)$, $0 \le \theta \le 1$, перетин неперервної границі Γ з лініями сітки знаходиться ближче до граничного вузла (рис. 19). Ці співвідношення означають, що значення $u(x_i, y_k)$ при $(x_i, y_k) \in \Gamma_h$ отримуються лінійною інтерполяцією значень $u(x, y)$ у внутрішньому вузлі і в точці перетину Γ з сіткою. Відкинувши в

останніх співвідношеннях похибку апроксимації, дістанемо вираз для невідомих $u_{i,k}$ в граничних вузлах

$$u_{i,k} = \frac{1}{\theta + 1}(\theta u_{i\pm 1,k} + \varphi_{i\pm\theta,k}) \tag{8.90}$$

або

$$u_{i,k} = \frac{1}{\theta + 1}(\theta u_{i,k\pm 1} + \varphi_{i,k\pm\theta}),$$

де введено позначення $\varphi_{i\pm\theta,k} = \varphi(x_i \pm \theta h, y_k)$, $\varphi_{i,k\pm\theta} = (x_i, y_k \pm \theta h)$. Зауважимо, що в (8.90) дробова частина кроку h (величина θ) залежить від вузла $\theta = \theta_{i,k}$. Щоб не ускладнювати запис в (8.90), індекси i та k випущені.

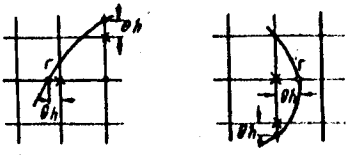


Рис. 19

Приєднуючи рівняння (8.90) до (8.89), отримаємо систему лінійних алгебраїчних рівнянь відносно $u_{i,k}$. У цій системі число рівнянь дорівнює числу невідомих і числу вузлів в G_h та Γ_h.

Система рівнянь (8.89), (8.90) — це різницева схема неперервної задачі (8.86), (8.87). Розв'язок цієї різницевої схеми — наближення до точного розв'язку у вузлах (x_i, y_k). Перепишемо систему рівнянь (8.89) у вигляді

$$L_h u_{i,k} = f_{i,k};$$

$$L_h u_{i,k} = A_{i,k}u_{i,k-1} + B_{i,k}u_{i-1,k} + C_{i,k}u_{i,k} + \\ + D_{i,k}u_{i+1,k} + E_{i,k}u_{i,k+1}, \tag{8.91}$$

де коефіцієнти задаються формулами

$$A_{i,k} = \frac{b_{i,k}}{h^2} - \frac{d_{i,k}}{2h}, \quad B_{i,k} = \frac{a_{i,k}}{h^2} - \frac{c_{i,k}}{2h}, \quad C_{i,k} = -\frac{2(a_{i,k} + b_{i,k})}{h^2} + g_{i,k},$$

$$D_{i,k} = \frac{a_{i,k}}{h^2} + \frac{c_{i,k}}{2h}, \quad E_{i,k} = \frac{b_{i,k}}{h^2} + \frac{d_{i,k}}{2h}.$$

За припущенням, гладкі функції $a(x, y) > 0$, $b(x, y) > 0$, $g(x, y) \le 0$, тому

$$g_{i,k} \le 0, \quad A_{i,k} > 0, \quad B_{i,k} > 0, \quad C_{i,k} < 0, \quad D_{i,k} > 0, \quad E_{i,k} > 0 \tag{8.92}$$

для достатньо малого кроку h. Зауважимо, що має місце рівність

$$A_{i,k} + B_{i,k} + C_{i,k} + D_{i,k} + E_{i,k} = g_{i,k}. \tag{8.93}$$

Для розв'язання системи лінійних рівнянь отриманої різницевої схеми можуть застосовуватись методи, викладені в главі 2 (метод простої

ітерації). Перепишемо систему рівнянь (8.91), (8.90) у вигляді, зручному для застосування методу простої ітерації:

для внутрішніх вузлів

$$u_{i,k} = -\frac{A_{i,k}}{C_{i,k}}u_{i,k-1} - \frac{B_{i,k}}{C_{i,k}}u_{i-1,k} - \frac{D_{i,k}}{C_{i,k}}u_{i+1,k} - \frac{E_{i,k}}{C_{i,k}}u_{i,k+1} + \frac{f_{i,k}}{C_{i,k}}; \quad (8.94)$$

для граничних вузлів

$$u_{i,k} = \frac{\theta}{\theta+1}u_{i\pm1,k\pm1} + \frac{1}{\theta+1}\varphi_{i\pm\theta,k\pm\theta}. \quad (8.95)$$

Тут для внутрішніх вузлів використовувався п'ятиточковий шаблон, зображений (див. рис. 16, *г*).

Припустимо, що $g_{i,k} < 0$ і виконуються умови (8.92). Розв'яжемо систему рівнянь відносно $u_{i,k}$ методом простої ітерації згідно з ітераційним процесом:

для внутрішніх вузлів

$$u_{i,k}^{(p+1)} = -\frac{A_{i,k}}{C_{i,k}}u_{i,k-1}^{(p)} - \frac{B_{i,k}}{C_{i,k}}u_{i-1,k}^{(p)} - \frac{D_{i,k}}{C_{i,k}}u_{i+1,k} - \frac{E_{i,k}}{C_{i,k}}u_{i,k+1}^{(p)} + \frac{f_{i,k}}{C_{i,k}};$$

для граничних вузлів

$$u_{i,k}^{(p+1)} = \frac{\theta}{\theta+1}u_{i\pm1,k\pm1}^{(p)} + \frac{1}{\theta+1}\varphi_{i\pm\theta,k\pm\theta},$$

$p = 0, 1, 2, \ldots$, $u_{i,k}^{(0)}$ задане.

Доведено, що якщо $g_{i,k} < 0$ та виконується умова (8.92), то послідовні наближення $u_{i,k}^{(p)}$ збігаються до точного розв'язку різницевої схеми $u_{i,k}$ або системи рівнянь (8.94), (8.95) і має місце оцінка

$$\max_{i,k} |u_{i,k}^{(p)} - u_{i,k}| \le \frac{q^p}{1-q}\max_{i,k}|u_{i,k}^{(0)}|,$$

де $q = \max\limits_{i,k}\left(\dfrac{\theta_{i,k}}{1+\theta_{i,k}}, -\dfrac{A_{i,k}+B_{i,k}+D_{i,k}+E_{i,k}}{C_{i,k}}\right)$.

Доведення цього твердження полягає в перевірці умови збіжності методу простої ітерації для системи лінійних рівнянь, при цьому мається на увазі, що невідомий вектор $\bar{x}$ утворює елементи $u_{i,k}$. Наприклад, компоненти вектора $\bar{x}$ можна перенумерувати таким чином: нехай $1 \le i \le N_1$, $1 \le k \le N_2$; тоді

$$x_1 = u_{1,1}, \quad x_2 = u_{2,1}, \quad \ldots, x_{N_1} = u_{N_1,1};$$

$$x_{N_1+1} = u_{2,1}\ x_{N_1+2} = u_{2,2}, \ldots, x_{2N_1} = u_{N_1,2};$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots x_{N_1N_2} = u_{N_1N_2}.$$

Відносно вектора $\bar{x} = \{x_1, x_2, \ldots, x_N\}$ різницева схема є системою лінійних рівнянь в матричному записі $A\bar{x} = \bar{b}$, де матриця A має в кожному рядку не більше п'яти ненульових елементів

$$A = \begin{pmatrix} * & * & . & . & . & . & . & . & . & . & . \\ * & * & * & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & * & . & . & * & * & * & . & * & . \\ . & . & * & . & . & * & * & * & . & * & . \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & * & * & \end{pmatrix}$$

Це пов'язано з тим, що похідні в кожному внутрішньому вузлі (i, k) апроксимувались за п'ятьма сусідніми вузлами.

Розв'язання різницевих рівнянь при $h \to 0$ збігається до точного розв'язання крайової задачі зі швидкістю, яка визначається порядком апроксимації рівнянь та крайових умов. Таким чином, для точного розв'язання $(u(x, y) \in C^4[G])$ оцінка похибки

$$\max_{i,k} |u_{i,k} - u(x_i, y_k)| = O(h^2), \quad h \to 0. \tag{8.96}$$

Оцінка похибки (8.96) є справедливою, якщо точний розв'язок неперервно диференційований чотири рази в області G. Для областей з кутовими точками, наприклад, прямокутника, взагалі кажучи, $u(x, y) \notin C^4[G]$. Але якщо гранична функція, тобто $\varphi(x, y)$ задовольняє в кутах спеціальні умови узгодження, то точний розв'язок $u(x, y) \in C^4[G]$ і є вірною оцінка (8.96).

Для прямокутної області $G = \{x_0 \le x \le x_1, \ y_0 \le y \le y_1\}$ такими умовами узгодження можуть бути:

1) достатня гладкість $\varphi(x, y)$;

2) функція $\varphi(x, y)$ повинна задовольняти в кутах прямокутника диференціальне рівняння.

Оцінка похибки (8.96) має в основному теоретичне значення, оскільки містить константу C, яку практично важко визначити

$$\max_{i,k} |u_{i,k} - u(x_i, y_k)| = ch^2 + O(h^2), \quad h \to 0.$$

Тому в реальних розрахунках використовується правило Рунге оцінки похибки, аналогічне тому, яке використовується в чисельному розв'язанні задачі Коші і розв'язанні звичайних диференціальних рівнянь. Робиться два варіанти розрахунку $u^h_{i,k}$ з кроком h та $u^{h/2}_{i,k}$ з кроком $h/2$; тоді похибка має вигляд

$$\max_{i,k} |u^{h/2}_{i,k} - u(x_i, y_k)| = \frac{1}{3} \max_{i,k} |u^{h/2}_{i,k} - u^h_{i,k}| + O(h^2)$$

і головна частина похибки визначається на вузлах, що збігаються.

Потрібно зазначити, що рівномірними прямокутними сітками найбільш зручно користуватися при розв'язанні задач у прямокутних областях. Якщо область має форму паралелограма (скошена система), то користуються координатами, осі яких паралельні сторонам цього паралелограма. Декартові прямокутні координати пов'язані з косокутними координатами (ξ, η) співвідношеннями $x = \xi + \eta \cos \alpha$; $y = \eta \sin \alpha$, де α — кут між ξ та η. У диференціальних виразах похідні за x та y замінюються похідними за ξ та η. Усі похідні апроксимуються за допомогою центральних різниць. Якщо область має форму кола, зручно користуватись полярними координатами $x = \rho \cos \theta$, $y = \rho \sin \theta$.

Наведемо деякі загальні зауваження. При чисельному розв'язанні крайових задач для диференціальних рівнянь в частинних похідних методом сіток можуть бути використані тільки різницеві схеми, які збігаються, оскільки в цьому разі можна розраховувати на отримання наближеного розв'язку задачі, достатньо близького до точного. Але й різницеві схеми, що збігаються, не завжди можуть бути використані при практичному розв'язанні задачі, оскільки при використанні методу сіток при обчисленні значень граничних функцій та правої частини виникають похибки. Щоб ці похибки не спотворили істинного розв'язання різницевої схеми, остання повинна бути стійкою за граничними умовами і за правою частиною. При використанні нестійкої різницевої схеми спотворення істинного розв'язку тим сильніше, чим дрібніша сітка; при використанні ж великої сітки не можна розраховувати на те, що розв'язок різницевої схеми буде близький до точного розв'язку крайової задачі для диференціального рівняння в силу поганої різницевої апроксимації рівняння.

Крім того, під час розв'язання різницевої задачі в процесі розрахунків нам обов'язково доведеться округляти значення розв'язків у вузлах сітки. Ці помилки можуть значно спотворити картину розв'язання, тому необхідною вимогою є стійкість різницевої схеми щодо помилок, які виникають в результаті округлення значень розв'язку у вузлах сітки. Оскільки помилки округлення значень розв'язку в вузлах сітки, принаймні, в найпростіших випадках можна компенсувати зміною правої частини різницевого рівняння, то особливо суттєвою є вимога до стійкості правої частини. Необхідно взяти до уваги й числовий алгоритм, який використовується для розв'язання різницевої схеми. Навіть у випадку, коли різницева схема стійка за граничними умовами і за правою частиною, при невдалому виборі алгоритму для розрахунку розв'язання цієї різницевої схеми може відбутися сильне накопичення обчислювальної похибки, у цьому разі нестійким буде сам процес розрахунку. Нестійкі алгоритми розрахунку практично непридатні у випадку дрібної сітки.

Приклад. Розглянемо рівняння

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + e^{-x} \frac{\partial u}{\partial x} - u = -f(x, y) \qquad (8.97)$$

у квадраті $G = \{0 \le x \le 1,\ 0 \le y \le 1\}$ з крайовими умовами

$$u(x, y) = \varphi(x, y) = \exp(x - y) \tag{8.98}$$

на сторонах квадрата.

Щоб функція $\varphi(x, y)$ задовольняла умови узгодження, підставимо (8.98) в (8.97) і, таким чином, визначимо функцію $f(x, y)$:

$$f(x, y) = e^{-y}(e^x + 1). \tag{8.99}$$

Отже, точний розв'язок рівняння (8.97) з крайовими умовами (8.98) — це й є функція (8.99), яка визначена всередині області G.

Метод побудови цього прикладу з відомим точним розв'язком крайової задачі є традиційним при відпрацюванні програми розв'язку розглядуваного класу задач.

Оберемо крок $h = 1/3$. Перенумеруємо вузли таким чином, як показано на рис. 20. Для значень $u_{i,k}$ у граничних вузлах із (8.98) дістаємо

$$u_{1,1} = 1{,}0,\ u_{1,2} = \exp(1/3),$$

$$u_{1,3} = \exp(2/3),\ u_{1,4} = \exp(1);$$

$$u_{2,4} = \exp(2/3),\ u_{3,4} = \exp(1/3),$$

$$u_{4,4} = 1{,}0,\ u_{4,3} = \exp(-1/3);$$

$$u_{4,2} = \exp(-2/3),\ u_{4,1} = \exp(-1),$$

$$u_{3,1} = \exp(-2/3),\ u_{2,1} = \exp(-1/3).$$

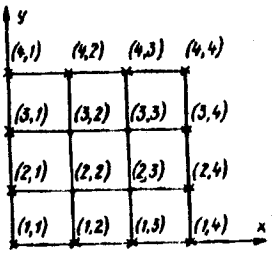


Рис. 20

Для значень у внутрішніх вузлах із (8.97), (8.99) отримуємо систему чотирьох рівнянь з чотирма невідомими:

$$\frac{u_{2,3} - 2u_{2,2} + u_{2,1}}{(1/3)^2} + \frac{u_{3,2} - 2u_{2,2} + u_{1,2}}{(1/3)^2} +$$

$$+ e^{-1/3}\frac{u_{2,3} - u_{2,1}}{(2/3)} - u_{2,2} = e^{-1/3}(e^{1/3} + 1);$$

$$\frac{u_{2,4} - 2u_{2,3} + u_{2,2}}{(1/3)^2} + \frac{u_{3,3} - 2u_{2,3} + u_{1,3}}{(1/3)^2} +$$

$$+ e^{-2/3}\frac{u_{2,4} - u_{2,2}}{(2/3)} - u_{2,3} = e^{-1/3}(e^{2/3} + 1);$$

$$\frac{u_{3,4} - 2u_{3,3} + u_{3,2}}{(1/3)^2} + \frac{u_{4,3} - 2u_{3,3} + u_{2,3}}{(1/3)^2} +$$

$$+ e^{-2/3}\frac{u_{3,4} - u_{3,2}}{(2/3)} - u_{3,3} = e^{-2/3}(e^{2/3} + 1);$$

$$\frac{u_{3,3} - 2u_{3,2} + u_{3,1}}{(1/3)^2} + \frac{u_{4,2} - 2u_{3,2} + u_{2,2}}{(1/3)^2} +$$

$$+ e^{-1/3}\,\frac{u_{3,3} - u_{3,1}}{(2/3)} - u_{3,2} = e^{-2/3}(e^{1/3} + 1).$$

Розв'язавши цю систему рівнянь відносно $u_{2,2}$, $u_{2,3}$, $u_{3,2}$, $u_{3,3}$, відшукаємо наближені значення до відповідних точних значень розв'язку $u(1/3, 1/3)$, $u(2/3, 1/3)$, $u(1/3, 2/3)$, $u(2/3, 2/3)$.

§ 8.6. ПОНЯТТЯ ПРО МЕТОД ПРЯМИХ РОЗВ'ЯЗАННЯ ГРАНИЧНИХ ЗАДАЧ

Розглянемо близький до різницевих чисельний метод прямих, в якому сітка вводиться тільки для частини змінних. Ці змінні розглядаються як дискретні, а одна змінна залишається неперервною. Похідні за дискретними змінними заміняються різницями. При цьому рівняння в частинних похідних апроксимується диференційно-різницевими рівняннями, які є системою великої кількості звичайних диференціальних рівнянь.

Нехай у прямокутній області $G\{\alpha \le x \le \beta,\ y_0 \le y \le y_0 + l\}$ необхідно знайти розв'язок еліптичного диференціального рівняння (8.86), що вдовольняє граничні умови:

$$u(x, y_0) = \varphi_0(x),\ u(x, y_0 + l) = \varphi_1(x) \quad (\alpha \le x \le \beta);$$

$$u(\alpha, y) = \psi_0(y),\ u(\beta, y) = \psi_1(y) \quad (y_0 \le y \le y_0 + l), \qquad (8.100)$$

де $\varphi_i(x)$, $\psi_i(y)$ $(i = 0, 1)$ — задані функції.

Розглянемо метод прямих наближеного розв'язання задачі, запропонований М.Г.Слободянським. На відрізку $[y_0, y_0 + l]$ візьмемо точки $y_k = y_0 + kh$ $(k = 0, 1, 2, \dots, n)$, $h = \dfrac{l}{n + l}$ і проведемо прямі $y = y_k$. Припускаючи існування достатньо гладкого розв'язання $u(x, y)$ задачі (8.86), (8.100), покладемо в рівнянні (8.86) $y = y_k$ $(k = 1, 2, \dots, n)$ і замінимо похідні за y різницевими відношеннями, наприклад, скористаємося рівностями

$$\left.\frac{\partial u}{\partial y}\right|_{y = y_k} = \frac{1}{2h}\,[u(x, y_{k+1}) - u(x, y_{k-1})] + O(h^2) =$$

$$= \frac{1}{2h}\,[u_{k+1}(x) - u_{k-1}(x)] + O(h^2);$$

$$\left.\frac{\partial^2 u}{\partial y^2}\right|_{y = y_k} = \frac{1}{h^2}\,[u(x, y_{k+1}) - 2u(x, y_k) + u(x, y_{k-1})] + O(h^2) =$$

$$= \frac{1}{h^2}\,[u_{k+1}(x) - 2u_k(x) + u_{k-1}(x)] + O(h^2), \qquad (8.101)$$

де $u_k(x) = u(x, y_k)$. Отримаємо систему n звичайних диференціальних рівнянь другого порядку:

$$a_k(x)u_k''(x) + \frac{b_k(x)}{h^2}\,[u_{k+1}(x) - 2u_k(x) + u_{k-1}(x)] + c_k(x)u_k'(x) +$$

$$+ \frac{d_k(x)}{2h}\,[u_{k+1}(x) - u_{k-1}(x)] + g_k(x)u_k(x) = f_k(x) + O(h^2).$$

Нехтуючи членами $O(h^2)$ та позначаючи через $u_k(x)$ наближене значення розв'язку $u(x, y)$ на прямій $y = y_k$, для визначення їх отримаємо систему рівнянь

$$a_k(x)u_k''(x) + \frac{b_k(x)}{h^2}\,[u_{k+1}(x) - 2u_k(x) + u_{k-1}(x)] + c_k(x)u_k'(x) +$$

$$+ \frac{d_k(x)}{2h}\,[u_{k+1}(x) - u_{k-1}(x)] + g_k(x)u_k(x) = f_k(x)$$

$$(k = 1, 2, \ldots, n). \qquad (8.102)$$

Використовуючи граничні умови на Г, маємо:

$$u_0(x) = \varphi_0(x) \quad (\alpha \le x \le \beta);$$

$$u_{n+1}(x) = \varphi_1(x) \quad (\alpha \le x \le \beta);$$

$$u_k(\alpha) = \psi_0(y_k),\ u_k(\beta) = \psi_1(y_k)\ \ (k = 1, 2, \ldots, n). \qquad (8.103)$$

Система (8.102) звичайних диференціальних рівнянь з граничними умовами (8.103) апроксимує з точністю до h^2 диференціальне рівняння (8.86) з граничними умовами (8.100) і має назву *системи рівнянь методу прямих.*

Загальний розв'язок системи (8.102) буде лінійно залежати від $2n$ довільних змінних. Використовуючи граничні умови (8.103) для відшукання цих змінних, отримаємо систему $2n$ лінійних алгебраїчних рівнянь, розв'язавши яку ми й знайдемо функції $u_k(x)$ $(k = 1, 2, \ldots, n)$, що апроксимують розв'язання задачі (8.86), (8.100) за прямими $y = y_k$ $(k = 1, 2, \ldots, n)$.

В залежності від способу заміни похідних за y різницевими відношеннями ми матимемо різні системи методу прямих, що з різною точністю апроксимують диференціальне рівняння (8.86).

Метод прямих можна розглядати як граничний випадок методу сіток, якщо, використовуючи прямокутну сітку, крок сітки за віссю x спрямувати до нуля.

При розв'язанні задач дискретизації змінних слід обирати в тому напрямі, де вихідні характеристики (геометричні, механічні параметри, фактори навантаження) змінюються більш гладко, а інтегрування ведеться за неперервною змінною.

§ 8.7. МЕТОД РІТЦА

Метод Рітца відноситься до варіаційних і базується на відповідності між знаходженням розв'язання крайової задачі для диференціального рівняння в частинних похідних і розв'язанням задачі про мінімум функціоналу, який відповідає розглядуваній крайовій задачі.

Зупинимось на основній ідеї методу Рітца на прикладі задачі про мінімум подвійного інтеграла

$$J(U) = \int\limits_D \int F(x, y, u, u_x', u_y', \dots)dx\,dy \qquad (8.104)$$

при умові

$$U = \varphi(s) \text{ на } \Gamma, \qquad (8.105)$$

де Γ — границя області D.

Нехай $U^*(x, y)$ — точний розв'язок задачі (8.104), (8.105), а $J(U^*) = m$ — мінімум. Якщо можемо побудувати функцію $\bar{U}(x, y)$, для якої значення інтеграла (8.104) близьке до значення m, то ця функція буде достатнім наближенням до шуканого розв'язку $u^*(x, y)$. Для знаходження такої функції $\bar{U}(x, y)$ методом Рітца запропоновано такий підхід. Вибирається функція, що залежить від кількох параметрів

$$U = \Phi(x, y, a_1, a_2, \dots, a_n), \qquad (8.106)$$

яка задовольняє умову (8.105). За рахунок вибору параметрів a_i $(i = 1, 2, \dots, n)$ можна побудувати таку функцію, яка дає інтегралу $J(U)$ найменше значення. При цьому вихідна задача (8.104), (8.105) зводиться до такої: підставивши функцію Φ з (8.106) у вираз (8.104) і виконавши необхідні операції, задачу про знаходження мінімуму функціоналу зводимо до задачі про знаходження мінімуму функції n змінних $I(a_1, a_2, \dots, a_n)$. Тоді для знаходження мінімуму цієї функції необхідно задовольнити умову

$$\frac{\partial I}{\partial a_i} = 0 \quad (i = 1, 2, \dots, n). \qquad (8.107)$$

Тобто для знаходження параметрів a_i $(i = 1, 2, \dots, n)$ маємо систему рівнянь (8.107). В залежності від того, чи функція Φ лінійна або нелінійна, маємо лінійну систему рівнянь (8.107). Розв'язавши цю систему і підста-

вивши знайдені значення параметрів a_i $(i = 1, 2, \ldots, n)$ у вираз (8.106), знаходимо наближене значення $\tilde{u}(x, y)$ шуканого розв'язку. Можна чекати, що зі збільшенням числа параметрів a_i $(i = 1, 2, \ldots, n)$ кожне наступне наближення буде наближатися до точного розв'язку.

Вираз (8.106) можна шукати у вигляді відрізка ряду

$$U(x, y) = \sum_{i=1}^{n} a_i u_i(x, y), \tag{8.108}$$

де $U_i(x, y)$ — задані координатні функції, що задовольняють граничні умови. Тоді, якщо функції u_i утворюють повну систему функцій при $n \to \infty$, то нескінченний ряд (8.108) був би точним розв'язком задачі.

Але при скінченному числі координатних функцій маємо наближений розв'язок.

Розглянемо, як приклад, задачу про згин прямокутної пластини сталої товщини зі сторонами a і b під дією навантаження q, всі сторони якої жорстко закріплені.

Така задача описується рівнянням

$$\Delta\Delta w = (q/D_m), \quad D_m = (Eh^3)/(12(1 - \nu^2)) \tag{8.109}$$

з граничними умовами: при

$$x = 0, a: \quad w = 0, \ \partial w/\partial x = 0;$$

при

$$y = 0, b: \quad w = 0, \ \partial w/\partial y = 0. \tag{8.110}$$

При цьому координатні осі віднесені до сторін пластини $x = 0$, $y = 0$.

Згідно з методом Рітца розв'язок задачі (8.109), (8.110) шукаємо у вигляді

$$w(x, y) = \sum_{m=1}^{M} \sum_{n=1}^{N} C_{mn} \left(1 - \cos \frac{2m\pi x}{a}\right) \left(1 - \cos \frac{2n\pi y}{b}\right). \tag{8.111}$$

Функціонал у даному випадку набуває вигляду

$$I(w) = \frac{1}{2} \int_D \int D_m \left\{ \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} \right)^2 + \right.$$

$$\left. + 2(1 - \nu) \left[\frac{\partial^2 w}{\partial x^2} \frac{\partial^2 w}{\partial y^2} - \left(\frac{\partial^2 w}{\partial x \partial y} \right)^2 \right] - qw \right\} dx\, dy. \tag{8.112}$$

Якщо обмежитися у виразі (8.111) одним членом, то після підстановки його у функціонал (8.112) і виконання всіх операцій дістанемо

$$I = C_n^2 D_m 2\pi \frac{3a^4 + 3b^4 + 2a^2b^2}{a^3b^3} - C_n abq. \qquad (8.113)$$

Звідки, задовольняючи умови стаціонарності (8.107), маємо

$$C_n = \frac{a^4b^4}{4\pi^4(3a^4 + 3b^4 + 2a^2b^2)} \frac{q}{D_m}. \qquad (8.114)$$

Остаточно одержуємо розв'язок задачі (8.109), (8.110) у вигляді

$$w(x, y) = \frac{a^4b^4}{4\pi^4(3a^4 + 3b^4 + 2a^2b^2)} \left(1 - \cos\frac{2\pi x}{a}\right)\left(1 - \cos\frac{2\pi y}{b}\right)\frac{q}{D_m}. \qquad (8.115)$$

Для квадратної пластини ($a = b$) максимальний прогин у центрі пластини $\left(x = y = \frac{a}{2}\right)$ дорівнює

$$w_{max} = 0{,}00128\,\frac{qa^4}{D_M}. \qquad (8.116)$$

Точне значення прогину, одержане в рядах, дорівнює $w_{max} = 0{,}00126\,\frac{qa^4}{D_m}$, тобто похибка становить 1,5%. Якщо у розв'язку (8.111) врахувати сім членів ряду, то одержаний результат збігається з точним.

§ 8.8. МЕТОД БУБНОВА—ГАЛЬОРКІНА

Поряд з методом Рітца для розв'язання крайових задач для диференціальних рівнянь у частинних похідних застосовується метод Бубнова—Гальоркіна, який має тісний зв'язок з варіаційними проблемами і, зокрема, з методом Рітца.

Викладемо основну ідею методу Бубнова—Гальоркіна. Нехай треба знайти розв'язок задачі для диференціального рівняння в частинних похідних

$$L\,[U(x, y)] = f(x, y) \qquad (8.117)$$

з відповідними граничними умовами. Шукатимемо наближений розв'язок крайової задачі у вигляді

$$\tilde{U}(x, y) = \sum_{i=1}^{n} c_i\varphi_i(x, y). \qquad (8.118)$$

де $\varphi_i(x, y)$ $(i = 1, 2, \ldots, n)$ — деяка система заданих функцій, що задовольняють граничні умови, а c_i — невизначені коефіцієнти. Вважаємо, що функції $\varphi_i(x, y)$ лінійно незалежні і є перші n функцій деякої повної системи функцій. Для того щоб $\bar{U}(x, y)$ була точним розв'язком крайової задачі для рівняння (8,118), треба, щоб вираз $L(\bar{U})$ дорівнював нулю точно, а ця вимога рівносильна вимозі ортогональності виразу $L(\bar{U})$ до всіх функцій системи $\{\varphi_i(x, y)\}$ $(i = 1, 2, \ldots, n, \ldots)$. Але ж маючи у своєму розпорядженні тільки n коефіцієнтів $c_1, c_2, \ldots, c_n$, можна лише задовольняти n умов ортогональності. Задовольняючи ці умови, приходимо до системи

$$\int_D \int \left\{L\left[\bar{U}(x, y)\right] - f(x, y)\right\} \varphi_i(x, y)dx\, dy =$$

$$= \int_D \int \left\{L\left[\sum_{i=1}^{n} c_i\varphi_i(x, y)\right] - f(x, y)\right\} \varphi_i(x, y)dx\, dy = 0 \quad (i = 1, 2, \ldots, n). \tag{8.119}$$

Розв'язуючи систему рівнянь (8.119), знаходимо коефіцієнти c_i $(i = 1, 2, \ldots, n)$. Підставляючи знайдені коефіцієнти c_i у вираз (8.118), отримаємо шуканий розв'язок крайової задачі.

Метод Бубнова—Гальоркіна взагалі не зв'язаний з варіаційними проблемами і може використовуватись для розв'язання тих задач, де не можна побудувати функціонал, що відповідає крайовій задачі для рівняння (8.117). Але для задач, які зв'язані з варіаційними проблемами, метод Бубнова—Гальоркіна знаходиться у тісному зв'язку і з методом Рітца (див. гл. 6).

Застосування методу Бубнова—Гальоркіна до розв'язання крайових задач розглянемо на прикладі розв'язання тієї ж самої задачі, що розв'язувалась методом Рітца, тобто задачі про згин прямокутної пластини при рівномірному навантаженні q з жорстко закріпленими сторонами (§ 8.7). Задача описується рівнянням

$$L[w(x, y)] = \Delta\Delta w = q/D_M \tag{8.120}$$

з граничними умовами:
при

$$x = 0, a: \quad w = 0, \partial w/\partial x = 0;$$

при

$$y = 0, b: \quad w = 0, \partial w/\partial y = 0. \tag{8.121}$$

Розв'язок задачі шукаємо, як і в § 8.7, у вигляді

$$w(x, y) = \sum_{m=1}^{M}\sum_{n=1}^{N} A_{mn}\left(1 - \cos\frac{2m\pi x}{a}\right)\left(1 - \cos\frac{2n\pi y}{b}\right). \tag{8.122}$$

При цьому задовольняються всі граничні умови (8.121). Тоді ліва частина виразу (8.119) після підстановки в неї розв'язку (8.122) з урахуванням одного члена набуває значення:

$$\int_0^a \int_0^b \left(\Delta\Delta w - \frac{q}{D_M}\right)\left(1 - \cos\frac{2\pi x}{a}\right)\left(1 - \cos\frac{2\pi y}{b}\right) dx\, dy =$$

$$= A_{11} 4\pi^4 ab\left(\frac{3}{a^4} + \frac{3}{b^4} + \frac{2}{a^2 b^2}\right) - ab\frac{q}{D_M}. \qquad (8.123)$$

Прирівнюючи одержану величину нулеві, знаходимо те ж саме значення коефіцієнта A_{11}, що й за методом Рітца, тобто

$$A_{11} = C_{11} = \frac{a^4 b^4}{4\pi^4(3a^4 + 3b^4 + 2a^2 b^2)} \frac{q}{D_M}. \qquad (8.124)$$

Таким чином, для даної задачі метод Бубнова—Гальоркіна дає розв'язок, який повністю збігається з розв'язком, отриманим за методом Рітца (§ 8.7).

§ 8.9. МЕТОД ВЛАСОВА—КАНТОРОВИЧА

Метод Власова—Канторовича — один з наближених методів, що дозволяє звести задачу, що описується диференціальним рівнянням в частинних похідних, до системи звичайних диференціальних рівнянь з відповідними граничними умовами,

Розглянемо основні положення методу Власова—Канторовича. Нехай в області $D = \{x_1 \le x \le x_2;\ y_1 \le y \le y_2\}$ система диференціальних рівнянь в частинних похідних має вигляд

$$\overline{F}\left(x, y, \overline{u}, \frac{\partial \overline{u}}{\partial x}, \frac{\partial \overline{u}}{\partial y}, \frac{\partial^2 \overline{u}}{\partial x^2}, \frac{\partial^2 \overline{u}}{\partial x \partial y}, \frac{\partial^2 \overline{u}}{\partial y^2}, \cdots\right) = \overline{f}(x, y) \qquad (8.125)$$

і на границі області Г задані граничні умови

$$\overline{G}\left(x, y, \overline{u}, \frac{\partial \overline{u}}{\partial x}, \frac{\partial \overline{u}}{\partial y}, \frac{\partial^2 \overline{u}}{\partial x^2}, \frac{\partial^2 \overline{u}}{\partial x \partial y}, \frac{\partial^2 \overline{u}}{\partial y^2}, \cdots\right)\Bigg|_{\Gamma} = 0, \qquad (8.126)$$

де вектор-функції $\overline{F}$ і $\overline{G}$ можуть бути лінійними або нелінійними відносно своїх аргументів.

Розв'язок крайової задачі (8.125), (8.126) для компонент вектора $\overline{u}(x, y)$ шукаємо у вигляді відрізка ряду

$$u_i(x, y) = \sum_{j=1}^{m} u_{ij}(x)\varphi_{ij}(y) \quad (i = 1, 2, \ldots, l), \qquad (8.127)$$

де $\varphi_{ij}(y)$ — апроксимуючі функції однієї змінної, що вибираються з певних міркувань, утворюють ортогональні системи лінійно незалежних функ-

цій в інтервалі $y_1 \le y \le y_2$ і задовольняють граничні умови (8.126); функції $u_{ij}(x)$ повинні бути визначені в інтервалі $x_1 \le x \le x_2$.

Підставляючи вирази (8.127) у вихідну систему рівнянь (8.125), знаходимо відхил

$$\bar{\varepsilon}\left(x, y, \bar{u}_{ij}, \frac{\partial \bar{u}_{ij}}{\partial x}, \frac{\partial^2 \bar{u}_{ij}}{\partial x^2}, \dots\right) =$$

$$= \bar{F}\left(x, y, \bar{u}, \frac{\partial \bar{u}}{\partial x}, \frac{\partial^2 \bar{u}}{\partial x^2} \frac{\partial^2 \bar{u}}{\partial x \partial y}, \frac{\partial^2 \bar{u}}{\partial y^2}, \dots\right) - \bar{f}, \qquad (8.128)$$

де замість $\bar{u}$ підставлені компоненти (8.127). Далі за допомогою процедури Бубнова—Гальоркіна проектуємо відхил $\bar{\varepsilon}$ на систему функцій $\varphi_{ij}(y)$, тобто задовольняємо умови

$$\int_{y_1}^{y_2} \varepsilon_i\left(x, y, u_{ij}, \frac{\partial u_{ij}}{\partial x}, \dots\right)\varphi_{ij}(y) = 0$$

$$(i = 1, 2, \dots, l;\ j = 1, 2, \dots, m), \qquad (8.129)$$

де ε_i — компоненти відхилу $\bar{\varepsilon}$.

Аналогічні перетворення виконуємо з граничними умовами на контурах $x = \text{const}$. У результаті приходимо до системи звичайних диференціальних рівнянь відносно невідомих функцій $u_{ij}(x)$:

$$\Phi_{ij}\left(x, u_{ij}, \frac{du_{ij}}{dx}, \frac{d^{m}u_{ij}}{dx^{2}}, \dots\right) = 0\ (i = 1, 2, \dots, l;\ j = 1, 2, \dots, m). \qquad (8.130)$$

Додаючи до системи (8.130) граничні умови на контурах $x = x_1$ і $x = x_2$, дістанемо одновимірну крайову задачу для функцій $u_{ij}(x)$. Розв'язавши цю систему і підставивши функції $u_{ij}(x, y)$ у вираз (8.127), знаходимо розв'язок вихідної крайової задачі (8,125), (8.126).

Для одержання системи рівнянь (8.130) також можна застосувати варіаційний метод Рітца, за допомогою якого задача про мінімум подвійного інтеграла зводиться до задачі про мінімум простого інтеграла. При цьому розв'язок задачі шукаємо у вигляді відрізка ряду (8.127).

Задачі (8.125), (8.126) відповідає інтеграл

$$J(u) = \int_D \int F(x, y, u, u_x', u_y')dx\,dy, \qquad (8.131)$$

мінімум якого дає розв'язок даної задачі.

Після підставлення виразу (8.127) у (8.131) і виконання всіх операцій інтегрування по координаті y, отримаємо вираз для функціоналу у вигляді

$$J(u) = \int_{x_1}^{x_2} \Phi(x, u_1, u_1', u_1'', u_2, u_2', u_2', \ldots)dx. \qquad (8.132)$$

З умови стаціонарності інтеграла (8.132) випливає, що функції u_m ($m = 1, 2, \ldots$) повинні задовольняти систему рівнянь Ейлера:

$$\Phi'_{u_1} - \frac{d}{dx}\Phi'_{u'_1} + \frac{d^2}{dx^2}\Phi'_{u''_1} - \ldots = 0;$$

$$\Phi'_{u_2} - \frac{d}{dx}\Phi'_{u'_2} + \frac{d^2}{dx^2}\Phi'_{u''_2} - \ldots = 0. \qquad (8.133)$$

До цієї системи звичайних диференціальних рівнянь треба додати граничні умови на контурах $x = \text{const}$, після чого приходимо до одновимірної крайової задачі відносно функції $u_m(x)$ ($m = 1, 2, \ldots$).

Проілюструємо застосування методу Власова—Канторовича на прикладі розв'язання задачі про згин прямокутної пластини сталої товщини зі сторонами $2a$ і $2b$ під дією рівномірного навантаження q при жорсткому закріпленні всіх сторін. Задача описується рівнянням

$$\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4} = \frac{q}{D_M} \qquad (8.134)$$

з граничними умовами: при

$$y = \pm b; \quad w = 0, \quad \frac{\partial w}{\partial y} = 0; \qquad (8.135)$$

при

$$x = \pm a; \quad w = 0, \quad \frac{\partial w}{\partial x} = 0. \qquad (8.136)$$

У відповідності з першим підходом за даним методом розв'язок (8.127), задовольняючи умови (8.135), шукаємо у вигляді

$$w(x, y) = \sum_{j=1}^{m} w_j(x)\varphi_j(y), \qquad (8.137)$$

де $\varphi_i(y)$ повинні задовольняти умови

$$\varphi_j(\pm b) = \varphi_j'(\pm b) = 0, \quad j = 1, 2, \ldots, m). \qquad (8.138)$$

Вважаючи, що задача симетрична відносно осі $0x$, функції $\varphi_j(y)$ вибираємо таким чином:

$$\varphi_j(y) = (y^2 - b^2)^2 y^{2j-2}. \qquad (8.139)$$

Використовуючи процедуру (8.129) і обмежуючись одним членом ряду (8.137), маємо

$$\int_{-b}^{b}\left(\Delta\Delta w-\frac{q}{D_M}\right)\varphi_1(y)dy=\int_{-b}^{b}\left[24w_1+2(12y^2-4b^2)w_1''+\right.$$

$$\left.+(y^2-b^2)^2w_1^{(4)}-\frac{q}{D_m}\right](y^2-b^2)dy=0,$$

або

$$\frac{128}{315}b^9w_1^{(4)}-\frac{256}{105}b^7w_1''+\frac{64}{5}b^5w_1=\frac{1}{2}q_1,$$

де

$$q_1=\int_{-b}^{b}\frac{q(x,y)}{D_M}(y^2-b^2)^2dy. \tag{8.140}$$

Загальний розв'язок рівняння (8.140) має вигляд

$$w_1(x)=C_1\operatorname{ch}\frac{\alpha x}{b}\cos\frac{\beta x}{b}+C_2\operatorname{ch}\frac{\alpha x}{b}\sin\frac{\beta x}{b}+$$

$$+C_3\operatorname{sh}\frac{\alpha x}{b}\sin\frac{\beta x}{b}+C_4\operatorname{sh}\frac{\alpha x}{b}\cos\frac{\beta x}{b}+w_0(x), \tag{8.141}$$

де $w_0(x)$ — частинний розв'язок рівняння (8.140); $\alpha=2{,}075$; $\beta=1{,}143$. Якщо $q=\text{const}$ і мають місце при $x=\text{const}$ умови (8.136), то $q_1=\frac{16}{15}qb^5$ і розв'язок (8.141) запишемо у вигляді

$$w(x,y)=(b^2-y^2)^2\frac{q}{24\gamma_0D_M}\left[\gamma_1\operatorname{ch}\frac{\alpha x}{b}\cos\frac{\beta x}{b}+\gamma_2\operatorname{sh}\frac{\alpha x}{b}\sin\frac{\beta x}{b}+\gamma_0\right], \tag{8.142}$$

де

$$\gamma_0=\beta\operatorname{sh}\alpha\mu\operatorname{ch}\alpha\mu+\alpha\sin\beta\mu\cos\beta\mu;$$

$$\gamma_1=-(\alpha\operatorname{ch}\alpha\mu\sin\beta\mu+\beta\operatorname{sh}\alpha\mu\cos\beta\mu);$$

$$\gamma_2=\alpha\operatorname{sh}\alpha\mu\cos\beta\mu-\beta\operatorname{ch}\alpha\mu\sin\beta\mu;$$

$$\mu=a/b.$$

Для квадратної пластини ($a=b$) маємо

$$w(0,0)=0{,}0136\,(2b)^4\frac{q}{Eh^3},\quad \nu=0{,}3, \tag{8.143}$$

а точний розв'язок $w(0,0)=0{,}0138(2b^4)\frac{q}{Eh^3}$.

На підставі другого підходу, вибираючи прогин теж у вигляді (8.137) і обмежуючись одним членом ряду, запишемо вираз для функціоналу (8.132)

$$I(\varphi) = \int_{-a}^{a} \left[\frac{128}{315} b^9 (w_1'')^2 - \frac{256}{105} b^7 w_1'' w_1 + \frac{64}{5} b^5 w_1^2 - \frac{16}{15} b^5 \frac{q}{D_M} \right] dx. \quad (8.144)$$

Звідси диференціальне рівняння Ейлера (8.133) набуває вигляду

$$\frac{256}{315} b^9 w_1^{(4)} - \frac{2256}{105} b^7 w_1'' + \frac{128}{5} b^5 w_1 = \frac{16}{15} b^5 \frac{q}{D_M}. \quad (8.145)$$

Порівняння одержаного рівняння (8.145) з рівнянням (8.140) показує, що вони ідентичні, тому можна зробити висновок, що обидва підходи методу Власова—Канторовича приводять до одного результату.

§ 8.10. МЕТОД, ЩО БАЗУЄТЬСЯ НА СПЛАЙН-АПРОКСИМАЦІЇ ФУНКЦІЙ В ОДНОМУ НАПРЯМІ

Основні відомості про поліноміальні сплайни наведено в § 1.8. Для застосування зазначеного методу використовуються базисні сплайни — B-сплайни. Множина всіх сплайнів $S(x)$ n-го степеня є лінійним скінченновимірним простором розмірності $N+n$. Найуважнішим базисом у цьому просторі є B-сплайни n-го степеня — $B_n^i(x)$ $(i = -n, \ldots, N-1)$, тобто будь-який сплайн $S(n)$ n-го степеня можна подати у вигляді

$$S(x) = \sum_{i=-n}^{N-1} b_i B_n(x), \quad (8.146)$$

де b_i — деякі сталі коефіцієнти; i — номер сплайна (вони нумеруються за лівим вузлом їхніх носіїв). B-сплайни нульового степеня на сітці Δ визначаються таким чином:

$$B_0^i(x) = \begin{cases} 1, & x \in [x_i, x_{i+1}], \\ 0, & x \notin [x_i, x_{i+1}], \end{cases}$$

$$i = \overline{0, N-1}. \quad (8.147)$$

Для побудови B-сплайнів степеня $n (n \geq 1)$ сітку Δ слід розширити, додаючи ще точки $x_{-n} < x_{-n+1} < \ldots < x_{-1} < x_0$; $x_N < x_{N+1} < \ldots < x_{N+n}$. Зокрема, можна покласти, що $x_i = x_0 + i(x_1 - x_0)$. Отримаємо сітку Δ'

$$x_{-n} < \ldots < x_{-1} < x_0 < \ldots x_N < x_{N+1} < \ldots < x_{N+n}.$$

Для B-сплайнів степеня n, що визначені на сітці Δ', має місце рекурентне співвідношення

$$B_n^i(x) = \frac{x - x_i}{x_{i+n} - x_i} B_{n-1}^i(x) + \frac{x_{i+n+1} - x}{x_{i+n+1} - x_{i+1}} B_{n-1}^{i+1}(x),$$

$$n = 1, 2, \ldots; \quad i = \overline{-n, N-1}. \tag{8.148}$$

З усіх B-сплайнів степеня n, які не дорівнюють нулеві на інтервалі $[x_i, x_{i+1}]$ $(i = \overline{0, N-1})$, можна скласти таблицю:

$$\begin{array}{r} B_0^i; \\ B_1^{i-1} B_1^i; \\ B_2^{i-2} B_2^{i-1} B_2^i; \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ B_n^{i-1} \ \ldots \ B_n^{i-2} B_n^{i-1} B_n^i. \end{array} \tag{8.149}$$

Сплайни $B_n^i(x)$ $(i = \overline{-n, N-1})$ мають такі властивості:

$$B_n^i(x) > 0, \quad x \in (x_i, x_{i+n+1});$$

$$B_n^i(x) \equiv 0, \quad x \notin (x_i, x_{i+n+1}); \tag{8.150}$$

$$\int_{-\infty}^{\infty} B_n^i(x)dx = 1.$$

Зокрема, при $h = x_{k+1} - x_k = \text{const}$ $(k = \overline{0, N-1})$, враховуючи (8.147)—(8.149), для B-сплайнів третього та п'ятого степенів дістанемо

$$B_3^i(x) = \frac{1}{6} \begin{cases} 0, & x < x_{j-2}; \\ t^3, & x_{j-2} \le x \le x_{j-1}; \\ 1 + 3t + 3t^2(1-t), & x_{j-1} \le x \le x_j; \\ 1 + 3(1-t) + 3t(1-t)^2, & x_j \le x \le x_{j+1}; \\ (1-t)^3, & x_{j+1} \le x \le x_{j+2}; \\ 0, & x > x_{j+2}, \end{cases} \tag{8.151}$$

де $t = (x - x_m)/h$ на інтервалі, $[x_m, x_{m+1})$, $m = \overline{j-2, j+1}$; $j = \overline{-1, N+1}$

$$B_5^j(x) = \frac{1}{120}\begin{cases} 0, & x < x_{j-3}; \\ t^5, & x_{j-3} \le x \le x_{j-2}; \\ -5t^5 + 5t^4 + 10t^3 + 10t^2 + 5t + 1, & x_{j-2} \le x \le x_{j-1}; \\ 10t^5 - 20t^4 - 20t^3 + 20t^2 + 50t + 26, & x_{j-1} \le x \le x_j; \\ -10t^5 + 30t^4 - 60t^2 + 66, & x_j \le x \le x_{j+1}; \\ 5t^5 - 20t^4 + 20t^3 + 20t^2 - 50t + 26, & x_{j+1} \le x \le x_{j+2}; \\ (1-t)^5, & x_{j+2} \le x \le x_{j+3}; \\ 0, & x > x > x_{j+3}, \end{cases} \tag{8.152}$$

де $t = (x - x_m)/h$ на інтервалі $[x_m, x_{m+1}]$, $m = \overline{j-3, j+2}$; $j = \overline{-2, N+2}$ (тут і надалі сплайни нумеруватимемо за середнім вузлом їхніх носіїв). У цьому випадку сплайни третього й п'ятого степенів, що визначаються формулами (8.151) і (8.152), запишемо у вигляді

$$S_3(x) = \sum_{j=-1}^{N+1} b_j B_3^j(x);$$

$$S_5(x) = \sum_{j=-2}^{N+2} b_j B_5^j(x).$$

При розв'язанні задач математичної фізики, зокрема теорії пластин і оболонок, у багатьох випадках виникає необхідність побудови сплайна, що задовольняє наперед різні граничні умови на кінцях інтервалу $[x_0, x_n]$. Такі сплайни у вигляді лінійної комбінації B-сплайнів можна побудувати для деяких варіантів граничних умов. Зокрема для сплайнів третього й п'ятого степенів їх можна подати у вигляді

$$S_3(x) = \sum_{j=0}^{N} \alpha_j \varphi_j(x), \quad N \ge 4; \tag{8.153}$$

$$S_5(x) = \sum_{j=0}^{N} \beta_j \psi_j(x), \quad N \ge 6, \tag{8.154}$$

де α_j і β_j — деякі сталі коефіцієнти; φ_j, ψ_j — лінійні комбінації B-сплайнів третього й п'ятого степенів відповідно.

Викладемо тепер основні положення методу зведення двовимірних крайових задач до одновимірних за допомогою сплайн-апроксимації функцій в одному координатному напрямі й чисельному інтегруванні в другому. Нехай у прямокутній області $D = \{x_1 \le x \le x_2;\ y_1 \le y \le y_2\}$ задана система диференціальних рівнянь у частинних похідних у вигляді

$$F\left(x, y, \bar{u}, \frac{\partial \bar{u}}{\partial x}, \frac{\partial \bar{u}}{\partial y}, \frac{\partial^2 \bar{u}}{\partial x^2}, \frac{\partial^2 \bar{u}}{\partial x \partial y}, \frac{\partial^2 \bar{u}}{\partial y^2}, \cdots\right) = \bar{f}(x, y) \tag{8.155}$$

з граничними умовами

$$G\left(x, y, \bar{u}, \frac{\partial \bar{u}}{\partial x}, \frac{\partial \bar{u}}{\partial y}, \cdots\right)\Big|_{\Gamma} = 0. \qquad (8.156)$$

де Γ — границя області D. Тут $\bar{F}$ і $\bar{G}$ — лінійні вектор-функції своїх аргументів,

Розв'язання крайової задачі (8.155), (8.156) для компонент вектора $\bar{u}$ шукатимемо у вигляді

$$u_j(x, y) = \sum_{i=0}^{N} u_{ij}(x)\varphi_{ij}(y), \quad j = \overline{1, l}, \qquad (8.157)$$

де функції $u_{ij}(x)$ підлягають визначенню, а $\varphi_{ij}(y)$ задаються у вигляді лінійних комбінацій B-сплайнів. При цьому B-сплайни слід вибирати таким чином, щоб їхній степінь був більшим за порядок старшої похідної відповідної компоненти вектора $\bar{u}$ в рівняннях системи (8.155), а лінійні комбінації задовольняли граничні умови при $y = \text{const}$ в (8.156). Після цього підставляємо вирази (8.157) у систему рівнянь (8.155) і вимагаємо, щоб рівняння задовольнялись у точках колокації $\xi \in [y_1, y_2]$. Тоді з (8.155) отримуємо систему звичайних диференціальних рівнянь відносно функцій $u_{ij}(x)$ $(i = \overline{0, N};\ j = \overline{0, l})$. Додаючи до них граничні умови, що одержуються з (8.156), приходимо до крайової задачі для системи звичайних диференціальних рівнянь, яку розв'язуємо методом дискретної ортогоналізації (глава 6).

Для прикладу розглянемо задачу про згин прямокутних пластин змінної товщини $h(x, y)$ під дією поперечного навантаження $q(x, y)$. У системі координат xoy — довжина сторін пластини a і b, а задача описується диференціальним рівнянням

$$D_M \Delta\Delta\omega + 2\frac{\partial D_M}{\partial x}\frac{\partial \Delta\omega}{\partial x} + 2\frac{\partial D_M}{\partial y}\frac{\partial \Delta\omega}{\partial y} + \Delta D_M \Delta\omega -$$

$$- (1 - \nu)\left(\frac{\partial^2 D_M}{\partial x^2}\frac{\partial^2 \omega}{\partial y^2} - 2\frac{\partial^2 D_M}{\partial x \partial y}\frac{\partial^2 \omega}{\partial x \partial y} + \frac{\partial^2 D_M}{\partial y^2}\frac{\partial^2 \omega}{\partial x^2}\right) = q, \qquad (8.158)$$

$$0 \le x \le a; \quad 0 \le y \le b,$$

де Δ — оператор Лапласа; $D_M = \dfrac{Eh^3}{12(1 - \nu^2)}$ — жорсткість пластини.

На контурах пластини $y = \text{const}$ задаються такі граничні умови:

1) контури жорстко закріплені

$$\omega = 0, \quad \frac{\partial \omega}{\partial y} = 0 \text{ при } y = 0, \ y = b; \qquad (8.159)$$

2) контури шарнірно оперті

$$\omega = 0, \quad \frac{\partial^2 \omega}{\partial y^2} = 0, \text{ при } y = 0, \ y = b; \tag{8.160}$$

3) один контур шарнірно опертий, а другий — жорстко закріплений:

$$\omega = 0, \quad \frac{\partial^2 \omega}{\partial y^2} = 0 \text{ при } y = 0;$$

$$\omega = 0, \quad \frac{\partial \omega}{\partial y} = 0 \text{ при } y = b. \tag{8.161}$$

Аналогічні умови можуть бути задані і на контурах $x = \text{const}$.
Розв'язок рівняння (8.158) шукатимемо у вигляді

$$\omega = \sum_{i=0}^{N} \omega_i(x)\psi_i(y), \tag{8.162}$$

де $\omega_i(x)$ — невідомі функції, а функції $\psi_i(y)$ визначаються виразами через B-сплайни п'ятого степеня:

$$\begin{aligned}
\psi_0 &= \alpha_{11}B_5^{-2} + \alpha_{12}B_5^{-1} + B_5^0;\\
\psi_1 &= \alpha_{21}B_5^{-1} + \alpha_{22}B_5^0 + B_5^1;\\
\psi_2 &= \alpha_{31}B_5^{-2} + \alpha_{32}B_5^0 + B_5^2;\\
\psi_i &= B_5^i, \quad i = \overline{3, N-3}; \qquad (8.163)\\
\psi_{N-2} &= \beta_{31}B_5^{N+2} + \beta_{32}B_5^N + B_5^{N-2};\\
\psi_{N-1} &= \beta_{21}B_5^{N+1} + \beta_{22}B_5^N + B_5^{N-1};\\
\psi_N &= \beta_{11}B_5^{N+2} + \beta_{12}B_5^N + B_5^{N-1},
\end{aligned}$$

де B_5^i ($i = \overline{-2, N+2}$, i — номер сплайна) — сплайни, що побудовані на рівномірній сітці Δ: $y_{-5} < y_{-4} < \ldots < y_N < y_{N+1} < \ldots < y_{N+5}$; $y_0 = 0$, $y_N = b$, α_{ij} ($i = 1, 2, 3$; $j = 1, 2$) — сталі коефіцієнти, які визначаються наперед у залежності від заданих граничних умов на краю пластини $y = 0$, а β_{ij} ($i = 1, 2, 3$; $j = 1, 2$) — аналогічно в залежності від граничних умов, що задані на краю $y = b$. Позначимо

$$A_\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix}; \quad A_\beta = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix}.$$

Тоді, якщо

$$A_\alpha = A_\beta = \begin{bmatrix} \frac{165}{4} & -\frac{33}{8} \\ 1 & -\frac{26}{33} \\ 1 & -\frac{1}{33} \end{bmatrix},$$

то функції $\psi_j(y)$ задовольняють умови жорсткого закріплення контурів (8.159). Якщо

$$A_\alpha = \begin{bmatrix} \frac{165}{4} & -\frac{33}{8} \\ 1 & -\frac{26}{33} \\ 1 & -\frac{1}{33} \end{bmatrix}; \quad A_\beta = \begin{bmatrix} 12 & -3 \\ -1 & 0 \\ -1 & 0 \end{bmatrix},$$

то задовольняються умови (8.161). При

$$A_\alpha = A_\beta = \begin{bmatrix} 12 & -3 \\ -1 & 0 \\ -1 & 0 \end{bmatrix}$$

задовольняються умови (8.160).

Перетворимо рівняння (8.158) до вигляду

$$\frac{\partial^4 \omega}{\partial x^4} = a_1 \frac{\partial^3 \omega}{\partial x^3} + a_2 \frac{\partial^4 \omega}{\partial x^2 \partial y^2} + a_3 \frac{\partial^3 \omega}{\partial x^2 \partial y} + a_4 \frac{\partial^2 \omega}{\partial x^2} +$$

$$+ a_5 \frac{\partial^3 \omega}{\partial x \partial y^2} + a_6 \frac{\partial^2 \omega}{\partial x \partial y} + a_7 \frac{\partial^4 \omega}{\partial y^4} + a_8 \frac{\partial^3 \omega}{\partial y^3} + a_9 \frac{\partial^2 \omega}{\partial y^2} + f, \qquad (8.164)$$

де $a_i = a_i(x, y)$ $(i = \overline{1, 9})$ і $f = f(x, y)$.

Вираз (8.162) при певних граничних умовах на контурах $y = \text{const}$ підставляємо в рівняння (8.164) і вимагаємо задоволення його в точках колокації $\eta_k \in [0, b]$ $(k = \overline{0, N})$:

$$\sum_{i=0}^{N} \omega_i^4(x)\psi_i(\eta_k) = a_1(x, \eta_k) \sum_{i=0}^{N} \omega_i'''(x)\psi_i(\eta_k) + a_2(x, \eta_k) \times$$

$$\times \sum_{i=0}^{N} \omega_i''(x)\psi_i''(\eta_k) + a_3(x, \eta_k) \sum_{i=0}^{N} \omega_i''(x)\psi_i'(\eta_k) +$$

$$+ a_4(x, \eta_k) \sum_{i=0}^{N} \omega_i''(x)\psi_i(\eta_k) + a_5(x, \eta_k) \sum_{i=0}^{N} \omega_i'(x)\psi_i''(\eta_k) +$$

$$+ a_6(x, \eta_k) \sum_{i=0}^{N} \omega_i'(x)\psi_i'(\eta_k) + a_7(x, \eta_k) \sum_{i=0}^{N} \omega_i(x)\psi_i^{(4)}(\eta_k) +$$

$$+ a_8(x, \eta_k) \sum_{i=0}^{N} \omega_i(x)\psi_i'''(\eta_k) + a_9(x, \eta_k) \sum_{i=0}^{N} \omega_i(x)\psi_i''(\eta_k) + f(x, \eta_k). \quad (8.165)$$

Співвідношення (8.165) — система $N + 1$ лінійних звичайних диференціальних рівнянь відносно функцій ω_i $(i = \overline{0, N})$. Введемо позначення:

$$\Psi_i = [\psi_i^{(j)}(\eta_k)], \quad k, i = \overline{0, N}, \quad j = \overline{0, 4};$$

$$\overline{\omega}^T = \{\omega_0, \omega_1, .., \omega_N\};$$

$$\overline{a_r}^T = \{a_r(x, \eta_0), a_r(x, \eta_1), \ldots, a_r(x, \eta_N)\}, \quad r = \overline{1, 9};$$

$$\overline{f}^T = \{f(x, \eta_0), f(x, \eta_1), \ldots, f(x, \eta_N)\}.$$

При $A = [a_{ij}]$ $(i, j = \overline{0, N})$ і $\overline{c} = \{c_0, c_1, \ldots, c_N\}$ через $\overline{c} * A$ позначимо матрицю $[c_i a_{ij}]$, тобто $\overline{c} * A = [c_i a_{ij}]$.

З урахуванням прийнятих позначень систему (8.165) запишемо у вигляді

$$\Psi_0\overline{\omega}^{(4)} = (\overline{a}_7 * \Psi_4 + \overline{a}_8 * \Psi_3 + \overline{a}_9 * \Psi_2)\overline{\omega} + (\overline{a}_5 * \Psi_2 + \overline{a}_6 * \Psi_1)\overline{\omega}' +$$

$$+ (\overline{a}_2 * \Psi_2 + \overline{a}_3 * \Psi_1 + \overline{a}_4 * \Psi_0)\overline{\omega}'' + \overline{a}_1 * \Psi_0\omega'' + \overline{f}. \quad (8.166)$$

При цьому покладаємо, що точки колокації вибрані таким чином, щоб матриця Ψ_0 була невиродженою. При використанні такого підходу за рахунок спеціального вибору точок колокації η_k $(k = \overline{0, N})$ при незначному N можна істотно збільшити порядок точності апроксимації. Зокрема, за точки колокації можна взяти корені поліномів Лежандра другого степеня на відрізку $[0, 1]$ так, що інтервал (y_{2i}, y_{2i+1}) матиме дві точки колокації, а сусідні інтервали (y_{2i-1}, y_{2i}) і (y_{2i+1}, y_{2i+2}) їх не матимуть. Такі точки колокації називаються *оптимальними*. Із (8.166) знаходимо

$$\omega^{IV} = \Psi^{-1}(\overline{a}_7 * \Psi_4 + \overline{a}_8 * \Psi_3 + \overline{a}_9 * \Psi_2)\overline{\omega} + \Psi_0^{-1}(\overline{a}_5 * \Psi_2 + \overline{a}_6 * \Psi_1)\overline{\omega}' +$$

$$+ \Psi_0^{-1}(\overline{a}_2 * \Psi_2 + \overline{a}_3 * \Psi_1 + \overline{a}_4 * \Psi_0)\overline{\omega}'' + \Psi_0^{-1}(\overline{a} * \Psi_0)\overline{\omega}'' + \Psi_0^{-1}\overline{f}_1. \quad (8.167)$$

Звідси видно, що матриці Ψ_j $(j = \overline{0, 4})$ мають стрічкову структуру. Систему (8.167) можна звести до нормального вигляду

$$\frac{d\overline{S}}{dx} = A(x)\overline{S} + \overline{F}(x), \quad 0 \le x \le a, \quad (8.168)$$

де $\overline{S}^T = \{\overline{\omega}, \overline{\omega}', \overline{\omega}'', \overline{\omega}'''\}$ — вектор-стовпець розмірності $4(N+1)$; $\overline{F}(x)$ — вектор-стовпець правої частини; $A(x)$ — квадратна матриця.

Граничні умови для системи (8.168) формулюються на основі граничних умов, що задані на краях $x = \text{const}$. Наприклад, нехай краї пластини $x = \text{const}$ жорстко закріплені, тобто при $x = 0$, $x = a$

$$\omega = 0, \quad \frac{\partial \omega}{\partial x} = 0. \tag{8.169}$$

Підставляючи вираз (8.162) в (8.169), одержуємо граничні умови для системи (8.168) при $x = 0$, $x = a$

$$\omega = 0, \quad \omega' = 0. \tag{8.170}$$

У загальному випадку граничні умови для системи (8.168) мають вигляд

$$D_1 \overline{S} = \overline{d}_1, \quad x = 0; \quad D_2 \overline{S} = \overline{d}_2, \quad x = a, \tag{8.171}$$

де D_1 і D_2 — задані прямокутні матриці відповідно порядку $2(N+1) \times 4(N+1)$; $\overline{d}_1$ і $\overline{d}_2$ — задані вектори. Одержана крайова задача (8.168), (8.171) розв'язується чисельним методом дискретної ортогоналізації.

Нехай квадратна пластина сталої товщини h зі стороною a перебуває під дією навантаження $q = q_0 \sin(\pi x/a)$. Два протилежних краї пластини жорстко закріплені, а два інші — шарнірно оперті, тобто маємо граничні умови

$$\omega = 0, \quad \frac{\partial \omega}{\partial y} = 0 \text{ при } y = 0, \quad y = a;$$

$$\omega = 0, \quad \frac{\partial \omega}{\partial y^2} = 0 \text{ при } x = 0, \quad x = a.$$

Апроксимація розв'язку B-сплайнами п'ятого степеня проводилась за координатою y за формулою (8.162). При цьому граничні умови на краях $y = \text{const}$ задовольнялись точно. При розв'язанні задачі використані дані: $a = 1$; $h = 0{,}1$; $\nu = 0{,}3$; $N = 7{,}9$. Як відомо, в цьому випадку, використовуючи метод відокремлення змінних у напрямі x, можна одержати точний розв'язок задачі. Цей приклад і розглядається, щоб порівняти наближений розв'язок із точним. Результати розв'язання задачі для прогину і других похідних наведено для $x = l/2$ і $y = l/2$ в табл. 32 і 33. Як видно з таблиць, відносна похибка не перевищує 0,12% при $N = 7$ і 0,06%; при $N = 9$, що свідчить про достатній ступінь точності даного підходу.

Розглянемо задачу про напружено-деформований стан прямокутних у плані пологих оболонок зі змінними параметрами. У системі координат xoy задача описується диференціальними рівняннями:

$$D_N\left(\frac{\partial^2 u}{\partial x^2}+\frac{1-\nu}{2}\frac{\partial^2 u}{\partial y^2}+\frac{1+\nu}{2}\frac{\partial^2 v}{\partial x\partial y}\right)+\frac{\partial D_N}{\partial x}\left[\frac{\partial u}{\partial x}+\frac{\omega}{R_1}+\nu\left(\frac{\partial v}{\partial y}+\frac{\omega}{R_2}\right)\right]+$$

$$+\frac{1-\nu}{2}\frac{\partial D_N}{\partial y}\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)=-D_N\left[\frac{1}{R_1}\frac{\partial \omega}{\partial x}+\omega\frac{\partial}{\partial x}\left(\frac{1}{R_1}\right)+\right.$$

$$\left.+\frac{\nu}{R_2}\frac{\partial v}{\partial x}+\nu\omega\frac{\partial}{\partial x}\left(\frac{1}{R_2}\right)\right];$$

$$D_N\left(\frac{\partial^2 v}{\partial y^2}+\frac{1-\nu}{2}\frac{\partial^2 v}{\partial x^2}+\frac{1+\nu}{2}\frac{\partial^2 u}{\partial x\partial y}\right)+\frac{\partial D_N}{\partial y}\left[\frac{\partial v}{\partial y}+\frac{\omega}{R_2}+\nu\left(\frac{\partial u}{\partial x}+\frac{\omega}{R_1}\right)\right]+$$

$$+\frac{1-\nu}{2}\frac{\partial D_N}{\partial x}\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)=-D_N\left[\frac{\nu}{R_1}\frac{\partial \omega}{\partial y}+\nu\omega\frac{\partial}{\partial y}\left(\frac{1}{R_1}\right)+\right.$$

$$\left.+\frac{1}{R_2}\frac{\partial \omega}{\partial y}+\omega\frac{\partial}{\partial y}\left(\frac{1}{R_2}\right)\right];$$

$$D_M\Delta\Delta\omega+2\frac{\partial D_M}{\partial x}\frac{\partial \Delta\omega}{\partial x}+\frac{\partial D_M}{\partial y}\frac{\partial \Delta\omega}{\partial y}+D_M\Delta\omega-$$

$$-(1-\nu)\left(\frac{\partial^2 D_M}{\partial x^2}\frac{\partial^2\omega}{\partial y^2}-2\frac{\partial^2 D_M}{\partial x\partial y}\frac{\partial^2\omega}{\partial x\partial y}+\frac{\partial^2 D_M}{\partial y^2}\frac{\partial^2\omega}{\partial x^2}\right)=$$

$$=-D_N\left[\frac{\partial u}{\partial x}+\frac{\omega}{R_1}+\nu\left(\frac{\partial v}{\partial y}+\frac{\omega}{R_2}\right)\right]\frac{1}{R_1}- \qquad (8.172)$$

$$-D_N\left[\frac{\partial v}{\partial y}+\frac{\omega}{R_2}+\nu\left(\frac{\partial u}{\partial x}+\frac{\omega}{R_1}\right)\right]\frac{1}{R_2}+q,$$

$$0\le x\le a;\quad 0\le y\le b.$$

Таблиця 32

$t=12y$	$w\left\vert\frac{10^{-3}q_0}{Eh^3}\right.$			$\frac{\partial^2 w}{\partial x^2}\left\vert\frac{10^{-3}q_0}{Eh^3}\right.$			$\frac{\partial^2 w}{\partial y^2}\left\vert\frac{10^{-3}q_0}{Eh^3}\right.$		
	Розв'язок в сплайнах		Точний розв'язок	Розв'язок в сплайнах		Точний розв'язок	Розв'язок в сплайнах		Точний розв'язок
	$N=7$	$N=9$		$N=7$	$N=9$		$N=7$	$N=9$	
0	0	0	0	0	0	0	632,7	633,0	633,2
1	1,757	1,758	1,759	−17,34	−17,35	−17,36	287,6	287,9	288,0
2	5,579	5,582	5,585	−55,06	−55,09	−55,12	58,33	58,13	58,18
3	9,851	9,857	9,861	−97,22	−97,27	−97,32	−90,67	−90,74	−90,74
4	13,53	13,53	13,54	−133,5	−133,6	−133,6	−182,2	−182,2	−182,3
5	15,96	15,97	15,98	−157,5	−157,6	−157,7	−231,3	−231,4	−231,5
6	16,81	16,82	16,83	−165,9	−166,0	−166,1	−246,9	−247,0	−247,0

Таблиця 33

$l = 12x$	$w \left\vert \frac{10^{-3}q_0}{Eh^3} \right.$			$\frac{\partial^2 w}{\partial x^2} \left\vert \frac{10^{-3}q_0}{Eh^3} \right.$			$\frac{\partial^2 w}{\partial y^2} \left\vert \frac{10^{-3}q_0}{Eh^3} \right.$		
	Розв'язок в сплайнах		Точний розв'язок	Розв'язок в сплайнах		Точний розв'язок	Розв'язок в сплайнах		Точний розв'язок
	$N = 7$	$N = 9$		$N = 7$	$N = 9$		$N = 7$	$N = 9$	
0	0	0	0	0	0	0	0	0	0
1	4,351	4,353	4,356	− 42,94	− 42,96	− 42,98	− 63,89	− 63,92	− 63,94
2	8,406	8,410	8,413	− 82,96	− 82,99	− 83,03	− 123,4	− 123,5	− 123,5
3	11,89	11,89	11,90	− 117,3	− 117,4	− 117,4	− 174,6	− 174,6	− 174,7
4	14,56	14,57	14,57	− 143,7	− 143,7	− 143,8	− 213,8	− 213,8	− 213,9
5	16,24	16,25	16,25	− 160,3	− 160,3	− 160,4	− 238,5	− 238,6	− 238,6
6	16,81	16,82	16,83	− 165,9	− 166,0	− 166,1	− 246,9	− 247,0	− 247,0

На контурах $y = \text{const}$ оболонки можуть бути задані такі граничні умови:

1) контури жорстко закріплені

$$u = v = \omega = 0, \quad \frac{\partial \omega}{\partial y} = 0 \tag{8.173}$$

при $y = 0$ і $y = b$;

2) контури шарнірно оперті

$$u = 0; \quad v = 0; \quad \omega = 0, \quad \frac{\partial^2 \omega}{\partial y^2} = 0 \tag{8.174}$$

при $y = 0$ і $y = b$;

3) один край контура шарнірно опертий, а другий — жорстко закріплений

$$u = 0; \quad v = 0; \quad \omega = 0, \quad \frac{\partial^2 \omega}{\partial y^2} = 0 \text{ при } y = 0;$$

$$u = 0; \quad v = 0; \quad \omega = 0, \quad \frac{\partial \omega}{\partial y} = 0 \text{ при } y = b. \tag{8.175}$$

Аналогічні граничні умови можуть бути задані й на контурах $x = \text{const}$.

Розв'язок задачі будемо шукати у вигляді

$$u = \sum_{i=0}^{N} u_i(x)\varphi_i(y);$$

$$v = \sum_{i=0}^{N} v_i(x)\varphi_i(y);$$

$$\omega = \sum_{i=0}^{N} \omega_i(x)\varphi_i(y), \tag{8.176}$$

де $u_i(x)$, $v_i(x)$, $\omega_i(x)$ — невідомі функції; $\Psi(y)$ визначаються за формулами (8.163), а функції $\varphi(y)$ виражаються через B-сплайни третього степеня таким чином:

$$\begin{gathered}
\varphi_0 = \gamma_{11}B_3^0 + \gamma_{12}B_3^{-1};\\
\varphi_1 = \gamma_{21}B_3^1 + \gamma_{22}B_3^0 + B_3^{-1};\\
\varphi_i = B_3^i, \quad i = \overline{2, N-2};\\
\varphi_{N-1} = \delta_{2i}B_3^{N-1} + \delta_{22}B_3^N + B_3^{N+1};\\
\varphi_N = \delta_{11}B_3^N + \delta_{12}B_3^{N+1},
\end{gathered} \tag{8.177}$$

де B_3^i $(i = \overline{-1, N+1})$ — кубічні B-сплайни, що побудовані на рівномірній сітці Δ: $y_{-3} < y_{-2} < \ldots < y_0 < y_1 < \ldots < y_N < y_{N+1} < y_{N+2} < y_{N+3}$; $y_0 = 0$, $y_N = b$, γ_{ij} $(i = 1, 2)$ — деякі сталі коефіцієнти, які визначаються в залежності від заданих граничних умов на краю оболонки $y = 0$, а δ_{ij} $(i, j = 1, 2)$ визначаються аналогічно в залежності від граничних умов, що задані на краю $y = b$.

Позначаючи

$$A = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}; \quad A = \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix},$$

маємо для випадку граничних умов (8.173)

$$A_\alpha = A_\beta = \begin{bmatrix} \frac{165}{4} & -\frac{33}{8} \\ 1 & -\frac{26}{33} \\ 1 & -\frac{1}{33} \end{bmatrix}, \quad A_\gamma = A_\delta = \begin{bmatrix} 1 & -4 \\ 1 & -1/2 \end{bmatrix};$$

для граничних умов (8.174)

$$A_\alpha\, A_\beta = \begin{bmatrix} 12 & -3 \\ -1 & 0 \\ -1 & 0 \end{bmatrix}, \quad A_\gamma = A_\delta = \begin{bmatrix} 1 & -4 \\ 1 & -1/2 \end{bmatrix};$$

для граничних умов (8.175)

$$A_\alpha = A_\beta = \begin{bmatrix} \frac{165}{4} & -\frac{33}{8} \\ 1 & -\frac{26}{33} \\ 1 & -\frac{1}{33} \end{bmatrix}; \quad A_\alpha A_\beta = \begin{bmatrix} 12 & -3 \\ -1 & 0 \\ -1 & 0 \end{bmatrix},$$

$$A_\gamma = A_\delta = \begin{bmatrix} 1 & -4 \\ 1 & -1/2 \end{bmatrix}.$$

Підставляючи вираз (8.176) у систему (8.172) і вимагаючи задоволення рівнянь системи в точках колокації $\eta_k \in [a, b]$ $(k = \overline{0, N})$ після ряду перетворень, як і у випадку згину пластин, дістанемо

$$\frac{d\bar{S}}{dx} = A(x)\bar{S} + \bar{F}(x), \quad 0 \le x \le a, \tag{8.178}$$

де $\bar{S}^T = \{\bar{u}, \bar{u}', \bar{v}, \bar{v}', \bar{\omega}, \bar{\omega}', \bar{\omega}'', \bar{\omega}'''\}$ — вектор-стовпець розмірності $8(N+1)$; $A(x)$ — квадратна матриця розмірності $8(N+1)$; $\bar{F}(x)$ — вектор-стовпець правої частини. Граничні умови для системи (8.178) знаходимо, враховуючи вирази (8.176), з умов (8.173) — (8.175). У загальному випадку їх можна записати у вигляді

$$D_1\bar{S} = \bar{d}_1, \quad x = 0; \quad D_2\bar{S} = \bar{d}_2, \quad x = a, \tag{8.179}$$

де D_1 і D_2 — прямокутні матриці розмірності $4(N+1) \times 8(N+1)$; d_1, d_2 — задані вектори. Крайову задачу (8.178), (8.179) розв'яжемо методом дискретної ортогоналізації.

Нехай полога сферична оболонка сталої товщини h і радіуса R рівномірно навантажена $q = \text{const}$. У плані оболонка має квадратну форму зі стороною a. На краях оболонки задані умови:

$$\begin{aligned} &u = v = w = \frac{\partial w}{\partial x} = 0 \text{ при } x = 0; \\ &u = v = w = \frac{\partial^2 w}{\partial x^2} = 0 \text{ при } x = a; \\ &u = v = w = \frac{\partial^2 w}{\partial y^2} = 0 \text{ при } y = 0; \\ &u = v = w = \frac{\partial w}{\partial y} = 0 \text{ при } y = a, \end{aligned} \tag{8.180}$$

тобто краї $y = 0$ і $x = a$ шарнірно оперті, а $x = 0$ і $y = b$ жорстко закріплені. Задача розв'язувалась при $a = 10$; $h = 0{,}4$; $R = 13$; $\nu = 0{,}3$; $N = 9$.

Результати розв'язання задачі наведено для прогину w і моментів M_1 і M_2 (табл. 34). Як випливає з таблиці, відносно діагоналі зберігається симетрія, що свідчить про їхню достатню точність.

Таблиця 34

$l = \begin{cases} y & x = 5 \\ 10 - x & y = 5 \end{cases}$	Значення прогину $-w/\frac{q}{E}$		Значення моментів			
			$M1/q$	$M2/q$	$M2/q$	$M1/q$
	$x = 5$	$y = 5$	$x = 5$	$y = 5$	$x = 5$	$y = 5$
0	0	0	0	0	0	0
0,625	71,38	71,50	0,0668	0,0678	0,1740	0,1750
1,250	132,1	132,3	0,1008	0,1020	0,2370	0,2386
1,875	178,0	178,2	0,1142	0,1154	0,2333	0,2342
2,500	209,6	209,7	0,1161	0,1171	0,1975	0,1980
3,125	229,2	229,3	0,1131	0,1139	0,1550	0,1551
3,750	239,9	240,0	0,1094	0,1100	0,1214	0,1213
4,375	244,1	244,1	0,1071	0,1077	0,1048	0,1045
5,000	242,8	242,8	0,1072	0,1077	0,1077	0,1072
5,625	235,8	235,9	0,1086	0,1091	0,1273	0,1271
6,250	221,9	222,0	0,1090	0,1097	0,1565	0,1566
6,875	198,9	199,1	0,1044	0,1053	0,1809	0,1816
7,500	165,1	165,4	0,0891	0,0901	0,1784	0,1794
8,125	120,5	120, 8	0,0552	0,0563	0,1154	0,1171
8,750	69,23	69,47	−0,0064	−0,0057	−0,0483	−0,0472
9,375	22,17	22,28	−0,1063	−0,1059	−0,3584	−0,3578
10,00	0	0	−0,2525	−0,2541	−0,8417	−0,8471

§ 8.11. МЕТОД, ЩО БАЗУЄТЬСЯ НА АПРОКСИМАЦІЇ ДИСКРЕТНИМИ РЯДАМИ ФУР'Є

Розглянемо підхід до розв'язання двовимірних крайових задач зі змінними параметрами в двох координатних напрямах під дією різних навантажень з довільними граничними умовами на контурах. Базується він на розкладанні в дискретні ряди Фур'є функцій, котрі входять у вихідні рівняння, що дає змогу звести задачу до одновимірної, яку можна розв'язати чисельно з достатньою точністю.

Нехай задача описується системою диференціальних рівнянь у частинних похідних

$$\frac{\partial Z_i}{\partial \alpha} = \psi_i\left(\alpha, \beta, \frac{\partial^k Z_j}{\partial \beta^k}\right) + f_i(\alpha, \beta), \quad i, j = \overline{1, l}; \ k = \overline{0, 4}, \qquad (8.181)$$

де $Z_i = Z_i(\alpha, \beta)$ $(\alpha_1 \le \alpha \le \alpha_2; \ \beta_1 \le \beta \le \beta_2)$ — шукані розв'язуючі функції; φ_i — лінійні оператори відносно своїх аргументів; $f_i(\alpha, \beta)$ — вільні члени; α, β — ортогональні криволінійні координати.

Для відкритих областей до цієї системи рівнянь додаються граничні умови на контурах $\alpha = \text{const}$ і $\beta = \text{const}$. Для замкнених в одному координатному напрямі областей граничні умови в цьому напрямі замінюються умовами періодичності.

Викладемо метод розв'язання даного класу задач. Шуканий розв'язок системи рівнянь подаємо у вигляді скінченного відрізка ряду

$$Z_i(\alpha, \beta) = \sum_{n=0}^{N} Z_{in}(\alpha)\psi_{in}(\beta), \quad i = \overline{1, l}, \qquad (8.182)$$

де $\psi_{in}(\beta$ — вибрані певним чином апроксимуючі функції, що залежать тільки від координати β; $Z_{in}(\alpha)$ — невідомі функціональні коефіцієнти, що підлягають визначенню і залежать від координати α.

Функції ψ_{in} повинні бути лінійно незалежні, належати повній системі функцій та задовольняти граничні умови на лініях $\beta = \text{const}$. Підставимо вирази (8.182) у рівняння системи (8.181) і подамо їхні праві частини у вигляді відрізка ряду за вибраними функціями $\psi_{in}(\beta)$:

$$\varphi_i\left(\alpha, \beta, \frac{\partial^k \sum_{n=0}^{N} Z_{jn}\psi_{jn}}{\partial \beta^k}\right) + f_i(\alpha, \beta) \equiv F_i(\alpha, \beta, Z_{in}) =$$

$$= \sum_{n=0}^{N} F_{in}(\alpha, Z_{jm})\psi_{in}(\beta), \quad i, j = \overline{1, l}; \ k = \overline{0, 4}; \ m = \overline{0, N}. \quad (*)$$

Помножуючи ліві й праві частини системи (8.181) на функції $\psi_{in}(\beta)$ та інтегруючи в інтервалі від β_1 до β_2, приходимо до зв'язаної системи звичайних диференціальних рівнянь порядку $l(N+1)$. Після деяких перетворень цю систему можна записати у вигляді

$$\frac{dZ_{in}}{d\alpha} = F_{in}(\alpha, Z_{im}), \quad i, j = \overline{1, l}; \ m, n = \overline{0, N}. \qquad (8.183)$$

Задача обчислення правих частин рівнянь системи (8.183) зводиться до знаходження відповідних коефіцієнтів Фур'є рядів (*). Для їх визначення пропонуємо алгоритм. Нехай відомо при деякому значенні змінної інтегрування $\alpha = \alpha_0$ поточні значення функцій $Z_{in}(\alpha_0)$ $(i = \overline{0, l}; \ n = \overline{0, N})$. Вибираємо на відрізку $[\beta_1, \beta_2]$ достатнє число точок β_s, в яких

$$\omega = \theta_y = 0 \text{ при } y = 0;$$

$$\hat{Q}_y = M_y = 0 \text{ при } y = a. \tag{8.185}$$

Розв'язок крайової задачі для системи (8.184), що задовольняє граничні умови на краях $x = 0$ і $x = a$, навантаження, а також функції φ_j ($i = 1, 2, 3$), які входять у праві частини рівнянь, подамо у вигляді розкладів

$$X(x, y) = \sum_{n=1}^{N} X_n(y) \sin \lambda_n x; \quad \varphi_2(x, y) = \sum_{n=1}^{n} \varphi_{2n}(y) \cos \lambda_n x;$$

$$X = \{\hat{Q}_y, M_y, \omega, \theta_y, \varphi_1, \varphi_3, q\}, \quad \lambda_n = \pi n / a. \tag{8.186}$$

Підставляючи розклади (8.186) в (8.184) і (8.185), дістанемо систему звичайних диференціальних рівнянь відносно амплітудних значень цих розкладів:

$$\frac{d\hat{Q}_{yn}}{dy} = \lambda_n^2 \left(\nu M_{ym} - \frac{E}{12} \varphi_{1n}\right) - q_n;$$

$$\frac{dM_{yn}}{dy} = \hat{Q}_{yn} + \frac{E}{6(1 + \nu)} \lambda_n \varphi_{2n};$$

$$\frac{\partial \omega_n}{\partial y} = -\theta_{yn};$$

$$\frac{\partial \theta_{yn}}{\partial y} = \frac{12(1 - \nu^2)}{E} \varphi_{3n} - \nu \lambda_n^2 \omega_n, \quad n = \overline{1, N}. \tag{8.187}$$

Граничні умови запишемо у вигляді

$$\omega_n = \theta_{yn} = 0 \text{ при } y = 0;$$

$$\hat{Q}_{yn} = M_{yn} = 0 \text{ при } y = a, \quad n = \overline{1, N}. \tag{8.188}$$

Тут для кожного n маємо:

$$\varphi_{1n} = \varphi_{1n}(y, \omega_m); \quad \varphi_{2n} = \varphi_{2n}(y, \theta_{ym});$$

$$\varphi_{3n} = \varphi_{3n}(y, M_{ym}), \quad m = \overline{1, N},$$

чим і визначається зв'язаність усіх $4N$ рівнянь системи (8.187). Одержані рівняння інтегруються одноразово для всіх гармонік. Для визначення в процесі інтегрування значень φ_{jn} ($j = 1, 2, 3$; $n = \overline{1, N}$) за поточним значенням амплітуд розв'язуваних функцій для фіксованого значення y обчислюємо в ряді точок x_i ($i = \overline{1, M}$) відрізка $[0, a]$ величини

$$h_i = h_0(1 + c_1 x_i + c_2 y);\quad \varphi_1^i(y) = -h_i^3(y) \sum_{n=1}^{N} \omega_n \lambda_n^2 \sin \lambda_n x_i;$$

$$\varphi_2^i(y) = h_i^3(y) \sum_{n=0}^{N} \theta_{y,n}(y) \lambda_n \cos \lambda_n x_i;$$

$$\varphi_2^i(y) = \frac{1}{h_i^3(y)} \sum_{n=0}^{N} M_{y,n}(y) \sin \lambda_n x_i.$$

Продовжуємо для фіксованого значення y функції φ_1, φ_3 непарним способом, а функції φ_2 — парним способом на відрізок $[a, 2a]$ і обчислюємо потім поточні значення $\varphi_{j,n}(y)$, використовуючи стандартну процедуру визначення коефіцієнтів Фур'є таблично заданої функції від змінної x. На початку інтегрування, враховуючи граничні умови, задаються початкові значення розв'язуючих функцій системи (8.187). Знайдені значення $\varphi_{i,n}(y)$ для фіксованого значення y підставляємо в (8.187) і продовжуємо інтегрування по y (виконуємо наступний крок).

Пластину розраховували при таких даних: $a = 1$; $h_0 = 0{,}1$; $c_1 = 0{,}25$; $c_2 = -0{,}5$; $\nu = 0{,}3$; $N = 6$; $M = 21$. Значення моментів M_x і M_y, прогину ω і кутів повороту θ_x, θ_y для середніх ліній пластини наведено в табл. 35. Одержані результати дозволяють судити про розподіл усіх факторів напружено-деформованого стану в залежності від умов закріплення країв і закону зміни товщини пластини.

Таблиця 35

$\frac{x}{a}$	$\frac{y}{a}$	$\frac{M_x}{q_0}$	$\frac{M_y}{q_0}$	$w\frac{E}{q_0}$	$\theta_x \frac{E}{q_0}$	$\theta_y \frac{E}{q_0}$
0	0,5	0	0	0	−264,43	0
0,2	0,5	0,02636	0,00298	48,306	−198,47	−150,36
0,4	0,5	0,04103	0,00416	74,219	−54,71	−228,56
0,6	0,5	0,03930	0,00368	70,158	90,24	−213,81
0,8	0,5	0,02337	0,00189	41,454	186,42	−125,24
1,0	0,5	0	0	0	217,47	0
0,5	0	−0,04158	−0,13861	0	0	0
0,5	0,2	0,00163	−0,04285	16,582	4,06	−146,71
0,5	0,4	0,03281	−0,00487	53,480	14,43	−213,73
0,5	0,6	0,04797	0,00928	99,754	30,14	−245,26
0,5	0,8	0,04941	0,01104	150,766	52,36	−265,04
0,5	1,0	0,04038	0	207,906	87,02	−321,46

КОНТРОЛЬНІ ЗАПИТАННЯ ТА ЗАВДАННЯ

До глави 1

1. У чому полягає задача інтерполяції функцій?
2. Дати визначення сплайна n-го ступеня.
3. Чим відрізняється інтерполяція сплайнами від інтерполяції многочленами?

До глави 2

1. Пояснити зв'язок методу Гаусса з розкладанням матриці на множники.
2. Як можна отримати обернену матрицю, застосовуючи прямі методи розв'язання СЛАР?
3. У чому суть загальної побудови ітераційних методів?

До глави 3

1. Сформулювати принцип стискаючих відображень.
2. У чому полягає основна ідея методу Ньютона?
3. У чому полягає ідея методу покоординатного спуску?
4. Який основний принцип методу продовження по параметру?

До глави 4

1. Дати загальну характеристику методів для вирішення повної проблеми власних чисел матриці.
2. У чому полягає основна ідея методу Лавер'е?
3. На якому принципі засновані прямий та ітераційний методи Якобі?
4. Як поліпшити проблему збіжності методу QR-алгоритму?

До глави 5

1. Яку геометричну інтерпретацію мають сталі в обчислювальному алгоритмі методу Рунге—Кутта?
2. Які особливості притаманні методам Рунге—Кутта та Ейлера?
3. Як пов'язані екстраполяційний та інтерполяційний методи Адамса з методами Адамса—Баушфорта та Адамса—Моултона?
4. У чому полягає проблема розв'язання задач для жорстких рівнянь?

До глави 6

1. Чим зумовлена нестійкість обчислювального процесу при розв'язанні лінійних крайових задач?
2. Яка ідея покладена в основу методів диференціальної та різницевої прогонки?
3. Чим відрізняється метод дискретної ортогоналізації від методів диференціальної та різницевої прогонки?
4. У чому полягає суть методу сплайн-колокації?

До глави 7

1. У чому полягає основна ідея методу зведення нелінійної крайової задачі до системи нелінійних рівнянь і задачі Коші?

2. Які методи ще застосовуються до розв'язання крайових задач для звичайних диференціальних рівнянь?

До глави 8

1. Чим відрізняється метод Власова—Канторовича від методів Ритца і Бубнова—Гальоркіна?

2. У чому полягає основна ідея методу, що базується на сплайн-апроксимації функцій в одному напрямі?

3. Як формулюються умови стійкої явної та неявної різницевих схем для рівнянь параболічного типу?

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Алберг Дэк, Нильсон Э., Уолт Дж. Теория сплайнов и ее приложение. М., 1972. 316с.

Березин И.С., Жидков Н.П. Методы вычислений: В 2 т. М., 1962. Т.1. 464 с.; Т.2. 620 с.

Бабушка И.Б., Витасек Э., Прагер М. Численные процессы решения дифференциальных уравнений. М., 1969. 368 с.

Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. М., 1987. 300 с.

Блехман И.И., Мышкис А.Д., Пановко Я.Г. Прикладная математика: предмет, особечности подходов. К., 1976. 270 с.

Вазов В., Форсайт Дж. Разностные методы решения дифференциальных уравнений в частных производных. М., 1963. 488 с.

Годунов С.К., Рябенький В.С. Разностные схемы, введение в теорию. М., 1973. 400 с.

Григоренко Я.М., Мукосд А.П. Розв'язання лінійних і нелінійних задач теорії оболонок на ЕОМ. К., 1992. 152 с.

Демидович Б.П., Марон И.А., Шувалова Э.З. Численные методы анализа. М., 1992. 368 с.

Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л. Методы сплайн-функций. М., 1980. 352 с.

Калиткин Н.Н. Численные методы. М., 1978. 512 с.

Коллатц Л. Численные методы решения дифференциальных уравнений. М., 1953. 503 с.

Крылов В.И., Бобков В.В., Монастырский П.И. Вычислительные методы: В 2 т. М., 1976., Т.1. 302 с.; Т2. 400 с.

Ланцош К. Практические методы прикладного анализа. М., 1961. 524 с.

Марчук Г.И. Методы вычислительной математики. М., 1977. 456 с.

На Ц. Вычислительные методы решения прикладных граничных задач. М., 1982. 296 с.

Ортега Д., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными. М., 1975. 558 с.

Ортега Д., Пул У. Введение в численные методы решения дифференциальных уравнений. М., 1986. 288 с.

Самарский А.А. Теория разностных схем. М., 1977. 656 с.

Самарский А.А., Гулин А.В. Численные методы. М., 1989. 432 с.

Современные численные методы решения обыкновенных дифференциальных уравнений / Под ред. Холл Дж., Уотт Дж. М., 1972. 312 с.

Уилкинсон Дж.Х. Алгебраическая проблема собственных значений. М., 1970. 364 с.

Хемминг Р.В. Численные методы. М., 1968. 400 с.

Шаманский В.Е. Методы численного решения краевых задач на ЭЦВМ. К., 1966. Ч.2. 244 с.

Навчальний посібник

Григоренко Ярослав Михайлович
Панкратова Наталія Дмитрівна

ОБЧИСЛЮВАЛЬНІ МЕТОДИ В ЗАДАЧАХ ПРИКЛАДНОЇ МАТЕМАТИКИ

Художник обкладинки *Г.Т. Задніпряний*
Художній редактор *Т.О. Щур*
Технічний редактор *Л.І.Швець*
Коректори *А.І.Бараз, Л.Ф.Іванова*

Здано до набору 29.09.94. Підп до друку 06.06.95.
Формат 60x84/16. Папір друк. №2. Гарн. Тип Таймс.
Офсет. друк. Ум.друк.арк. 16,27. Ум. Фарбовідб. 16,62.
Обл.-вид. арк. 18,55. Вид. №3584. Зам. № Д-220.

Оригінал-макет виготовлений у видавництві «Либідь»
на ПЕОМ типу IBM AT за допомогою програмного комплексу Xerox Ventura Publisher 2.0
інженером-програмістом *М.М.Білянською* та оператором *Т.В.Кулик*

Видавництво «Либідь» при Київському університеті ім. Тараса Шевченка.
252001 Київ, Хрещатик, 10

Київська книжкова друкарня наукової книги
252004 Київ, Б.Хмельницького, 19