МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Опорний конспект лекцій

з дисципліни

"Програмне забезпечення дискретних динамічних систем"

Тернопіль - 2013

Лекція 1

Вхід, стан та вихід системи. Класифікація та властивості систем

1.1. Вхід, стан і вихід системи

Динаміка системи описується її математичною моделлю. Така модель відображає математичні залежності між трьома множинами змінних: змінними входу, виходу і стану.

Вхід системи, що виражається або множиною тимчасових функції, або множиною тимчасових послідовностей вхідних значень, представляє зовнішні змінні, що діють на систему. **Вихід** системи, що виражається аналогічно входу, представляє опис безпосередньо спостережуваної поведінки системи.

Основна властивість будь-якої динамічної системи полягає в тому, що її поведінка у будь-який момент часу залежить не тільки від змінних, що діють на неї в даний момент часу, але і від змінних, що діяли на неї у минулому. Вважається, що така система володіє "пам'яттю", яка дозволяє враховувати внесок змінної, що діяла на неї з минулого моменту часу до моменту спостереження її поведінки. **Стан** системи, який визначається як множина значень так званих змінних стану, представляє миттєве значення "комірок" цій пам'яті. Якщо в довільний момент часу t_0 відомі стан і вхідний відрізок $u(t_0, t]$, то у будь-який момент часу $t \geq t_0$ можуть бути визначені вихід і стан системи.

Під **динамічною системою** розуміють об'єкт будь-якої природи, стан якого змінюється в часі у відповідності з деякою динамічною закономірністю, тобто в результаті дії детермінованого оператора еволюції. Таким чином, поняття динамічної системи є результат певної ідеалізації, яка полягає у зневажанні випадковими факторами, які присутні в будь-якій реальній системі. Відповідно, детермінований підхід правомірний у тих випадках, коли вплив випадкових сил (**флуктуацій**) дуже малий і їх статистичні характеристики не відіграють істотної ролі у визначенні характеристик поведінки системи.

Поняття "динамічна система" також дуже наближено. Звичайний сенс виразу "динамічний" майже самий, що і у виразу "причинний": минулі події впливають на майбутні події, але не навпаки. Математичний опис динамічної системи приводить

до підкреслення і формалізації напряму причинності від минулого до майбутньому. З математичної точки зору динамічна система є аксіоматичним поняттям. Щоб ввести математичну модель динамічної системи, необхідно визначити миттєвий стан системи у вигляді сукупності деяких величин або функцій та задати **оператор еволюції**, за допомогою якого знаходиться відповідність між станом системи в початковий момент часу і єдиним станом в кожний наступний момент часу. Оператор еволюції може бути заданий безліччю різних способів: за допомогою диференціальних чи інтегральних рівнянь, дискретних відображень послідовностей, за допомогою матриць, графів тощо.

Визначення. *Динамічною системою називається складний математичний об'єкт, який визначається наступними аксіомами.*

(а) Задані наступні множини: множина моментів часу T, множина станів X, множина вхідних значень U, множина допустимих вхідних функцій $\Omega = \{\omega : \mathrm{T} \to \mathrm{U}\}$, множина вихідних значень Y і множина вихідних функцій $\Gamma = \{\gamma : \mathrm{T} \to \mathrm{Y}\}$.

(б) (Напрям часу.) Множина T – впорядкована підмножина множини дійсних чисел.

(в) Простір допустимих вхідних функцій Ω, задовольняє наступним умовам:

(1) (Нетривіальність.) Множина Ω не порожня.

(2) (Зчленування вхідних дій.) Вхідний відрізок $\omega(\mathrm{t}_1, \mathrm{t}_2]$ – функція $\omega \in \Omega$, задана на тимчасовому інтервалі $(\mathrm{t}_1, \mathrm{t}_2] \cap \mathrm{T}$. Якщо $\omega, \omega' \in \Omega$ і $\mathrm{t}_1 < \mathrm{t}_2 < \mathrm{t}_3$, то знайдеться функція $\omega'' \in \Omega$, для якої $\omega''(\mathrm{t}_1, \mathrm{t}_2] = \omega(\mathrm{t}_1, \mathrm{t}_2]$ і $\omega''(\mathrm{t}_2, \mathrm{t}_3] = \omega'(\mathrm{t}_2, \mathrm{t}_3]$.

(г) Задана перехідна функція стану φ, яка визначає стан $\mathrm{x}(\mathrm{t}) = \varphi(\mathrm{t}; \mathrm{t}, \mathrm{x}, \omega) \in \mathrm{X}$, досягнутий у момент часу $\mathrm{t} \in \mathrm{T}$ при вхідній дії $\omega \in \Omega$, якщо в початковий момент часу $\mathrm{t} \in \mathrm{T}$ початковий стан $\mathrm{x} = \mathrm{x}(\tau) \in \mathrm{X}$. Функція φ володіє наступними властивостями:

(1) (Напрям часу.) Функція φ визначена для всіх значень $\mathrm{t} \geq \tau$ і необов'язково визначена для всіх значень $\mathrm{t} < \tau$.

(2) (Узгодженість.) Рівність $\varphi(\mathrm{t}; \mathrm{t}, \mathrm{x}, \omega) = \mathrm{x}$ виконується при всіх $\mathrm{t} \in \mathrm{T}$,

$x \in X$ і $\omega \in \Omega$.

(3) (Композиційна властивість.) Для будь-яких значень $t_1 < t_2 < t_3$ і будь-яких станів $x \in X$ і всіх входів $\omega \in \Omega$ має місце $\varphi(t_3; t_1, x, \omega) = \varphi(t_3; t_2, \varphi(t_2; t_1, x, \omega), \omega)$.

(4) (Причинність.) Якщо $\omega, \omega' \in \Omega$ і $\omega(t, \tau] = \omega'(t, \tau]$, то $\varphi(t; \tau, x, \omega) = \varphi(t; \tau, x, \omega')$.

(д) Існує відображення виходу $\eta : T \times X \to Y$, що визначає вихідну величину $y(t) = \eta(t, x(t))$. Відображення $\eta(\sigma, \varphi(\sigma; \tau, x, \omega))$ при $\sigma \in (\tau, t]$ є вихідним відрізком, тобто відрізком $\gamma \in (\tau, t]$ деякої вихідної функції, яка задана на інтервалі $(\tau, t]$.

Відзначимо, що пара (τ, x), де $\tau \in T$ і $x \in X$, представляє подію в динамічній системі. Множина $T \times X$ визначає простір подій в динамічній системі.

Система є такою, що **фізично реалізовується**, якщо її вихід і стан в довільний момент часу t_0 є функцією тільки тих входів, які впливали на систему до моменту часу t_0.

Також розглядається система, яка фізично реалізується, вихід і стан якої підлягають впливу $u(t, t_0]$, тобто вплив аж до моменту t_0 включно. В цьому випадку існує відображення виходу $\eta : T \times X \times U \to Y$, що визначає вихідну величину $y(t) = \eta(t, x(t), u(t))$ або $y(\sigma) = \eta(\sigma, \varphi(\sigma; \tau, , x, \omega), u(\sigma))$ при $\sigma \in (\tau, t]$.

Система називається **детермінованою**, якщо її вихід і стан у будь-який момент t можна достовірно визначити по її стану в деякий момент $t_0 < t$ і по відомому входу з напівзамкненого інтервалу $[t_0, t)$.

Система називається **стохастичною**, якщо інформація про її стан в деякий момент часу t_0 і про її вхід на інтервалі $[t_0, t)$ дозволяє визначити вихід системи і її стан у момент часу $t_0 > t$ тільки з певною вірогідністю або іншими статистичними засобами.

Оскільки вхід, стан і вихід описуються кінцевим числом змінних, то їх представляють векторами. Так, наприклад, вхід керованого об'єкту, що складається з m змінних, позначається вектором

$$u^{\gamma} = [u_1, u_2, ..., u_m] = [u_i], \; i = 1,2,..., m.$$

Аналогічно стан об'єкту виражається вектором змінних стану. Так, наприклад, в лінійному керованому об'єкті *n*-го порядку, який складається з послідовного з'єднання n ланок першого порядку, стан може бути представлений змінними, які вимірюються між окремими ланками і в кінці ланцюжка цих ланок. Позначимо ці змінні через $x_1, x_2, ..., x_n$.

В ролі змінних стану можуть бути вибрані, наприклад, похідні вихідної змінної $x_1(t), dx_1(t)/dt, ..., d^{n-1}x_1(t)/dt^{n-1}$ або будь-які інші функції, які зазвичай є лінійною комбінацією таких змінних системи, які утворюють вектор стану, що повністю визначає стан системи. Звідси витікає, що вибір компонент вектора стану достатньо довільний, важливо тільки, щоб ці компоненти повністю описували стан системи. При виборі компонент вектора стану віддається перевага таким змінним стану, які дозволяють спростити необхідні розрахунки або які легко можуть бути виміряні і т.д.

Вихід об'єкту також виражається вектором вихідних змінних. Цей вектор однозначно визначається вектором стану. Вихідний вектор може співпадати з вектором стану, або вихідні змінні можуть безпосередньо співпадати з деякими змінними вектора стану. Наприклад, в керованому об'єкті, який складається з ланцюжка послідовно сполучених ланок першого порядку, вихідна змінна ланцюга x_1, є однією з компонент як вектора стану

$$x^{\tau} = [x_1, x_2, ..., x_n],$$

так і вектора стану

$$x^{\tau} = [x_1(t), \frac{dx_1(t)}{dt}, ..., \frac{d^{n-1}x(t)}{dt^{n-1}}].$$

Множина всіх можливих входів u називається простором входів U. Аналогічна множина всіх можливих станів системи x називається простором станів X, а множина всіх можливих виходів системи y – простором виходів Y.

Якщо вектори входу, стану і виходу визначені в кожен момент часу t з деякого інтервалу (безперервний час), то мова йде про **безперервну** систему, безперервний керований об'єкт і т.д. Якщо вектори входу і стану визначені тільки в дискретні моменти часу t_k, де k є послідовністю чисел, зазвичай цілих з деякого інтервалу, то мова йде про **дискретні** системи.

1.2. Класифікація систем

В основу класифікації динамічних систем покладені визначення стану системи, властивості і спосіб задання оператора еволюції. Стан системи визначається сукупністю деяких величин $x_j, j=1,2,...,N$ чи функцій $x_j(r), r \in R^M$. Величини x_j являються **динамічними змінними**, які звичайно безпосередньо пов'язані з кількісними характеристиками, що спостерігаються і вимірюються в реальних системах (струм, напруга, швидкість, температура, концентрація речовини, чисельність популяції і т.д.). Множина всіх можливих станів динамічної системи називається її **фазовим простором**. Якщо x_j – величини, а не функції, та їх число N скінчене, то фазовий простір R^N цієї системи має скінчену розмірність. Системи з скінченою розмірністю фазового простору часто називають **системами із зосередженими параметрами**, так як їх параметри не повинні бути функціями просторових координат. Такі системи зазвичай описуються звичайними диференціальними рівняннями або відображеннями послідовностей.

Проте існує широкий клас систем з фазовим простором нескінченної розмірності. Так, якщо динамічні змінні x_j системи являються функціями деяких змінних $r_k, k=1,2,...,M$, то розмірність фазового простору нескінчена. Як правило, r_k – це просторові координати, тому параметри системи, яка моделюється, залежать від точки в просторі. Такі системи називаються **системами із розподіленими параметрами**, чи просто **розподіленими системами**. Розподілені системи частіше за все описуються диференціальними рівняннями в частинних похідних або інтегральними рівняннями. Ще одним прикладом систем з нескінченною розмірністю фазового простору слугують **системи, оператор еволюцій яких містить затримку в часі** T_3. В цьому випадку стан системи також задається набором функцій $x_j(t), t \in [0, T_3]$.

В залежності від властивостей оператора еволюції можна виділити декілька класів динамічних систем. Якщо оператор володіє властивістю суперпозиції (тобто являється лінійним), то система **лінійна,** якщо оператор цією властивістю не володіє, то система **нелінійна**. Якщо стан системи і оператор еволюції визначений для будь-якого моменту часу, то її називають системою з **неперервним часом** чи

потоком; якщо ж стан визначений лише в окремі моменти часу (тобто на дискретній множині значень), то говориться про систему з **дискретним часом** (**дискретну систему,** чи **каскад**). Для каскадів оператор еволюції зазвичай задається у вигляді дискретного відображення, яке часто називають **відображенням наслідування.** Якщо оператор еволюції не залежить від часу явним чином, то система являється **автономною,** тобто в ній відсутні адитивні чи мультиплікативні зовнішні впливи; в іншому випадку – система **неавтономна**. В залежності від того, володіє система властивістю зберігати елемент фазового об’єму (хоча б в середньому) під дією оператора еволюції чи не володіє, розрізняють **консервативні** (які зберігають енергію) і **неконсервативні** системи. Стиснення елемента фазового об’єму говорить про наявність втрат енергії. Системи із втратами називаються **дисипативними**. Ріст елементу фазового об’єму свідчить про додавання енергії в систему. Таку систему також називають **дисипативною, але з негативними втратами.**

1.3. Властивості систем

Досяжність, керованість і стабільність

При вирішенні завдань управління методами теорії простору станів враховуються деякі фундаментальні властивості динамічних систем, які не зустрічаються в класичній теорії управління, що оперує тільки вхідними і вихідними сигналами даної системи. Цими властивостями є досяжність, керованість і стабільність систем. Всі ці три поняття, а також і інші, визначаються строго. Вони представляють інструмент, який дозволяє стисло виразити і сформулювати умови, необхідні для вирішення завдань синтезу, тобто для розрахунку керованого пристрою, який забезпечує оптимальне в заданому сенсі управління.

Слід зазначити, що теорія простору станів, як і теорія управління в цілому, розглядає не реальні об'єкти, а їх математичні моделі. Отже, вживаний тут апарат може бути тільки математичним, і такими ж є рішення, які дає ця теорія. При розрахунку керуючого пристрою, який забезпечує оптимальне в певному значенні управління заданим об'єктом, теорія простору станів часто вимагає від керованого об'єкту виконання умов керованості і спостережуваності. Виконання цих вимог часто дозволяє

розраховувати оптимальне управління за допомогою простих математичних операцій. Це, проте, не означає, що некерована і неспостережувальна система не може бути керована субоптимально в практичному сенсі. Слід заздалегідь відзначити, що стабільність і виявлення об'єкту є вирішальними для реалізації управління ним.

Визначення. Стан $x(t_1)$ лінійної системи можна досягнути, якщо існує момент часу $t_0 < t_1$, де $(t_1 - t_0)$ – кінцевий інтервал, і такий вхід $u(t)$, який переводить початковий стан системи $x(t_0) = 0$ в бажаний стан $x(t_1)$.

Визначення. Стан $x(t_1)$ лінійної системи керований, якщо існує момент часу $t_2 > t_1$ і такий вхід $u(t)$, який переводить стан системи $x(t_1)$ у стан $x(t_2) = 0$ за умови, що інтервал $t_2 - t_1$ є кінцевий.

Існуюче поняття так званої повної досяжності системи у момент часу t означає, що будь-який стан $x(t) \in X$ досягнений або керований. У стаціонарних дискретних системах досяжність і керованість стану не залежать від моменту часу t_1. Якщо до того ж система безперервна, то кожен стан керований. Тому в літературі, присвяченій безперервним системам, зазвичай говорять тільки про керованість, оскільки тут відсутня відмінність між досяжністю і керованістю стану. Початковий стан $x(t_0)$ безперервної лінійної системи може бути переведений за кінцевий час в будь-який інший стан за допомогою відповідного вхідного сигналу. Для дискретної системи, що описується рівнянням

$$x(k+1) = Ax(k) + bu(k),$$

це справедливо, якщо

$$\det[A] \neq 0.$$

Вимога переходу системи з довільного стану в інший довільний стан можна замінити на вимогу, яка полягає в тому, що система, переведена з нульового стану в ненульовий стан $x(t_1)$, повинна бути такою, що переводиться в початковий нульовий стан за допомогою відповідного вхідного сигналу. Якщо система задовольняє такій умові, то її називають **оборотною**.

Визначення. Система $\{A, B\}$ стабільна, якщо існує така дійсна матриця K, що матриця $A - BK$ стійка, тобто що для всіх власних значень λ_i матриці $A - BK$ має місце $|\lambda_i| < 1, i = 1,2,..., n$.

Згідно іншому визначенню система $\{A, B\}$ стабільна, якщо і тільки якщо вихідні компоненти (моди), відповідні нестійким власним значенням матриці A, досяжні.

Дане визначення стверджує, що все, що необхідне для стабілізації, – це керуючий пристрій, вихідні керівні змінні якого представляють лінійне перетворення вектора стану, тобто

$$u(k) = -Kx(k).$$

Умови стабільності насправді – це умови стійкості замкнутої системи управління, що містить нестійкий об'єкт, в якій вхід елементу зворотного зв'язку (пристрою, що управляє) визначається вектором змінних стану керованого об'єкту.

Спостережуваність, відновлюваність і виявлення

У багатьох випадках стан системи не вимірюється і, отже, управління не може бути безпосередньо реалізоване. Таким чином, виникає питання, чи можна визначити вектор стану по вимірюваному виходу або по вимірюваних виходах об'єкту з багатьма входами і багатьма виходами? В зв'язку з цим розрізняють спостережуваність стану і відновлюваність стану.

Визначення. Стан $x(t_0)$ системи спостережуваний, якщо він може бути визначене по майбутніх значеннях вихідної змінної $y(t), t > t_0$, і якщо інтервал $t - t_0$ кінцевий.

Визначення. Стан $x(t_0)$ системи відновлюваний, якщо він може бути визначений по минулих значеннях вихідної змінної $y(t), t < t_0$ і якщо інтервал $t - t_0$ кінцевий.

Визначення. Пара $\{C, A\}$ системи виявлена, якщо існує така дійсна матриця R така, що матриця $A - RC$ стійка, тобто що для всіх власних значень λ_i матриці $A - RC$ має місце $|\lambda_i| < 1, i = 1,2,..., n$.

Лекція 2

Математичні моделі систем у просторі станів

2.1. Представлення звичайного диференціального рівняння рівняннями стану

Математична модель може бути отримана для керованого об'єкту, керуючого пристрою або для замкнутого контура управління. Якщо ми знаємо фізичний опис системи і можемо записати рівняння, що описують поведінку її окремих частин, то отримати рівняння стану системи звичайно дуже легко.

У багатьох випадках не всі змінні керованого об'єкту можна виміряти. Вимірними є лише ті, які складають вихідний вектор $y(t)$. Серед вимірних змінних зазвичай знаходяться і керовані змінні. Число взаємно незалежних керованих змінних визначає **розмірність системи**. Якщо кожна з N керованих змінних багатовимірної системи управляється незалежно, то число r вхідних змінних, тобто число керуючих змінних керованого об'єкту, повинне бути не менше N за умови, що всі вхідні змінні також взаємно незалежні. Вони зазвичай впливають на різні точки входу керованого об'єкту.

Порядок системи залежить від числа змінних стану. Цей порядок рівний числу змінних стану.

Якщо об'єкт з одним входом і одним виходом здатен відстежувати стрибкоподібні вхідні дії, то опис такого об'єкту в просторі станів має вигляд

$$x' = Fx + gu,$$

$$y = h^T x + ku.$$

У рівнянні виходу з'явився додатковий доданок ku. Якщо порядок об'єкту рівний n, то відповідні матриці в рівняннях мають розмірності $F(n;n)$, $g(n;1)$, $h^T(1;n)$ і $k(1;1)$, які узгоджені з розмінностями відповідних векторів. Система рівнянь стану складається з **рівняння динаміки** і **рівняння виходу**. У рівнянні динаміки F – матриця динаміки, а g – матриця-стовпець входу (матриця входу). У рівнянні виходу h^T – матриця-рядок виходу (матриця виходу), а k – коефіцієнт посилення по входу (матриця посилення по входу). Терміни в дужках відносяться до багатовимірних систем.

Іноді бажано знайти математичну модель в просторі станів за відомим диференціальним рівнянням, що описує динаміку об'єкту.

Процедура – змінні стану визначаються співвідношенням $x_{i+1} = x_j^{'}$

Розглянемо стаціонарний керований об'єкт з вхідною змінною u і вихідною змінною y, динаміка якого описується лінійним диференціальним рівнянням з постійними коефіцієнтами:

$$\sum_{i=0}^{n} a_i y^{(i)}(t) = \sum_{j=0}^{m} b_j u^{(j)}(t), \qquad n \geq m .$$

Дане рівняння заміняється наступними двома співвідношеннями (2.1, 2.2)

$$\tilde{A}x(t) = u(t), \tag{2.1}$$

$$\tilde{B}x(t) = y(t). \tag{2.2}$$

Якщо тепер ввести змінні стану

$$x(t) = x_1(t),$$

$$x^{(1)}(t) = x_2(t) = x_1^{(1)}(t),$$

.....

$$x^{(n-1)}(t) = x_n(t) = x_{n-1}^{(1)}(t),$$

то очевидно, що

$$x^{(n)}(t) = x_n^{(1)}(t),$$

і рівняння (2.1) може бути записане у вигляді

$$x_n^{(1)}(t) = \frac{1}{a_n} u(t) - \frac{a_{n-1}}{a_n} x_n(t) - \ldots - \frac{a_1}{a_n} x_2(t) - \frac{a_0}{a_n} x_1(t) \tag{2.3}$$

Співвідношення змінних стану і рівняння (2.3) можна записати одним рівнянням за допомогою постійних матричних операторів і вектора змінних стану:

$$\begin{bmatrix} x_1^{(1)}(t) \\ x_2^{(1)}(t) \\ \ldots \\ x_n^{(1)}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & \ldots & \ldots & -\frac{a_{n-1}}{a_n} \end{bmatrix} * \begin{bmatrix} x_1(t) \\ x_2(t) \\ \ldots \\ x_n(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \ldots \\ \frac{1}{a_n} \end{bmatrix} * u(t) \tag{2.4}$$

За допомогою змінних станів з рівняння (2.2) можна отримати рівняння, що визначає перетворення змінних стану $x_i(t), i = 1,2,\ldots, n$ у вихідну змінну $y(t)$:

$$b_0 x_1(t) + b_1 x_2(t) + \ldots + b_{n-1} x_n(t) + b_n x_n^{(1)}(t) = y(t), \quad n = m$$

Підставляючи вираз (2.3) в дане рівняння, отримаємо:

$$y(t) = (b_0 - \frac{a_0}{a_n} b_n)x_1(t) + (b_1 - \frac{a_1}{a_n} b_n)x_2(t) + ... + (b_{n-1} - \frac{a_{n-1}}{a_n} b_n)x_n(t) + \frac{b_n}{a_n} u(t). \qquad (2.5)$$

Рівняння (2.4) і (2.5) відповідають рівнянню динаміки та рівнянню виходу, де

$$F = \begin{bmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & ... & ... & -\frac{a_{n-1}}{a_n} \end{bmatrix}, \qquad g = \begin{bmatrix} 0 \\ 0 \\ ... \\ \frac{1}{a_n} \end{bmatrix},$$

$$h^T = \left[b_0 - \frac{a_0}{a_n} b_n, b_1 - \frac{a_1}{a_n} b_n, ..., b_{n-1} - \frac{a_{n-1}}{a_n} \right], \quad k = \frac{b_n}{a_n}.$$

Приклад. Застосувати процедуру для отримання рівняння стану, що відповідає диференціальному рівнянню

$$2y''(t) + 3y'(t) + 2y(t) = u(t) + 5u'(t).$$

Розв'язок. У відповідності з рівняннями (2.4) та (2.5) отримуємо

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} * u(t)$$

$$y(t) = [1,\ 5] * [x_1(t),\ x_2(t)]^T.$$

Лекція 3

Оператори. Z – перетворення

3.1. Основні властивості операторів

Розглянуті в другій лекції рівняння стану відображають перетворення векторів $u \in U$ і $x \in X$ у вектори $x' \in X'$ або $y \in Y$. Перетворення здійснюються за допомогою операторів F, g, h^T і k. Оскільки перетворення за допомогою операторів є однією з операцій, що найчастіше зустрічаються в просторі станів, то розглянемо основні типи і властивості цих операторів і деякі їх застосування у зв'язку з рівняннями стану.

Нехай X і Y – два векторні простори і кожен елемент $y \in Y$ однозначно відповідає елементу $x \in X$. Тоді цю відповідність стисло можна записати так:

$$x = Ay, \tag{3.1}$$

і говориться про оператор A, який визначений в просторі Y і який відображає Y в X.

Сума операторів A_1 і A_2 визначається співвідношенням

$$(A_1 + A_2)y = A_1y + A_2y \tag{3.2}$$

для всіх значень $y \in Y$.

Оскільки додавання у векторному просторі комутативне, то з (3.2) слідує, що

$$(A_1 + A_2)y = (A_2 + A_1)y, \tag{3.3}$$

$$[A_1 + (A_2 + A_3)]y = [(A_1 + A_2) + A_3)]y \tag{3.4}$$

Добуток операторів A_1A_2 - є послідовне застосування оператора A_2 до *y* і оператора A_1 до A_2y. Комутативний закон тут не виконується:

$$A_1A_2y \neq A_2A_1y \tag{3.5}$$

Якщо y *m*-вимірний вектор, який дорівнює сумі *m*-вимірних векторів, наприклад $y = y_1 + y_2$, то дистрибутивний закон в загальному випадку не виконується при перетворенні за допомогою оператора A:

$$A(y_1 + y_2) \neq Ay_1 + Ay_2 \tag{3.6}$$

Якщо помножити *m*-вимірний вектор на довільну константу (число) *a*, то знову отримуємо *m*-вимірний вектор:

$$f = ay \tag{3.7}$$

Оператор A називається *однорідним*, якщо

$$Aay = aAy \tag{3.8}$$

Спеціальними типами операторів є*: векторний* оператор, де

$$f = Ay, A = [A_1, A_2, ..., A_n], \tag{3.9}$$

і матричний оператор:

$$\begin{bmatrix} f_1 \\ f_2 \\ ... \\ f_n \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & ... & A_{1m} \\ A_{21} & A_{22} & ... & A_{2m} \\ ... & ... & ... & ... \\ A_{n1} & A_{n2} & ... & A_{nm} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_m \end{bmatrix} \tag{3.10}$$

де $f_i = \sum_{j=1}^{m} A_{ij}y_j, i = 1,2,...,n$ або $f = Ay$.

Лінійність

Оператор називається *лінійним,* якщо він володіє властивістю однорідності і дистрибутивності:

$$A(ay_1 + by_2) = aAy_1 + bAy_2 \tag{3.11}$$

Якщо умова (3.11) не виконується, то такий оператор – *нелінійний.*

Якщо математична модель об'єкта може бути описана диференціальними рівняннями, то рівняння стану з лінійним оператором приймає вигляд:

$$\frac{dx(t)}{dt} = F(t)x(t) + G(t)u(t) \tag{3.12}$$

$$y(t) = H(t)x(t) + K(t)u(t) \tag{3.13}$$

де $u(t)$ – r-вимірний вхідний вектор, $x(t)$ – n-вимірний вектор стану, $y(t)$ – p-вимірний вихідний вектор, а F, G, H і K – лінійні матричні оператори розміру $(n;n)$, $(n;r)$, $(p;n)$, і $(p;r)$ відповідно.

Для дискретних лінійних систем замість рівнянь (3.12) і (3.13) приймають

$$x(k+1) = A(k)x(k) + B(k)u(k) \tag{3.14}$$

$$y(k) = C(k)x(k) + D(k)u(k) \tag{3.15}$$

У рівняннях (3.12) – (3.15) всі лінійні оператори в загальному випадку є функціями незалежної змінної часу. Приведені рівняння стану описують системи із змінними в часі коефіцієнтами.

Стаціонарність

Оператор A називається *стаціонарним*, якщо він задовольняє співвідношення

$$x(t - t_1) = A[y(t - t_1)] \tag{3.16}$$

при всіх значеннях t і t_1, і для всіх значень $y \in Y$.

Система називається *стаціонарною*, якщо вона може бути описана стаціонарними операторами. Іншими словами, система, що описується диференціальними рівняннями з постійними коефіцієнтами, – стаціонарна система і, аналогічно, система описана рівняннями (3.12) і (3.13), – стаціонарна система, якщо оператори F, G, H і K – постійні матриці.

Подібним же чином ми можемо визначити стаціонарність дискретної системи.

Еквівалентність систем

Дві системи з векторами входу, стану і виходу u^1, x^1, y^1 і u^2, x^2, y^2 відповідно *еквівалентні по спостереженню*, якщо для всіх значень $u \in U$ і для всіх значень t з рівності

$$u^1(t) = u^2(t) \tag{3.17}$$

випливає рівність

$$y^1(t) = y^2(t) \tag{3.18}$$

Отже, у визначенні еквівалентності по спостереженню не проводиться порівняння стану систем, а порівнюються тільки зовнішні змінні, тобто входи і виходи систем.

При розгляді еквівалентності або відмінності внутрішніх властивостей систем потрібно ввести складніше визначення еквівалентності, яке повинне включати також стан системи. Проте таке визначення не повинне залежати від довільного вибору системи координат змінних стану. Іншими словами, вектори стану двох систем з еквівалентними станами у будь-який момент часу t можуть бути різними, проте повинна існувати можливість перетворення одного вектора стану в інший за допомогою лінійного постійного оператора.

Дві системи *строго еквівалентні*, якщо виконується умова еквівалентності по спостереженню і якщо із співвідношення (3.17) випливає

$$x^1(t) = Fx^2(t) \tag{3.19}$$

де F – невироджена постійна матриця.

3.2. Z-перетворення

Цифрова обробка сигналів є не що інше, як обробка послідовностей (дискретних значень сигналу). Для обробки безперервних функцій існує потужний математичний апарат на базі перетворення Лапласа. Проте застосування цього перетворення до послідовності неможливе. Воно проводиться т над функціями. Z-перетворення є, в деякому розумінні, аналогом перетворення Лапласа для послідовностей.

Z-перетворення над послідовністю $x(n)$ задається наступною формулою:

$$X(z) = \sum_{n=-\infty}^{n=\infty} x(n) z^{-n} \tag{3.20}$$

Що таке z? z- це звичайна комплексна змінна. Наприклад, існує послідовність, що складається всього з чотирьох членів:

$$x(0) = 8,$$
$$x(1) = -2$$
$$x(2) = 0$$
$$x(3) = 4$$

тоді Z-перетворення цієї послідовності згідно формули (3.20) буде наступним:

$$X(z) = 8 - 2z^{-1} + 4z^{-3} \tag{3.21}$$

Іншими словами, за допомогою Z-перетворення отримали з дискретної послідовності $x(n)$ безперервну функцію $X(z)$. При цьому необхідно відмітити, що $X(z)$ це не просто функція, а функція комплексної змінної. Тобто $X(z)$ визначена на комплексній площині z і значення $X(z)$ - теж комплексні величини. Дане перетворення називається *прямим*. Існує і *зворотне* Z-перетворення, коли з функції комплексної змінної $X(z)$ може бути отримана початкова послідовність $x(n)$, але таке перетворення використовується рідко, і тут його розглядувати не будемо.

3.3. Властивості Z-перетворення

Властивість 1: лінійність.

Цю властивість можна описати так: якщо послідовності $x(n)$ відповідає Z-перетворення $X(z)$, а послідовності $y(n)$ відповідає Z-перетворення $Y(z)$

$$x(n) \leftrightarrow X(z)$$
$$y(n) \leftrightarrow Y(z),$$

то суперпозиції цих послідовностей відповідає суперпозиція їх Z-перетворень:

$$ax(n) + by(n) \leftrightarrow aX(z) + bY(z), \tag{3.22}$$

a і b тут – звичайнікоефіцієнти.

Властивість 2: Z-перетворення затриманої послідовності.

Якщо послідовності $x(n)$ відповідає Z-перетворення $X(z)$

$$x(n) \leftrightarrow X(z),$$

то такій же послідовності, алезатриманою на k відліків відповідає Z-перетворення $z^{-k}X(z)$:

$$x(n-k) \leftrightarrow z^{-k}X(z). \qquad (3.23)$$

Тобто затримка послідовності призводитьдо домноження її Z-перетворення на z^{-k}.

Властивість 3: Z-перетворення згортки послідовностей.

Якщо послідовності $x(n)$ відповідає Z-перетворення $X(z)$, а послідовності $y(n)$ відповідає Z-перетворення $Y(z)$

$$\begin{matrix} x(n) \leftrightarrow X(z) \\ y(n) \leftrightarrow Y(z) \end{matrix},$$

то дискретній згортці послідовностей $x(n)$ і $y(n)$ відповідає добуток їх Z-перетворень:

$$x(n) * y(n) \leftrightarrow X(z)Y(z). \qquad (3.24)$$

3.4. Представлення різницевого рівняння рівняннями стану і зв'язок отриманого рішення із Z-перетворенням

В дискретному варіанті виводу рівнянь стану окремих частин замкнутої системи керування, які описуються лінійними різницевими рівняннями з постійними коефіцієнтами, необхідно розрізняти випадки функціонування окремих частин у дискретному часі та функціонування окремих частин у неперервному часі. Позначивши вхідну змінну контуру управління через u і вихідну змінну через y, можна записати відповідне різницеве рівняння у вигляді

$$\begin{matrix} a_n y(k) + a_{n-1} y(k+1) + ... + a_1 y(k+n-1) + a_0 y(k+n) = \\ = b_0 u(k+n) + b_1 u(k+n-1) + ... + b_n u(k) \end{matrix}. \qquad (3.25)$$

У випадку, коли окремі частини контуру управління функціонують у дискретному часі коефіцієнт b_0 може відрізнятись від нуля, тоді як у випадку неперервних елементів цей коефіцієнт завжди повинен дорівнювати нулю.

Введемо змінні стану

$$x_1(k) = a_0 y(k) - b_0 u(k), \qquad (3.26)$$

$$
\begin{aligned}
&x_2(k) = a_1 y(k) + a_0 y(k+1) - b_0 u(k+1) - b_1 u(k),\\
&x_3(k) = a_2 y(k) + a_1 y(k+1) + a_0 y(k+2) - b_0 u(k+2) - b_1 u(k+1) - b_2 u(k),\\
&\ldots\\
&x_n(k) = a_{n-1} y(k) + a_{n-2} y(k+1) + \ldots + a_0 y(k+n-1) -\\
&- b_0 u(k+n-1) - \ldots - b_{n-2} u(k+1) - b_{n-1} u(k).
\end{aligned}
\qquad (3.27)
$$

Знайшовши $y(k)$ з рівняння (3.26), підставивши отримане значення у всі рівняння (3.27) і замінивши в цих рівняннях всі члени, які містять $y(k+1)$ та $u(k+1),\ i = 1,2,\ldots, k+n-1$ на змінні стану $x_v(k+1),\ v = 1,2,\ldots, n-1$, отримаємо

$$
y(k) = \frac{1}{a_0} x_1(k) + \frac{b_0}{a_0} u(k), \qquad (3.28)
$$

$$
\begin{aligned}
&x_2(k) = \frac{a_1}{a_0} x_1(k) + x_1(k+1) - \left(b_1 - \frac{a_1}{a_0} b_0 \right) u(k),\\
&x_3(k) = \frac{a_2}{a_0} x_1(k) + x_2(k+1) - \left(b_2 - \frac{a_2}{a_0} b_0 \right) u(k),\\
&\ldots\\
&x_n(k) = \frac{a_{n-1}}{a_0} x_1(k) + x_{n-1}(k+1) - \left(b_{n-1} - \frac{a_{n-1}}{a_0} b_0 \right) u(k).
\end{aligned}
\qquad (3.29).
$$

Вихідне різницеве рівняння (3.25) може бути тепер записане за допомогою змінних стану наступним чином:

$$
0 = \frac{a_n}{a_0} x_1(k) + x_n(k+1) - \left(b_n - \frac{a_n}{a_0} b_0 \right) u(k). \qquad (3.30)
$$

Рівняння (3.29) та (3.30) визначають *перше рівняння стану (рівняння динаміки)*

$$
\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \ldots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} -\frac{a_1}{a_0} & 1 & 0 & \ldots & 0 \\ -\frac{a_2}{a_0} & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -\frac{a_n}{a_0} & 0 & 0 & \ldots & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \\ \ldots \\ x_n(k) \end{bmatrix} + \begin{bmatrix} b_1 - \frac{a_1}{a_0} b_0 \\ b_2 - \frac{a_2}{a_0} b_0 \\ \ldots \\ b_n - \frac{a_n}{a_0} b_0 \end{bmatrix} u(k). \qquad (3.31)
$$

Друге рівняння стану (рівняння виходу) визначається рівнянням (3.28).

Векторно – матрична форма рівнянь (3.31) та (3.28) має вигляд

$$
\begin{aligned}
&x(k+1) = Ax(k) + bu(k),\\
&y(k) = c^T x(k) + du(k),
\end{aligned}
\qquad (3.32)
$$

де структура окремих матричних операторів і векторів слідує із порівняння рівнянь (3.32) з рівняннями (3.31) та (3.28). Відмітимо, що тут A - матриця динаміки, $b(B)$ -

матриця-стовпчик входу (матриця входу), $c^T(C) = [1/a_0, 0, ..., 0]$ - матриця-стрічка виходу (матриця виходу) та $d(D) = b_0 / a_0$ - коефіцієнт входу (матричний коефіцієнт входу). Терміни у дужках відповідають багатовимірним системам.

Визначимо відповідне рівнянням (3.32) Z-перетворення:

$$X(z) = (E - z^{-1}A)^{-1} x(0) + z^{-1}(E - z^{-1}A)^{-1}(bU(z)). \qquad (3.33)$$

Звідси випливає, що Z-перетворення матриці A^k, де k - незалежна змінна рівне

$$Z[A^k] = (E - z^{-1}A)^{-1}. \qquad (3.34)$$

Z-перетворення вихідної змінної має вигляд

$$Y(z) = c^T X(z) + dU(z). \qquad (3.35)$$

Приклад. Визначити Z-перетворення різницевого рівняння $y(k) - 1{,}5y(k-1) + 0{,}5y(k-2) = u(k-1) + 3u(k-2)$.

Розв'язок. Матриці рівнянь стану дорівнюють

$$A = \begin{bmatrix} 1{,}5 & 1 \\ -0{,}5 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad c^T = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

$$Z[A^k] = \begin{bmatrix} 1-1{,}5z^{-1} & -z^{-1} \\ 0{,}5z^{-1} & 1 \end{bmatrix}^{-1} = \frac{1}{1-1{,}5z^{-1}+0{,}5z^{-2}} \begin{bmatrix} 1 & z^{-1} \\ -0{,}5z^{-1} & 1-1{,}5z^{-1} \end{bmatrix}.$$

$$X(z) = \begin{bmatrix} \dfrac{1}{\Delta(z)} & \dfrac{z^{-1}}{\Delta(z)} \\ -\dfrac{0{,}5z^{-1}}{\Delta(z)} & \dfrac{1-1{,}5z^{-1}}{\Delta(z)} \end{bmatrix} x(0) + \begin{bmatrix} \dfrac{z^{-1}+3z^{-2}}{\Delta(z)} \\ \dfrac{3z^{-1}-5z^{-2}}{\Delta(z)} \end{bmatrix} U(z),$$

де $\Delta(z) = 1 - 1{,}5z^{-1} + 0{,}5z^{-2}$.

Враховуючи вид матриці c^T, при $x(0) = 0$ отримуємо

$$Y(z) = \frac{z^{-1}+3z^{-2}}{1-1{,}5z^{-1}+0{,}5z^{-2}} U(z).$$

Лекція 4

Ідентифікація параметрів моделі. Метод найменших квадратів

4.1. Ідентифікація параметрів моделі

Ідентифікація динамічних об’єктів в загальному випадку складається з визначення їх структури і параметрів за даними спостережень – вхідному впливу і вихідній величині. Для рішення цієї задачі необхідно:

1) окреслити клас об’єктів;

2) вибрати модель;

3) вибрати критерій якості ідентифікації – середні втрати, які б характеризували різницю між вихідними величинами об’єкту і моделлю;

4) сформувати алгоритм ідентифікації, який, використовуючи доступні для спостереження значення вхідних та вихідних величин, змінював би параметри моделі так, щоб середні втрати з ростом часу досягали мінімуму.

Задача ідентифікації формулюється наступним чином: за результатами спостережень над вхідними й вихідними змінними об'єкта повинна бути побудована оптимальна в деякому сенсі модель, тобто формалізоване представлення об'єкта.

На вибір класу моделі, класу вхідних сигналів і критерію впливає апріорна інформація про об'єкт і вид його застосування (рис. 4.1).

При побудові моделей об'єктів слід приймати до уваги деякі їх особливості, зокрема:

1. *Наявність у системі елементів, сталі часу яких різко відрізняються внаслідок присутності в системі швидкої і повільної частин*, внаслідок чого:

– задача ідентифікації стає жорсткою, що у свою чергу приводить до повільної збіжності ітераційних процесів;

– тестові сигнали для таких об'єктів повинні містити велику кількість відліків, що у свою чергу збільшує об'єм розрахунків, необхідних для побудови моделі.

2. *Залежність поведінки об'єкта від форми механічних елементів*, яка проявляється у вигляді нелінійної зміни параметрів швидкої частини при зміні стану повільної, причому сама нелінійність часто важко апроксимується.

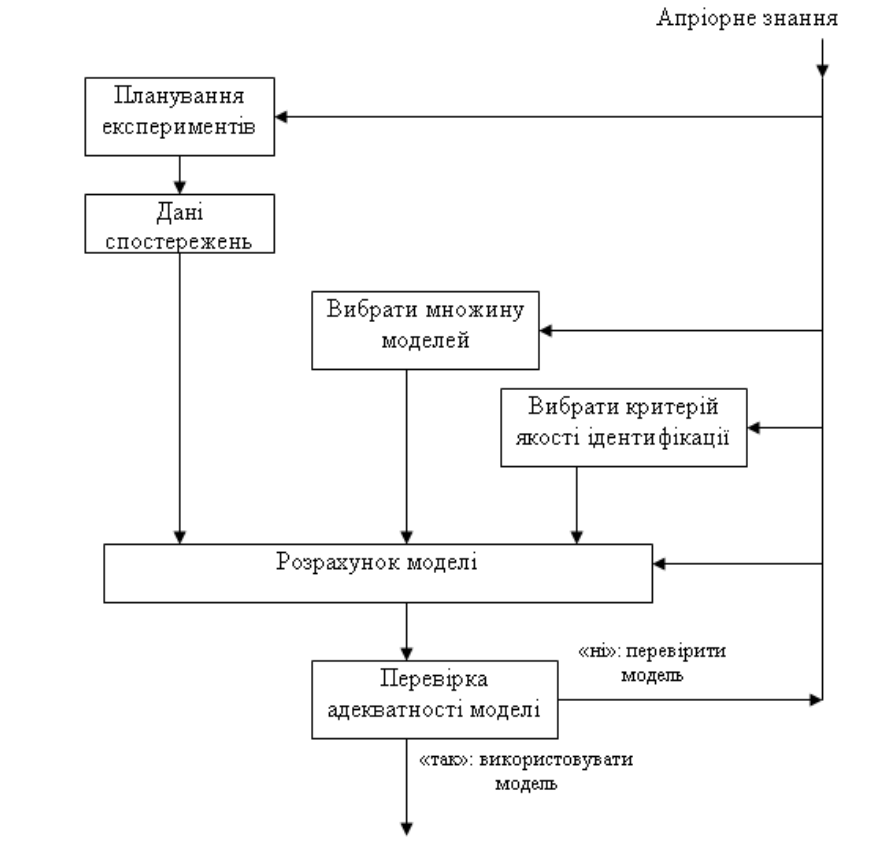

Рис. 4.1. Контур ідентифікації системи

2. *Залежність поведінки об'єкта від форми механічних елементів*, яка проявляється у вигляді нелінійної зміни параметрів швидкої частини при зміні стану повільної, причому сама нелінійність часто важко апроксимується.

3. *Наявність у системі елементів із розподіленими параметрами*, при моделюванні яких виникає велика кількість внутрішніх змінних, що приводить до ускладнення загальної моделі об'єкта.

4. *Неоднозначний початковий стан.* На відміну від чисто електричних систем, де початковий стан системи перед включенням живлення завжди є однаковий, економічна система може містити елементи, стан яких у початковий момент може бути різним. Цей фактор ускладнює ідентифікацію об'єкта, оскільки його початковий стан у деяких випадках виявляється невідомим.

Математичні макромоделі виявляються ефективними при проведенні моделювання систем з їх використанням внаслідок наступних особливостей:

1. Математичні моделі практично завжди є адаптованими для використання ЕОМ. Особливо це стосується дискретних моделей, для використання яких у

методах цифрового моделювання не потрібно використовувати жодних процедур апроксимації чи інтерполяції.

2. Математичні моделі здебільшого є набагато простіші за своєю структурою, ніж відтворюваний ними об'єкт. Це зумовлено тим, що математичні моделі здебільшого не включають в себе внутрішні особливості об'єкта, а відтворюють лише зовнішні його характеристики.

3. При побудові математичних моделей є можливість зробити акцент на певних характеристиках відтворюваного ними об'єкта, що дає змогу використовувати моделі, які особливо точно відтворюють ті характеристики об'єкта, які нас цікавлять в першу чергу.

В будь-якому динамічному об’єкті з плином часу відбуваються зміни. Ці зміни визначаються властивостями об’єкту, які відображаються в його перехідних характеристиках, тобто реакціях об’єкта на деякий вхідний вплив. Таким чином, динамічні режими, і як окремий випадок – перехідні характеристики, відображають поведінку об’єкту у ситуації, яка склалась, на основі причинно – наслідкової концепції системного аналізу.

Можна стверджувати, що існує визначений клас моделей поведінкового характеру. Ціллю створення таких моделей є відображення поведінки об’єкту при зовнішніх та внутрішніх змінах умов їх функціонування. Таким чином, модель такого виду може бути використана для зберігання динамічних властивостей об’єкта у компактній формі для імітації чи генерації реакцій об’єкта на різні зовнішні впливи, для прогнозування і можливого відновлення поведінки об’єкта у передуючі та майбутні проміжки часу. Все це обумовило широке застосування таких моделей у системах управління (вибір ціле направленого впливу), в системах діагностики (як еталон передуючих властивостей об’єкта), в системах моніторингу (як засіб оцінки і прогнозування виникаючих ситуацій) тощо.

У якості вхідних впливів при ідентифікації динамічних процесів для дискретних моделей використовуються або типові впливи (сходинкові, імпульсні), або будь-якого виду впливи (навіть випадкові послідовності).

Таким чином, у фіксовані моменти часу відбуваються зміни вхідних та вихідних змінних. Після отримання даних виникають наступні можливі задачі

ідентифікації. Відомі точні значення вхідних $u(t)$ та вихідних $y(t)$ змінних. Необхідно визначити дискретну модель у формі

- різницевого рівняння;
- дискретної передавальної функції;
- моделі у термінах простору станів;

які б найкращим чином апроксимували властивості і динамічні характеристики вихідного об’єкту.

При рішенні даної задачі можливі наступні варіанти:

1) задача параметричної ідентифікації – додатково відома структура моделі, і в даному випадку необхідно тільки визначити невідомі параметри моделі. У якості невизначеностей виступають: а) вибір періоду дискретизації; б) невідомі значення параметрів моделі;

2) задача непараметричної ідентифікації – коли невідома структура моделі, тобто до двох наперед вказаних невизначеностей додається ще третя – невизначеність структури;

3) окремий випадок задачі параметричної ідентифікації – значення вхідних $u(t)$ та вихідних $y(t)$ змінних містять певну похибку вимірювання. Тому тут, в порівнянні із задачею 1), додається ще одна (третя) невизначеність – невідома похибка вимірювань;

4) окремий випадок задачі непараметричної ідентифікації – тут, в порівнянні із задачею 2) додається четверта невизначеність – невідома похибка вимірювань.

Сформулюємо деякі зауваження про припущення відносно об’єкту управління, які найбільш часто використовуються при параметричній ідентифікації.

Як правило, передбачається, що об’єкт є лінійним або допускає лінеаризацію. Тобто об’єкт, який розглядається, повинен задовольняти принципу суперпозиції – якщо вихідна величина об’єкту $y(t)$ – це результат перетворення моделлю F^M вхідного впливу $u(t)$, тоді виконуються умови:

1) $F^M(au(t)) = ay(t)$, a – істотна константа;

2) $F^M(u_1(t) + u_2(t)) = y_1(t) + y_2(t)$.

Факт лінійності або нелінійності об’єкта може бути встановлений або експериментальним шляхом, або по рівняннях, які описують процеси перетворення інформаційних процесів.

Більш того, першочергово враховується, що об’єкт управління в динамічному відношенні є стійким, тобто починаючи з моменту часу t_0, до якого динамічна система знаходилась у стані рівноваги, реакція на кожну обмежену вхідну змінну $|u(t)| \le M_u < \infty$ буде також обмеженою вихідною змінною $|y(t)| = M_y < \infty$, а також стаціонарним – зсув вхідної функції на деяку часову константу призводить до зсуву вихідної функції на ту ж саму константу.

Стаціонарні лінійні системи, безперечно, являють собою найбільш важливий клас динамічних систем, які розглядаються в теорії та практиці. Треба розуміти, що такі системи відповідають ідеалізованому представленню про процеси, які реально протікають. Але не дивлячись на це, таке наближення виправдано, а проектні рішення, основані на використанні лінійної теорії, в багатьох випадках призводять до відмінних результатів.

У відповідності із розглянутими видами невизначеностей в задачі ідентифікації, сформулюємо основні допущення до побудови моделей динамічних об’єктів:

- об’єкт є стійким, керованим, спостережним, стаціонарним і лінійним, визначеного порядку;
- використовується лінійна за параметрами модель;
- вектор значень параметрів існує і єдиний;
- на вході динамічного об’єкту постійно діє вплив $u(t)$.

Розглянемо основні форми представлення неперервних та дискретних математичних моделей лінійних динамічних об’єктів. Використовуючи математику для опису фізичних явищ, ми лише проводимо наближення до реальності (моделювання). В даному випадку для опису поведінки системи в будь-який момент часу використовують єдину математичну характеристику – змінну стану.

Опис за допомогою змінної стану дозволяє представити більшість систем відповідними сукупностями диференціальних рівнянь першого порядку типу

$$x(t) = f(x(t), u(t), t) \tag{4.1}$$

і сукупностями функціональних співвідношень типу

$$y(t) = g(x(t), u(t), t) \tag{4.2}$$

де t - час, $u(t)$ – вхідні змінні, $y(t)$ – вихідні змінні та $x(t)$ – змінні стану.

Відомо, що рішення диференціального рівняння першого порядку залежить в першу чергу від початкової умови $x(t_0) = x_0$, де t_0 – початковий момент спостереження. Співвідношення (4.1) та (4.2) визначаються як стандартна форма рівнянь стану, а саме, вираз (4.1) є диференціальним рівнянням стану, а вираз (4.2) – рівняння типу вхід-стан-вихід.

Однак, в наш час відомий великий клас систем, в яких вхідні дії та вихідні сигнали визначені для деяких регулярних моментів часу на заданому інтервалі, і відповідно, ці системи не можуть бути описані диференціальними рівняннями. Такі системи називаються системами з дискретним часом (дискретні динамічні системи).

Оскільки вхідні та вихідні змінні дискретної динамічної системи визначаються тільки для деяких фіксованих точок часової осі, тобто $t_0, t_1, t_2, \ldots$, вони можуть бути представлені у вигляді послідовностей $(u(t_0), u(t_1), u(t_2), \ldots)$ та $(y(t_0), y(t_1), y(t_2), \ldots)$, або $(u_0, u_1, u_2, \ldots)$ та $(y_0, y_1, y_2, \ldots)$ відповідно, де $t_0, t_1, t_2, \ldots$ – точки на часовій осі, які нас цікавлять.

Лінійні стаціонарні системи, які функціонують у дискретному часі, можна описати лінійними різницевими рівняннями виду

$$y_{k+1} = ay_k + bu_k. \tag{4.3}$$

Відмітимо, що для рішення рівняння (4.3) необхідно єдина початкова умова y_0. Іншими словами, вихідна змінна y_n у будь-який момент часу $n \geq 0$ може бути виражена через початкову умову y_0 і вхідну послідовність $(u_0, u_1, \ldots, u_n)$ наступним чином:

$$\begin{aligned}
&y_1 = ay_0 + bu_0,\\
&y_2 = ay_1 + bu_1 = a(ay_0 + bu_0) + bu_1 = a^2y_0 + abu_0 + bu_1,\\
&y_3 = ay_2 + bu_2 = a(a^2y_0 + abu_0 + bu_1) + bu_2 = a^3y_0 + a^2bu_0 + abu_1 + bu_2,\\
&\vdots
\end{aligned}$$

Таким чином,

$$y_n = a^n y_0 + a^{n-1} b u_0 + a^{n-2} b u_1 + \ldots + b u_{n-1},$$

або

$$y_n = a^n y_0 + \sum_{k=1}^{n} a^{n-k} b u_{k-1}. \qquad (4.4)$$

Інший опис дискретної динамічної системи можна подати за допомогою різницевих рівнянь стану

$$x_{n+1} = a x_n + b u_n \qquad (4.5)$$

і співвідношення вхід-стан-вихід

$$y_n = c x_n + d u_n. \qquad (4.6)$$

Рішення рівняння (4.5) має вигляд

$$x_n = a^n x_o + \sum_{i=0}^{n-1} a^{n-i-1} b u_i \qquad (4.7)$$

та

$$y_n = c a^n x_0 + \sum_{i=0}^{n-1} c a^{n-i-1} b u_i + d u_n. \qquad (4.8)$$

Методи ідентифікації у більшості випадків при заданій обмеженій точності вимірювань не дозволяють побудувати складну модель, еквівалентну за структурою і параметрам реальному об’єкту. Але цей факт не заважає наступному використанню такої моделі, якщо, звичайно, вона відображає істотні сторони об’єкта. Більш того, саме в силу своєї простоти така модель найбільш придатна для наступного використання.

Ідентифікацію можна провести або методами фізико-математичного аналізу, або методами експериментального аналізу.

При ідентифікації методами фізико-математичного аналізу виходять з конструктивних даних і математичного опису найпростіших процесів, які мають місце в об’єкті, який вивчається. Таким чином отримують систему алгебраїчних і диференціальних рівнянь, які містять як вхідні і вихідні змінні, так і змінні стану. В ці рівняння іноді включаються надлишкові внутрішні змінні об’єкту, які можна і не враховувати. Якщо опис системи фізико-математичними методами повний, то

окремі рівняння стану можна впорядкувати і надати їм форму рівнянь стану. В цьому випадку компоненти вектора стану мають більш точний фізичний зміст.

При ідентифікації методами експериментального аналізу зазвичай знаходять математичну модель стійкого об’єкта за вимірами вхідних і вихідних величин. Для цієї цілі було розроблена досить велика кількість різних методів ідентифікації.

Методи ідентифікації систем можна розділити на детерміновані і статистичні методи. В детермінованих методах зазвичай припускають, що система повинна мати визначений початковий стан, наприклад, $x(t_0) = x_0$, і відносно простий вхідний сигнал, наприклад, прямокутний імпульс одиничної площини, одиничний стрибок, синусоїдальний сигнал і т.д.

В статистичних методах ідентифікації систем початковий стан і вхідний сигнал довільні. Крім корисного сигналу, на об’єкт діє завада, статистичні властивості якої можуть бути невідомі. Статистичні методи дозволяють виразити якість оцінювання через такі параметри, як, наприклад, дисперсія, коваріаційна матриця і т.д.

Різниця в окремих методах ідентифікації може бути викликана лінійністю або нелінійністю об’єкта, який ідентифікується, присутністю або відсутністю завад і можливістю їх виміру, присутністю або ні інформації про порядок або структуру моделі і т.д. Однак головне, що відрізняє окремі методи, – це тип математичної моделі і критерій якості побудови моделі. Відмітимо, що методи ідентифікації, які застосовуються для безпосереднього оцінювання матриць в рівняннях стану, розвинуті в меншій мірі, ніж методи, які застосовуються для побудови класичних моделей, таких як імпульсна перехідна функція, передавальна функція і т.д. Але це не є великим недоліком, оскільки вже відомі співвідношення, які досить добре пояснюють зв’язок між класичними моделями і описом систем в просторі станів. Однак будь-який з цих методів ідентифікації систем дозволяє знайти модель тільки тієї частини об’єкта, яка досягається і спостерігається, оскільки при обчисленнях можна використовувати тільки спостереження на вході і виході об’єкта.

В задачах параметричної ідентифікації при проведенні досліджень досить часто з’являються похибки вимірювань. Відносно цих похибок конкретної інформації немає, вони не ідентифікуються, проте їхній вплив на результат

дослідження є досить суттєвий. Так як природа похибки окремо від конкретного об’єкту не досліджується, то вона визнається випадковою, і для її статистичного опису використовуються апарат теорії ймовірностей.

Таким чином, задача параметричної ідентифікації з наявною похибкою вимірювань спирається на відомі в математичній статистиці методи: метод найменших квадратів (МНК) та метод максимальної правдоподібності (ММП).

4.2. Метод найменших квадратів

В методі найменших квадратів за критерій узгодження експериментальних і розрахункових даних прийнята сума квадратів відхилень

$$\psi = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \ .$$

За допомогою методу найменших квадратів при розв'язуванні задачі параметричної ідентифікації моделі оцінку $\vec{g}$ вектора невідомих параметрів моделі отримуємо за формулою:

$$\vec{g} = (F^T \cdot F)^{-1} \cdot F^T \cdot \vec{y},$$

де F - матриця значень базових функцій моделі, розрахованих в точках експерименту.

Лекція 5

Інтервальний підхід при ідентифікації параметрів моделі. Метод допустимого оцінювання параметрів моделі.

5.1. Інтервальний підхід при ідентифікації параметрів моделі

Застосування інтервального аналізу визначається цілим рядом переваг:

- не потребує знання імовірнісних характеристик невизначених факторів, які рідко бувають точно відомі на практиці;
- отримують чіткі оцінки для самих величин, які знаходяться, а не для ймовірностей або математичних сподівань, що має велике значення при наявності малої кількості вимірювань параметрів і одній або декількох реалізацій;

- статистичні характеристики не можуть гарантувати певний результат одного конкретного дослідження;
- у всіх випадках даються гарантовані двосторонні апроксимації шуканих рішень.

В загальному випадку точність інтервального результату повністю визначається наступними чотирма факторами:

1. Невизначеністю в задані вихідних даних.
2. Заокругленнями при виконанні операцій, які змінюють або породжують інтервальні об’єкти.
3. Наближеним характером методу, який використовується.
4. Ступеню врахування залежностей між інтервальними об’єктами, які приймають участь в розрахунках (змінними і константами).

У межах інтервального аналізу зручно виділити методи аналізу інтервальних даних, під якими будемо розуміти методи, направлені на розв’язування задач моделювання з інтервальними невизначеностями в експериментальних даних, дослідження механізмів впливу невизначеностей на їх формування, отримання та дослідження математичних моделей об’єктів з множинними оцінками параметрів.

Методи інтервального аналізу та їхній розвиток створили передумови розвитку трьох напрямків наукової та практичної діяльності, пов’язаної з математичним моделюванням об’єктів на основі інтервальних даних:

- математичний, в межах якого проводяться дослідження математичних проблем інтервальних обчислень;
- комп’ютерний, в межах якого досліджуються питання створення та використання інтервальних обчислень;
- прикладний, в межах якого відбувається використання методів інтервального аналізу і відповідних машинних засобів для побудови та дослідження математичних моделей широкого класу об’єктів.

Класична інтервальна арифметика являє собою алгебраїчну систему, носій якої – множина всіх дійсних інтервалів $x := [x^-, x^+] = \{x \in R \mid x^- \le x \le x^+\}$, а бінарні операції – додавання, віднімання, множення і ділення визначені у відповідності до наступного фундаментального принципу:

$$x * y := \{x * y \mid x \in X, \quad y \in Y\}, \tag{5.1}$$

для всіх інтервалів X, Y, таких, що виконання точкової операції $x * y$, $* \in \{+, -, \cdot, / \}$, має значення для будь-яких $x \in X$ і $y \in Y$. Розгорнуте визначення інтервальних арифметичних операцій таке:

$$X + Y = [x^- + y^-,\ x^+ + y^+], \tag{5.2}$$

$$X - Y = [x^- - y^+,\ x^+ - y^-], \tag{5.3}$$

$$X \cdot Y = [\min\{x^- y^-, x^- y^+, x^+ y^-, x^+ y^+\},\ \max\{x^- y^-, x^- y^+, x^+ y^-, x^+ y^+\}], \tag{5.4}$$

$$X / Y = x \cdot [1/y^+,\ 1/y^-], \quad \text{для } y \notin 0. \tag{5.5}$$

Ширина кінцевого інтервалу в інтервальних обчисленнях залежить від порядку здійснення операцій, які володіють певними властивостями, найголовнішими з яких є:

1. Комутативність:

$$\begin{aligned} &[x^-, x^+] + [y^-, y^+] = [y^-, y^+] + [x^-, x^+], \\ &[x^-, x^+] \cdot [y^-, y^+] = [y^-, y^+] \cdot [x^-, x^+] \end{aligned} \tag{5.6},$$

2. Асоціативність:

$$\begin{aligned} &([x^-, x^+] + [y^-, y^+]) + [z^-, z^+] = [x^-, x^+] + ([y^-, y^+] + [z^-, z^+]), \\ &([x^-, x^+] \cdot [y^-, y^+]) \cdot [z^-, z^+] = [x^-, x^+] \cdot ([y^-, y^+] \cdot [z^-, z^+]) \end{aligned} \tag{5.7},$$

3. Субдистрибутивність:

$$[x^-, x^+] \cdot ([y^-, y^+] + [z^-, z^+]) = [x^-, x^+] \cdot [y^-, y^+] + [x^-, x^+] \cdot [z^-, z^+]. \tag{5.8}$$

Основною властивістю інтервальних обчислень є **монотонність включення**. Нехай маємо наступні інтервали $\longrightarrow$ $[x^-, x^+]$, $[y^-, y^+]$, $[z^-, z^+]$, $[c^-, c^+]$, тоді отримаємо для них включення у такому вигляді:

$$[x^-, x^+] \subset [z^-, z^+],\ [y^-, y^+] \subseteq [c^-, c^+] \Rightarrow [x^-, x^+] \cap [y^-, y^+] \subseteq [z^-, z^+] \cap [c^-, c^+] \tag{5.9}$$

Саме ця властивість дозволяє побудувати ітераційні процедури наближення множин розв’язків для задач інтервальними даними. При цьому розміри множини визначаються шириною інтервалів вхідних даних та можливостями ітераційної процедури.

Операції над векторами і матрицями в інтервальній арифметиці визначаються аналогічно відповідним операціям класичної інтервальної арифметики.

Сумою (різницею) двох матриць однакової розмірності є інтервальна матриця того ж розміру, яка утворюється поелементними сумами (різницями) операндів. Якщо X – матриця розмірності $m \times l$ і Y – матриця розмірності $l \times n$, то добутком матриць X та Y буде матриця Z розмірності $m \times n$, така що

$$z_{ij} = \sum_{k=1}^{l} x_{ik} y_{kj}.$$

Відома особливість інтервального матричного множення в класичній арифметиці – відсутність асоціативності.

Операції $\cup$ та $\cap$ у застосуванні до інтервальних векторів відбуваються покомпонентно:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cup \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \cup y_1 \\ x_2 \cup y_2 \\ \vdots \\ x_n \cup y_n \end{pmatrix} \text{ та } \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cap \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \cap y_1 \\ x_2 \cap y_2 \\ \vdots \\ x_n \cap y_n \end{pmatrix}. \qquad (5.10)$$

Аналогічно в покомпонентному змісті буде виконуватись відношення "$\leq$" між інтервальними векторами.

Достатньо простий спосіб представлення результатів у випадку застосування інтервального підходу, а також властивість монотонного включення, яка притаманна усім чисельним процедурам інтервального аналізу, спричинюють розвиток його методів, а також широке застосування для розв'язування багатьох задач, пов'язаних із моделюванням систем.

5.2. Метод допустимого оцінювання параметрів моделі

Принципи побудови методу допустимого оцінювання параметрів моделей лінійних динамічних систем у випадку адитивних та обмежених за амплітудою похибок в каналах вимірювань базуються на властивостях множини допускових оцінок параметрів цих моделей. Реалізація методу передбачає два етапи: знаходження початкового наближення $\vec{g}_0$; покращення початкового наближення до забезпечення умови.

Перший етап ітераційного методу допустимого оцінювання лінійних динамічних систем - *вибір початкового наближення в ітераційному методі пошуку допустимого розв'язку ІСЛАР*.

Початкове наближення $\vec{g}_0$ до допустимого розв'язку $\vec{g}_{dop}$ обчислюється, виходячи із наближеного представлення множини допустимих оцінок параметрів, як розв'язок довільно вибраних *m*-рівнянь ІСЛАР (5.11):

$$x^-_{k+1} \le g_1 \cdot [x^-_{1,k}, x^+_{1,k}] + \ldots + g_i \cdot [x^-_{i,k}, x^+_{i,k}] + \ldots + g_m \cdot [x^-_{m,k}, x^+_{m,k}] + q \cdot u_k \le x^+_{k+1},\ k = 0,\ldots, N-1 \quad (5.11)$$

із заміною інтервалів $[x^-_{i,k}, x^+_{i,k}], \forall i = 1,.., m, \quad \forall\ k = 0,\ldots, m-1$ на їх точкові значення $x^+_{i,k}, \forall i = 1,.., m, \quad \forall\ k = 0,\ldots, m-1$. Аналіз властивостей допускової області параметрів, в цьому випадку розв'язок кожної нерівності сформованої у такий спосіб ІСЛАР в просторі оцінок параметрів $\vec{g}$ задає «гіперсмугу» $\overline{\Omega}_p$ (рис. 5.1).

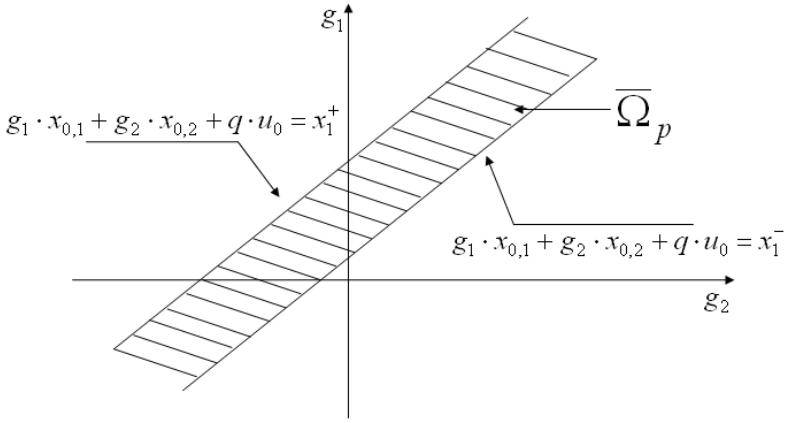

Рис.5.1. Ілюстрація розв'язку одного рівняння ІСЛАР у просторі параметрів $(m = 2)$ для $x^-_{ik} = x^+_{ik}$

Натомість перетин *m* таких «гіперсмуг» утворює множину Ω_m, яка в просторі параметрів є *m*-вимірним паралелепіпедом (рис. 5.2).

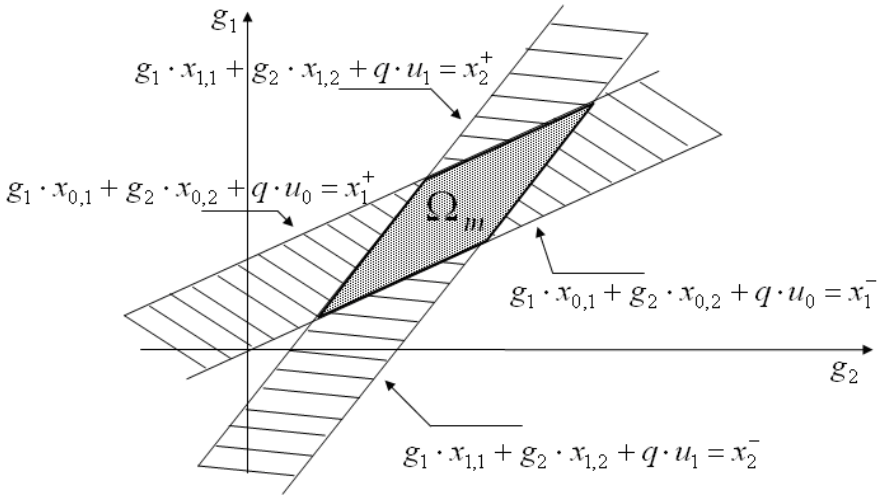

Рис. 5.2. Розв'язок системи рівнянь у вигляді *m*- вимірного паралелепіпеда $(\text{для } m = N-1 = 2)$

При цьому справедливим є таке включення $\vec{g}_{dop} \in \Omega_{dop} \subset \Omega_m$. Тоді за початкове наближення $\vec{g}_0$ доцільно вибрати центр симетрії m-вимірного паралелепіпеда (рис.5.3).

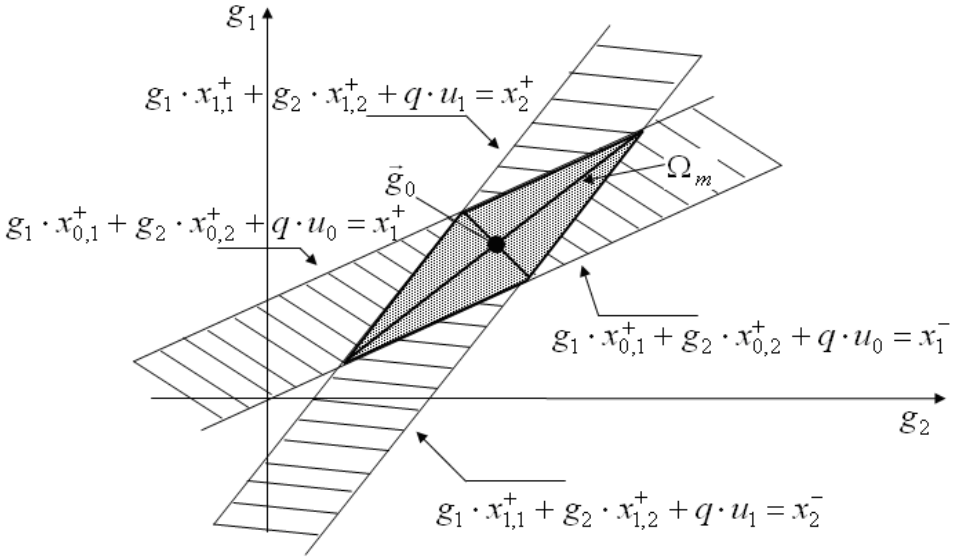

Рис. 5.3. Ілюстрація процедури вибору початкового наближення до допустимого розв'язку ІСЛАР (для $m = N = 2$)

Матричне представлення початкового наближення має такий вигляд::

$$\vec{g}_0 = (X^+)^{-1} \cdot \vec{\bar{x}}_{k+1},$$

$$X^+ = \begin{pmatrix} x_{0,1}^+ & \cdots & x_{0,m}^+ & u_0 \\ \vdots & & \vdots & \vdots \\ x_{m-1,1}^+ & \cdots & x_{m-1,m}^+ & u_m \end{pmatrix} \qquad (5.12)$$

де X^+ – матриця верхніх меж інтервалів m змінних стану для вибраних m рівнянь ІСЛАР (7.11); $\vec{\bar{x}}_{k+1} = \left(\frac{x_{1,k+1}^- + x_{1,k+1}^+}{2}, ..., \frac{x_{m,k+1}^- + x_{m,k+1}^+}{2} \right)^T$ – вектор, компоненти якого є середини відповідних інтервалів $[x_{k+1}^-, x_{k+1}^+]$, $k = 0,..., m-1$.

Отриманий за виразом (5.12) розв'язок схематично проілюстровано на рис.5.3.

Пропоноване вище значення $\vec{g}_0$ є достатньо грубим наближенням розв'язку $\vec{g}_{dop}$, однак його вибір зменшує кількість обчислень при пошуку розв'язку $\vec{g}_{dop}$ і його знаходження не вимагає використання складних обчислювальних алгоритмів.

Другий етап ітераційного методу допустимого оцінювання лінійних динамічних систем - *ітераційний метод покращення початкового наближення.*

Кожна *l*+1-а ітерація методу складається з трьох кроків.

Крок 1. Генерування випадкового вектора ξ в околі радіусом *r* наближеного розв'язку, отриманого на *l*-й ітерації ($\mathrm{r} = \mathrm{const}$).

Крок 2. Обчислення нового наближення $\vec{\mathrm{g}}_{l+1}$.

Крок 3. Перевірка "якості" отриманого наближення.

Графічно реалізації даного методу наведена на рис. 5.4.

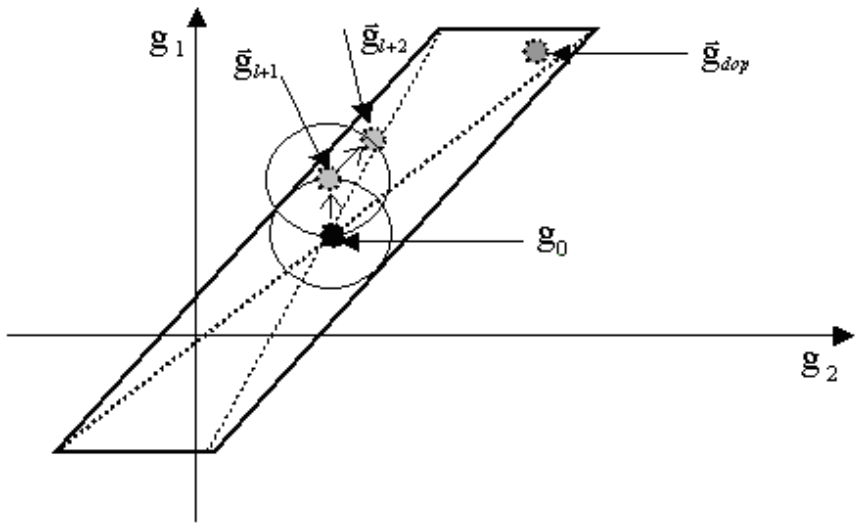

Рис. 5.4. Ілюстрація ітераційного методу пошуку допустимого розв'язку ІСЛАР (для $\mathrm{m} = \mathrm{N} = 2$)

При цьому на кожній $(l+1)$ –й ітерації знаходиться наближення $\vec{\mathrm{g}}_{l+1}$, яке задовольняє такій умові:

$$\left\|\vec{\mathrm{g}}_{l+1} - \vec{\mathrm{g}}_{\mathrm{dop}}\right\| \le \left\|\vec{\mathrm{g}}_{l} - \vec{\mathrm{g}}_{\mathrm{dop}}\right\|. \qquad (5.13)$$

Очевидно, що умову (5.13) на кожній ітерації перевірити не можливо, оскільки $\vec{\mathrm{g}}_{\mathrm{dop}}$ є невідомим.

Вводиться співвідношення, яке буде описувати оцінку якості поточного наближення до допустимого розв'язку ІСЛАР у такому вигляді

$$\delta = \max_{\mathrm{k+1=1...N}} \left\{ \mathrm{wid}([\hat{\mathrm{x}}_{\mathrm{k+1}}^{l+1}]) - \mathrm{wid}([\hat{\mathrm{x}}_{\mathrm{k+1}}^{l+1}] \bigcap [\mathrm{x}_{\mathrm{k+1}}]) \right\}, \qquad (5.14)$$

де δ – кількісна величина, яка задає якість наближення; $\mathrm{wid}(\bullet)$ – оператор виділення ширини інтервалу; $[\hat{x}_{k+1}^{l+1}]$, $k = 0,..., N-1$ - прогнозний коридор, знайдений на основі наближення $\vec{g}_{l+1}$ за формулою:

$$[\hat{x}_{k+1}^{l+1}] = g_{1,l+1} \cdot [\hat{x}_{1k}^{-}, \hat{x}_{1k}^{+}] + ... + g_{m,l+1} \cdot [\hat{x}_{mk}^{-} \hat{x}_{mk}^{+}] + q \cdot u_k . \qquad (5.15)$$

Інтервали $[\hat{x}_{k+1}^{l+1}]$, $k = 0,..., N-1$ є оцінками вектора параметрів стану динамічного об'єкта в *k*-й дискретний момент часу, знайдені на (*l*+1) -у кроці ітераційного методу.

Отже, для кожної ітерації умову (13) можна переписати у такому вигляді

$$\begin{aligned} &\max_{k+1=1...N} \left\{ \mathrm{wid}([\hat{x}_{k+1}^{l+1}]) - \mathrm{wid}([\hat{x}_{k+1}^{l+1}] \bigcap [x_{k+1}]) \right\} \leq \\ &\leq \max_{k+1=1...N} \left\{ \mathrm{wid}([\hat{x}_{k+1}^{l}]) - \mathrm{wid}([\hat{x}_{k+1}^{l}] \bigcap [x_{k+1}]) \right\}. \end{aligned} \qquad (5.16)$$

Прийнявши до уваги, що $\vec{g}_{l=1} = \vec{g}_0$, знайдемо вектор параметрів стану, отриманий на (*l*+1) -й ітерації, згенерувавши при цьому випадковий вектор пошуку $\vec{\xi}_1$ в околі початкового наближення радіусом *r* за формулою:

$$\vec{\xi}_1 = r \cdot \left(\frac{\Delta g_{1l}}{R_1}, \frac{\Delta g_{2l}}{R_1}, ..., \frac{\Delta g_{nl}}{R_1}, \frac{\Delta q_1}{R_1} \right), \qquad (5.17)$$

де $\Delta g_{1l}, \Delta g_{2l}, ..., \Delta g_{nl}, \Delta q_1$ – випадкові числа, згенеровані відповідно до рівномірного закону розподілу на інтервалі; $R_1 = \sqrt{\Delta g_{1l}^2 + \Delta g_{2l}^2 + ... + \Delta g_{nl}^2 + \Delta q_1^2}$, значення радіусу $r = \mathrm{const}$.

На наступному кроці обчислюємо наближення до допустимого розв'язку ІСЛАР:

$$\vec{g}_{l+1} = \vec{g}_l + \vec{\xi}_1 . \qquad (5.18)$$

Формально задача знаходження допустимого вектора параметрів $\vec{g}_{dop} \in \Omega_{dop}$ інтервальної моделі лінійної дискретної динамічної системи зводиться до задачі

$$\delta = \max_{k+1=1...N} \left\{ \mathrm{wid}([\hat{x}_{k+1}^{l+1}]) - \mathrm{wid}([\hat{x}_{k+1}^{l+1}] \bigcap [x_{k+1}]) \right\} \xrightarrow{\vec{g}_{l+1}} \min . \qquad (5.19)$$

Умова $\vec{g}_{l+1} \in \Omega_{dop}$ еквівалентна умові

$$\delta = \max_{k+1=1...N} \left\{ \mathrm{wid}([\hat{x}_{k+1}^{l+1}]) - \mathrm{wid}([\hat{x}_{k+1}^{l+1}] \bigcap [x_{k+1}]) \right\} = 0 . \qquad (5.20)$$

Для розв'язування мінімаксної задачі (5.19) запропонована процедура випадкового пошуку оптимального розв'язку. Алгоритм реалізації ітераційної процедури випадкового пошуку має такий вигляд.

Крок 1. Генерування випадкового вектора пошуку $\vec{\xi}_l$ за формулою (5.17).

Крок 2. Обчислення наступного наближення $\vec{g}_{l+1}$ за формулою (5.18).

Крок 3. Обчислення $[\hat{x}_{k+1}^{l+1}],\ k=0,...,N-1$ за формулою (5.15).

Крок 4. Перевірка якості наближення за умовою (5.16). У випадку не виконання умови, у формулі (5.17) замість r покладемо $(-r)$ і перехід на крок 2.

Крок 5. Перевірка умови (5.20) і у випадку її виконання завершення процедури. У протилежному випадку перехід на крок 1.

Проте недоліком методу по-ітераційного покращення початкового наближення допустимого розв'язку ІСЛАР є сталість параметра r в процесі реалізації ітераційної процедури, що призводить до поганої збіжності. Останнє зумовило введення адаптивної процедури настроювання параметра r у формулі (5.17).

Для початку модифікується формула (7.17) до вигляду:

$$\vec{\xi}_l = r_l \cdot (\frac{\Delta g_{1l}}{R_l}, \frac{\Delta g_{2l}}{R_l}, ..., \frac{\Delta g_{nl}}{R_l}, \frac{\Delta q_l}{R_l}), \qquad (5.21)$$

і знаходиться початкове наближення радіусу r_l, тобто $r_{l=0}$.

Процедура адаптивного настроювання параметра r_l на кожній ітерації виконується в два кроки.

Перший крок – формування початкового значення параметра $r_{l=0}$. Для цього із множини розв'язків Ω вибирається одна множина Ω_m (наприклад, ту, яка використовується для знаходження початкового наближення $\vec{g}_0$ до допустимого розв'язку $\vec{g}_{dop}$ ІСЛАР). Конфігурація вибраної множини відома – m- вимірний паралелепіпед. Початкове значення параметра $r_{l=0}$ приймається рівним половині довжини найменшої діагоналі множини Ω_m.

Вершини і довжина діагоналей множини Ω_m розраховується згідно формули (5.22):

$$\vec{g}_p = (X^+)^{-1} \cdot \vec{x}_{ik+1}^{p},\ p=1,...,2^m, i=1,...,m, \qquad (5.22)$$

де вектор $\vec{g}_p$ є однією із вершин многогранника Ω_m, утвореного перетином відповідних площин, заданих інтервальними рівняннями системи, причому $\vec{x}_{ik+1}^{p}$ -

вектор, складений із нижніх $x^{-}_{i,k+1}$ та верхніх $x^{+}_{i,k+1}$ меж відповідних інтервалів $[x^{-}_{k+1}, x^{+}_{k+1}]$, $k = 0,..., m-1$; та формули (5.23):

$$l = \left\| \vec{g}_p - \vec{g}_s \right\|, \tag{5.23}$$

де $\vec{g}_p, \vec{g}_s$, $p,s = 1...2^m$, $p \neq s$, $\vec{g}_s$ – відповідні вершини многогранника Ω_m.

Найменша діагональ вибирається відповідно до формули

$$r_{l=0} = \frac{\min\limits_{p,s=1,...,2^m, p \neq s} \left\| \vec{g}_p - \vec{g}_s \right\|}{2}. \tag{5.24}$$

Знайдений на першому кроці адаптивної процедури параметр $r_{l=0}$ підставляємо у формулу (5.21) для знаходження випадкового вектора $\vec{\xi}_l$ і обчислюємо наближення до розв'язку ІСЛАР $\vec{g}_{l+1}$ за формулою (5.18). Генерування випадкового вектора $\vec{\xi}_l$ із параметром $r_{l=0}$ здійснюємо до тих пір, доки відбувається виконання умови (5.19). У разі виникнення зациклення, тобто, коли на протязі декількох ітерацій генерування $\vec{\xi}_l$ не призводить до покращення наближення до розв'язку ІСЛАР – переходимо до другого кроку процедури адаптивного настроювання параметра випадкового пошуку r_l - зменшення параметра r_l шляхом його половинного ділення:

$$r_l = \frac{r_{l-1}}{2}. \tag{5.25}$$

На рисунку 5.5 схематично зображено процедуру випадкового пошуку з адаптивною процедурою пошуку.

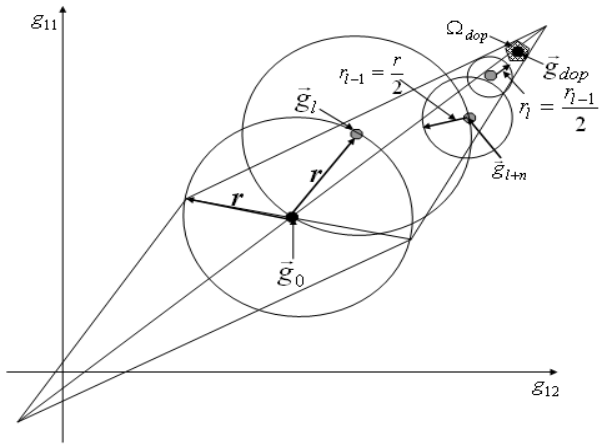

Рис. 5.5. Ілюстрація ітераційного методу пошуку з адаптивною процедурою настроювання параметрів r_l

Лекція 6

Чисельне диференціювання та інтегрування

6.1 Чисельне диференціювання

Нехай є функція $f(x)$ яку необхідно продиференціювати кілька разів і знайти цю похідну в деякій точці.

Якщо заданий явний вид функції, то вираз для похідної часто виявляється достатньо складним і бажано його замінити простішим. Якщо ж функція задана тільки в деяких точках (табличний), то отримати явний вид її похідних взагалі неможливо. У цих ситуаціях виникає необхідність наближеного (чисельного) диференціювання.

Ідея чисельного диференціювання полягає в тому, що функція замінюється інтерполяційним поліномом (Лагранжа, Ньютона) і похідна функції наближено замінюється відповідною похідною інтерполяційного многочлена.

$$f^{(m)}(x) \approx L_n^{(m)}(x),$$
$$f^{(m)}(x) \approx l_n^{(m)}(x), 0 \le m \le n.$$

Розглянемо прості формули чисельного диференціювання. Припустимо, що функція задана у рівновіддалених вузлах

$$x_i = x_0 + ih, h > 0, i = 0, \pm 1, \pm 2, \ldots.$$

Її значення і значення похідних у вузлах позначимо:

$$f(x_i) = f_i, \quad f'(x_i) = f_i', \quad f''(x_i) = f_i''.$$

Нехай функція задана у двох точках x_0 і $x_1 = x_0 + h$, та її значення f_0, f_1. Побудуємо інтерполяційний многочлен першого степеня

$$l_1(x) = f(x_0) + (x - x_0) f(x_0; x_1).$$

Похідна $l_1'(x)$ дорівнює $l_1'(x) = f(x_0; x_1) = \dfrac{f_1 - f_0}{h}$.

Похідну функції *f*(*x*) у точці x_0 наближено замінюємо похідною інтерполяційного многочленна

$$f_0'(x) \approx \frac{f_1 - f_0}{h}. \qquad (6.1)$$

Величина $\frac{f_1 - f_0}{h}$ називається ***першою різницевою похідною***. Нехай $f(x)$ задана у трьох точках $x_0, x_1 = x_0 + h,\ x_{-1} = x_0 - h.$

Інтерполяційний многочлен Ньютона другого степеня має вигляд:

$$l_2(x) = f(x_0) + (x - x_0)f(x_0; x_1) + (x - x_0)(x - x_1)f(x_0; x_1; x_{-1}).$$

Беремо похідну

$$l_2'(x) = f(x_0; x_1) + (2x - x_0 - x_1)f(x_0; x_1; x_{-1}).$$

у точці x_0, яка дорівнює

$$l_2'(x_0) = \frac{f_1 - f_0}{x_1 - x_0} + (x_0 - x_1) \times \left[\begin{array}{l} \frac{f_0}{(x_0 - x_1)(x_0 - x_{-1})} + \frac{f_1}{(x_1 - x_0)(x_1 - x_{-1})} + \\ + \frac{f_{-1}}{(x_{-1} - x_0)(x_{-1} - x_1)} \end{array} \right] = \frac{f_1 - f_{-1}}{2h}.$$

Отримуємо наближену формулу

$$f_0' \approx \frac{f_1 - f_{-1}}{2h}. \qquad (6.2)$$

Величина $\frac{f_1 - f_{-1}}{2h}$ називається ***центральною різницевою похідною***.

Нарешті, якщо розглянути другу похідну

$$l_2''(x) = 2f(x_0; x_1; x_{-1}) = 2\left[\begin{array}{l} \frac{f_0}{(x_0 - x_1)(x_0 - x_{-1})} + \frac{f_1}{(x_1 - x_0)(x_1 - x_{-1})} + \\ \frac{f_{-1}}{(x_{-1} - x_0)(x_{-1} - x_1)} \end{array} \right] = \frac{f_1 - 2f_0 + f_{-1}}{h^2},$$

отримаємо наближену формулу:

$$f_0'' \approx \frac{f_1 - 2f_0 + f_{-1}}{h^2}. \qquad (6.3)$$

Величина $\frac{f_1 - 2f_0 + f_{-1}}{h^2}$ називається ***другою різницевою похідною***.

Формули (6.1 - 6.3) називаються ***формулами чисельного диференціювання.***

Припускаючи функцію f такою, що безперервно диференціюється достатнє число разів, отримаємо похибки наближених формул (6.1 - 6.3).

Сформулюємо наступну лему.

Лема 1. Нехай $f \in C[a, b]$, $\xi_i \in [a, b]$ - довільні точки, $i = \overline{1, n}$. Тоді існує така точка $\xi \in [a, b]$, що $\frac{f(\xi_1) + f(\xi_2) + \ldots + f(\xi_n)}{n} = f(\xi).$

Похибки формул чисельного диференціювання дає наступна лема.

Лема 2. Припустимо, що $f \in C_2[x_0, x_1]$. Тоді існує така точка ξ, що

$$f_0' = \frac{f_1 - f_0}{h} - \frac{h}{2} f''(\xi), \quad x_0 < \xi < x_1. \tag{6.4}$$

Якщо $f \in C_3[x_{-1}, x_1]$, то існує така точка ξ, що

$$f' = \frac{f_1 - f_{-1}}{2h} - \frac{h^2}{6} f'''(\xi), \quad x_{-1} < \xi < x_1. \tag{6.5}$$

Коли $f \in C_4[x_{-1}, x_1]$, то існує ξ така, що

$$f_0'' = \frac{f_{-1} - 2f_0 + f_1}{h^2} - \frac{h^2}{12} f^{(4)}(\xi), \quad x_{-1} < \xi < x_1. \tag{6.6}$$

Формули (6.4-6.6) називають ***формулами чисельного диференціювання із залишковими членами***.

Похибки формул (6.1-6.3) оцінюються за допомогою наступних нерівностей, які витікають із співвідношень (6.4-6.6):

$$\begin{aligned}
&\left| f_0' - \frac{f_1 - f_0}{h} \right| \le \frac{h}{2} \max_{[x_0, x_1]} |f''(x)|, \\
&\left| f_0' - \frac{f_1 - f_{-1}}{2h} \right| \le \frac{h^2}{6} \max_{[x_{-1}, x_1]} |f'''(x)|, \\
&\left| f_0'' - \frac{f_{-1} - 2f_0 + f_1}{h^2} \right| \le \frac{h^2}{12} \max_{[x_{-1}, x_1]} |f^{(4)}(x)|.
\end{aligned} \tag{6.7}$$

Говорять, що похибка формули (6.1) має ***перший порядок відносно h***, а похибка формул (6.2) і (6.3) має ***другий порядок відносно h***. Також говорять, що формула чисельного диференціювання (6.1) ***першого порядку точності*** (відносно *h*), а формули (6.2) і (6.3) мають ***другий порядок точності***.

Вказаним способом можна отримувати формули чисельного диференціювання для більш старших похідних і для більшої кількості вузлів інтерполяції.

Вибір оптимального кроку. Допустимо, що межа абсолютної похибки при обчисленні функції f в кожній точці задовольняє нерівності

$$\Delta f_i \le \overline{\Delta}. \tag{6.8}$$

Нехай в окрузі точки x_0 похідні, через які подаються кінцеві члени в формулах (6.5) - (6.6), неперервні і задовольняють нерівностям:

$$|f'''(x)| \le \overline{M}_3, \quad |f^{(4)}(x)| \le \overline{M}_4, \tag{6.9}$$

де $\overline{M}_3, \overline{M}_4$ - деякі числа. Тоді повна похибка формул (6.2) - (6.3) (без урахування похибки заокруглень) відповідно до (6.5), (6.6), (6.8), (6.9) не перевищує відповідно величин:

$$\varepsilon_1 = \frac{\overline{\Delta} + \overline{\Delta}}{2h} + \frac{h^2}{6}\overline{M}_3, \qquad \varepsilon_2 = \frac{\overline{\Delta} + 2\overline{\Delta} + \overline{\Delta}}{h^2} + \frac{h^2}{12}\overline{M}_4. \tag{6.10}$$

Мінімізація по *h* цих величин приводить до наступних значень *h*

$$h_1 = \left(\frac{3\Delta}{\overline{M}_3}\right)^{1/3}, h_2 = 2\left(\frac{3\Delta}{\overline{M}_4}\right)^{1/4}. \tag{6.11}$$

При цьому

$$\varepsilon_1 = \frac{3}{2}\left(\frac{\overline{M}_3\Delta^{-2}}{3}\right)^{1/3}, \qquad \varepsilon_2 = 2\left(\frac{\overline{M}_4\overline{\Delta}}{3}\right)^{1/2}. \tag{6.12}$$

Якщо при вибраному значенні *h* з формул (6.2) - (6.3) відрізок $[x_{-1}, x_1]$ не виходить за межі околиці точки x_0, в якій виконується відповідна нерівність (6.9), то знайдене ***h*** є **оптимальним** і повна похибка чисельного диференціювання оцінюється відповідною величиною (6.12).

6.2 Таблиця основних інтегралів

Щоб успішно застосовувати інтегральне числення під час розв'язування задач, необхідно, насамперед, оволодіти технікою знаходження невизначених інтегралів від елементарних функцій. Одним з основних моментів успішного оволодіння технікою інтегрування елементарних функцій є досконале знання таблиці основних інтегралів. Ця таблиця складена за таблицею похідних з використанням властивості інваріантності формули інтегрування. Справедливість формул таблиці можна перевірити диференціюванням.

Нехай $u = u(x)$ – довільна функція, що на проміжку *X* має неперервну похідну $u'(x)$. Тоді на цьому проміжку справедливі такі формули:

Таблиця 6.1

Основні інтеграли

$\int 0 \cdot du = C$
$\int 1 \cdot du = \int du = u + C$
$\int u^{\alpha} du = \frac{u^{\alpha+1}}{\alpha+1} + C,\ (\alpha \neq -1)$
$\int \frac{du}{u} = \ln \lvert u \rvert + C$ або $\int \frac{du}{u} = \ln \lvert Cu \rvert,\ (u \neq 0)$
$\int a^u du = \frac{a^u}{\ln a} + C,\ (a > 0, a \neq 1)$ при $a = e \cdot \int e^u du = e^u + C$
$\int \sin u du = -\cos u + C$
$\int \cos u du = \sin u + C$
$\int \frac{du}{\cos^2 u} = tgu + C$, де $\cos u \neq 0$
$\int \frac{du}{\sin^2 u} = -ctgu + C$, де $\sin u \neq 0$
$\int \frac{du}{1+u^2} = arctgu + C$, або $\int \frac{du}{1+u^2} = -arcctgu + C$
$\int \frac{du}{\sqrt{1-u^2}} = \arcsin u + C$, або $\int \frac{du}{\sqrt{1-u^2}} = -\arccos u + C$
$\int \frac{du}{u^2+a^2} = \frac{1}{a} arctg \frac{u}{a} + C$, або $\int \frac{du}{u^2+a^2} = -\frac{1}{a} arcctg \frac{u}{a} + C \quad (a \neq 0)$
$\int \frac{du}{\sqrt{a^2-u^2}} = \arcsin \frac{u}{a} + C$, або $\int \frac{du}{\sqrt{a^2-u^2}} = -\arccos \frac{u}{a} + C$ $(a \neq 0$ в інтервалі $u \in (-a,a))$
$\int tgudu = -\ln \lvert \cos u \rvert + C,$ де $\cos u \neq 0$
$\int ctgudu = \ln \lvert \sin u \rvert + C,$ де $\sin u \neq 0$
$\int \frac{du}{\sqrt{u^2 \pm a^2}} = \ln \left\lvert u + \sqrt{u^2+a^2} \right\rvert + C$, якщо під коренем знаходиться $u^2 - a^2$, то $\lvert u \rvert > \lvert a \rvert$.

Продовження таблиці 6.1

$\int \frac{du}{u^2-a^2} = \frac{1}{2a}\ln\left\|\frac{u-a}{u+a}\right\| + C \quad (a \neq 0, \quad u \neq \pm a)$
$\int chudu = shu + C.$
$\int shudu = chu + C$

Ця таблиця має такий вигляд і у випадку, якщо $u=x$, тобто u є незалежною змінною інтегрування.

Зупинимося детальніше на деяких формулах.

За формулою маємо $\int \frac{dx}{x} = \ln|x| + C.$ Функція $f(x) = \frac{1}{x}$ визначена і неперервна $\forall x \in (-\infty,0) \cup (0,\infty)$.

Якщо $x>0$, то однією з первісних є $F(x)=\ln x$, оскільки $(\ln x)' = \frac{1}{x}$. Отже, для $x>0$ $\int \frac{du}{u} = \ln|u| + C.$

Якщо $x<0$, то однією з первісних для $f(x) = \frac{1}{x}$ є $F(x)=\ln(-x)$, оскільки $[\ln(-x)]' = \frac{1}{-x}\cdot(-1) = \frac{1}{x}$. Отже, для $x<0$ $\int \frac{dx}{x} = \ln(-x) + C.$

Об'єднуючи ці два випадки, одержуємо формулу

$$\int \frac{dx}{x} = \begin{cases} \ln x + C & \text{при} \quad x > 0, \\ \ln(-x) + C & \text{при} \quad x < 0. \end{cases}$$

або

$$\int \frac{dx}{x} = \ln|x| + C.$$

Маємо $\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C.$ Щоб переконатися у справедливості цієї формули, знайдемо похідну від правої частини

$$\left(\arcsin\frac{x}{a} + C\right)' = \frac{1}{\sqrt{1-\frac{x^2}{a^2}}}\cdot\frac{1}{a} + 0 = \frac{1}{\sqrt{\frac{a^2-x^2}{a^2}}}\cdot\frac{1}{a} = \frac{1}{\sqrt{a^2-x^2}}.$$

Маємо $\int \frac{du}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C \quad (a \neq 0), \quad (x \neq \pm a).$ Доведемо її справедливість. Для цього перетворимо підінтегральну функцію

$$\frac{1}{x^2-a^2}=\frac{1}{(x-a)(x+a)}=\frac{1}{2a}\left(\frac{1}{x-a}-\frac{1}{x+a}\right).$$

Оскільки

dx=d(x-a)=d(x+a),

маємо

$$\int\frac{dx}{x^2-a^2}=\int\frac{1}{2a}\left(\frac{1}{x-a}-\frac{1}{x+a}\right)dx=\frac{1}{2a}\left[\int\frac{d(x-a)}{x-a}-\int\frac{d(x+a)}{x+a}\right]=$$
$$=\frac{1}{2a}\left(\ln|x-a|-\ln|x+a|\right)+C=\frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right|+C.$$

Інтеграли називаються ***табличними***, за їх допомогою можна знаходити й інші інтеграли, і мета існуючих методів інтегрування полягає в тому, щоб звести шуканий інтеграл до табличного.

6.3 Основні методи інтегрування

Інтегрувати функції значно складніше, ніж диференціювати. При диференціюванні функції безпосередньо застосовуються основні формули диференціювання. При інтегруванні функцій безпосередньо застосувати основні формули можливо лише в окремих випадках.

Як правило, підінтегральну функцію доводиться перетворювати для зведення інтеграла до табличного.

Розглянемо зараз основні методи інтегрування, які спрощують зведення підінтегральної функції до такого вигляду, що дає змогу застосувати безпосереднє інтегрування, тобто обчислювати інтеграл за допомогою таблиці інтегралів і основних властивостей невизначених інтегралів.

Метод розкладання на суму

Цей метод грунтується на розкладанні підінтегральної функції в лінійну комбінацію більш простих функцій і застосування властивості лінійності інтеграла:

$$\int\sum_{i=1}^{n}a_i f_i(x)dx=\sum_{i=1}^{n}a_i\int f_i(x)dx\quad\left(\sum_{i=1}^{n}|a_i|>0\right).$$

Приклад. 1. Знайти інтеграл $I=\int\left(3\sin x-5+4x^3-\frac{1}{x}+\frac{6}{\sqrt{1+x^2}}\right)dx$.

Застосовуючи властивість лінійності невизначеного інтеграла, маємо

$$I = 3\int \sin x dx - 5\int dx + 4\int x^3 dx - \int \frac{dx}{x} + 6\int \frac{dx}{\sqrt{1-x^2}}.$$

Використовуючи формули основних інтегралів, знаходимо

$$3\int \sin x dx = 3(-\cos x + C_1) = -3\cos x + 3C_1;$$

$$-5\int dx = -5(x + C_2) = -5x - 5C_2;$$

$$4\int x^3 dx = 4\left(\frac{x^{3+1}}{3+1} + C_3\right) = x^4 + 4C_3;$$

$$-\int \frac{dx}{x} = -(\ln|x| + C_4);$$

$$6\int \frac{dx}{\sqrt{1-x^2}} = 6(\arcsin x + C_5) = 6\arcsin x + 6C_5).$$

Таким чином,

$$I = -3\cos x - 5x + x^4 - \ln x + 6\arcsin x + (3C_1 - 5C_2 + 4C_3 - C_4 + 6C_5).$$

Всі довільні сталі підсумовуємо, результат позначаємо однією літерою, тому $C = 3C_1 - 5C_2 + 4C_3 - C_4 + 6C_5$ і остаточно отримуємо

$$\int \left(3\sin x - 5 + 4x^3 - \frac{1}{x} + \frac{6}{\sqrt{1-x^2}}\right)dx = -3\cos x - 5x + x^4 - \ln|x| + C.$$

У правильності отриманого результату легко переконатись диференціюванням.

Метод підстановки або заміни змінної інтегрування

У багатьох випадках введення нової змінної інтегрування дозволяє звести знаходження шуканого інтеграла до табличного, тобто перейти до безпосереднього інтегрування. Такий метод називається методом підстановки, або заміни змінної.

Теорема (*про інтегрування за допомогою підстановки*). Нехай $F(x)$ первісна функції $f(x)$ на проміжку X, тобто

$$\int f(x)dx = F(x) + C, \qquad \forall x \in X,$$

а функція $x=\varphi(t)$ визначена і диференційована на проміжку T, множиною значень якої є проміжок X. Тоді справджується рівність

$$\int f(\varphi(t))\varphi'(t)dt = F(\varphi(t)) + C, \qquad \forall t \in T.$$

Нехай інтеграл $\int f(x)dx$ не є табличним. Тоді для його знаходження теорема застосовується одним з таких двох способів.

1. Припустимо, що від підінтегральної функції $f(x)$ можна відокремити функцію $\varphi(x) = t$ таку, що підінтегральний вираз запишеться у вигляді

$$f(x)dx = (g(x))\varphi'(x)dx = g(t)dt.$$

Тоді за теоремою маємо

$$\int f(x)dx = \int g(t)dt.$$

Якщо інтеграл у правій частині зводиться до табличного, то для нього можна записати первісну $G(t)$ або $G(\varphi(x))$ і тоді

$$\int f(x)dx = G(\varphi(x)) + C.$$

При цьому може бути зручним формалізований запис:

$$\int f(x)dx = \int g(\varphi(x))\varphi'(x)dx = \left|\begin{matrix} t = \varphi(x) \\ dt = \varphi'(x)dx \end{matrix}\right| = \int g(t)dt = G(t) + C = G(\varphi(x)) + C.$$

Або запис у формі введення функції під знак диференціала

$$\int f(x)dx = \int g(\varphi(x))\varphi'(x)dx = \int g(\varphi(x))d\varphi(x) = G(\varphi(x)) + C.$$

Тут диференційована функція $\varphi(x)$ є змінною інтегрування.

2. При знаходженні невизначеного інтеграла $\int f(x)dx$ користуються підстановкою $x=\psi(t)$, де функція $\psi(t)$ є диференційованою $\forall t \in T$, $\psi'(t)\neq 0$ $\forall t \in T$ і має обернену функцію $t=\psi^{-1}(x)$.

Таким чином, приходимо до попередньої підстановки.

При цьому формалізований запис буде таким:

$$\int f(x)dx = \left|\begin{matrix} x = \psi(t) \\ dx = \psi'(t)dt \end{matrix}\right| = \int f(\psi(t))\psi'(t)dt = \int g(t)dt = G(t) + C =$$

$$= \left|t = \psi^{-1}(x)\right| = G(\psi^{-1}(x)) + C.$$

Отже, при інтегруванні заміною змінної виконуються підстановки двох видів: $t=\varphi(x)$ і $x=\psi(t)$. Підстановки треба підбирати так, щоб одержані після перетворень нові інтеграли зі змінною інтегрування t були табличними. І після їх знаходження

від введеної змінної інтегрування *t* потрібно перейти до заданої змінної інтегрування *x*.

Приклад.

Знайти $I = \int e^{x^2} x dx$.

По-перше, можна застосувати підстановку $t = x^2$, звідки $dt = 2xdx,\ xdx = \frac{1}{2}dt$.

Підставимо в інтеграл і матимемо

$$I = \frac{1}{2}\int e^t dt = \frac{1}{2}e^t + C.$$

Повернемося до попередньої змінної x

$$I = \frac{1}{2}e^{x^2} + C.$$

По-друге, можна ввести функцію під знак диференціала, тобто записати

$$xdx = \frac{1}{2}dx^2.$$

Тоді на підставі властивості інваріантності маємо

$$I = \int e^{x^2} x dx = \frac{1}{2}\int e^{x^2} dx^2 = \frac{1}{2}e^{x^2} + C.$$

Метод інтегрування частинами

Цей метод базується на використанні формули диференціювання добутку двох функцій.

Теорема (*про формулу інтегрування частинами*). Нехай функції $u(x)$ і $v(x)$ такі, що $\forall x \in X$ існують $u'(x)$ і $v'(x)$. Крім того, функція $u'(x)v(x)$ має первісну на X, тобто існує $\int v(x)u'(x)dx$. Тоді функція $u(x)v'(x)$ також має первісну на X і справджується формула

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx,\ \forall x \in X.$$

Первісною функції $(u(x)v(x))'$ на проміжку X є функція $u(x)v(x)$. Функція $u'(x)v(x)$ має первісну за умовою теореми. Отже, і функція $u(x)v'(x)$ як різниця інтегрованих функцій має первісну. Інтегруючи обидві частини цієї тотожності, дістаємо потрібну формулу. Оскільки $v'(x)dx = dv,\ u'(x)dx = du$, то її можна переписати у вигляді

$$\int \mathrm{udv} = \mathrm{uv} - \int \mathrm{vdu}\,.$$

Ця формула і називається формулою інтегрування частинами невизначеного інтеграла.

Назва інтегрування частинами пояснюється тим, що формула не дає остаточного результату, а лише зводить задачу відшукання інтеграла $\int \mathrm{udv}$ до задачі відшукання іншого інтеграла $\int \mathrm{vdu}$, яка при вдалому виборі *u* і *dv* має виявитись простішою.

Якщо *u* і *dv* вибрані невдало, то замість спрощення задача ускладнюється. Для знаходження функції *v* за диференціалом *dv* можна брати будь-яку довільну сталу, оскільки в остаточний результат вона не входить. Справді

$$\int udv = u(v+c) - \int (v+c)du = uv + cu - \int vdu - c\int du =$$
$$= uv - \int vdu + cu - cu = uv - \int vdu.$$

Щоб не проводити зайвих обчислень, можна завжди покласти *c*=0.

У деяких випадках формула інтегрування частинами є тільки допоміжною при відшуканні інтеграла, вона приводить до алгебраїчного рівняння відносно шуканого інтеграла.

Формулу інтегрування частинами інколи доводиться застосовувати декілька разів.

Приклад.

$\mathrm{I} = \int (\mathrm{x}+1)\sin \mathrm{xdx}\,.$

Введемо позначення *u* =*x*+1, *dv* =*sinxdx.* Тоді $\mathrm{I} = \mathrm{uv} - \int \mathrm{vdu}$. Знайдемо *u* і *v,* які містяться в правій частині. З рівності *dv=sinxdx* знаходимо: $\mathrm{v} = \int \sin \mathrm{xdx} = -\cos \mathrm{x}$, а *du=dx.*

Отже,

$\mathrm{I} = -(\mathrm{x}+1)\cos \mathrm{x} - \int (-\cos \mathrm{x})\mathrm{dx} = -\mathrm{x}\cos \mathrm{x} + \sin \mathrm{x} + \mathrm{C}\,.$

Нехай тепер *u=sinx, du=xdx.* Звідси $\mathrm{v} = \int \mathrm{xdx} = \frac{\mathrm{x}^2}{2}$, $\mathrm{du} = \cos \mathrm{xdx}$. Отже,

$$\mathrm{I} = \frac{1}{2}\mathrm{x}^2 \sin \mathrm{x} - \frac{1}{2}\int \mathrm{x}^2 \cos \mathrm{xdx}\,.$$

У правій частині дістали невизначений інтеграл, який є складнішим, ніж заданий.

Можна вказати на деякі типи інтегралів, які зручно інтегрувати частинами.

1. Інтеграли виду $\int P(x)e^{ax}dx$, $\int P(x)\sin bxdx$, $\int P(x)\cos bxdx$, де $P(x)$ – многочлен n-го степеня від x, $a\neq 0$, $b\neq 0$ – дійсні числа. У цих інтегралах за u слід взяти множник $P(x)$. Після застосування n разів формули інтегрування частинами ці інтеграли зводяться відповідно до інтегралів $\int e^{ax}dx$, $\int \sin bxdx$, $\int \cos bxdx$ як у прикладі.

2. Інтеграли виду $\int P(x)\ln xdx$, $\int P(x)\arcsin x$, $\int P(x)\arccos xdx$, $\int P(x)arctgxdx$, $\int P(x)arctgxdx$, де $P(x)$ – многочлен n-го степеня від x.

Тут за u слід брати множники *lnx, arcsinx, arccosx, arctgx, arcctgx*.

3. Інтеграли виду $\int e^{ax}\sin bxdx$, $\int e^{ax}\cos bxdx$, $a\neq 0$, $b\neq 0$ – дійсні числа.

Лекція 7

Власні значення та власні вектори матриці

7.1 Знаходження власних векторів і власних значень матриць

Якщо A — квадратна матриця n-го порядку і $Ax=\lambda x$ при $x\neq 0$, то число λ називається ***власним значенням*** матриці, а ненульовий вектор x — відповідним йому ***власним вектором***. Перепишемо задачу в такому вигляді

$$(A-\lambda E)x=0, \qquad x\neq 0. \tag{7.1}$$

Для існування нетривіального розв'язку задачі (7.1) має виконуватися умова

$$\det(A-\lambda E)=0. \tag{7.2}$$

Цей визначник являє собою многочлен n-ї степені від λ. Його називають ***характеристичним многочленом***. Існує n власних значень - коренів цього многочлена, серед яких можуть бути однакові (кратні).

Якщо знайдено деяке власне значення, то, при підстановці його в однорідну систему (7.1), можна визначити відповідний власний вектор. Будемо нормувати власні вектори. Нормуванням (на одиницю) вектора x називають множення його

на $\|x\|^{-1}$. Нормований вектор має одиничну довжину. Тоді кожному простому (не кратному) власному значенню відповідає один (з точністю до напрямку) власний вектор, а сукупність всіх власних векторів, що відповідають 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. Наприклад, розглянемо такі матриці четвертого порядку:

$$A=\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{bmatrix}, \quad C_4=\begin{bmatrix} a & 1 & 0 & 0 \\ 0 & a & 1 & 0 \\ 0 & 0 & a & 1 \\ 0 & 0 & 0 & a \end{bmatrix}, \quad B=\left[\begin{array}{c|ccc} a & 0 & 0 & 0 \\ \hline 0 & a & 1 & 0 \\ 0 & 0 & a & 1 \\ 0 & 0 & 0 & a \end{array}\right] \tag{7.3}$$

В кожній з них характеристичне рівняння приймає вигляд $\det(A-\lambda E)=(a-\lambda)^4=0$, а отже, власне значення $\lambda=a$ і має кратність p=4. Проте в першій матриці є чотири лінійно-незалежних власних вектора

$$e_1=\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad e_2=\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_3=\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad e_4=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \tag{7.4}$$

У другій матриці є тільки один власний вектор e_1. Другу матрицю називають ***простою жордановою*** (або класичною) ***підматрицею***.

Третя матриця має так звану ***канонічну жорданову форму*** (по діагоналі стоять або числа, або жорданові підматриці, а інші елементи дорівнюють нулю).

Таким чином, якщо серед власних значень матриці є кратні, то її власні вектори не завжди утворюють базис. Однак і в цьому випадку власні вектори, що відповідають різним власним значенням, являються лінійно-незалежними.

При розв’язуванні теоретичних і практичних задач часто виникає потреба визначити власні значення даної матриці *A*, тобто обчислити корені її характеристичного рівняння (7.2), а також знайти відповідні власні векторі матриці *A*. Друга задача є простішою, оскільки якщо корені характеристичного рівняння

відомі, то знаходження власних векторів зводиться до відшукання ненульових розв'язків деяких однорідних лінійних систем. Тому, в першу чергу, будемо займатися першою задачею - відшуканням коренів характеристичного рівняння (7.2).

Тут в основному використовують два прийоми:

1) розгортання характеристичного визначника в поліном n-го степеня

$$\mathrm{D}(\lambda) = \det(\mathrm{A} - \lambda \mathrm{E})$$

з подальшим розв'язком рівняння $\mathrm{D}(\lambda) = 0$ одним з відомих наближених способів.

Розгортання характеристичного визначника.

Як відомо, характеристичним визначником матриці $A = [\mathrm{a}_{ij}]$ називається визначник вигляду

$$\mathrm{D}(\lambda) = \det(\mathrm{A} - \lambda \mathrm{E}) = \begin{vmatrix} \mathrm{a}_{11} - \lambda & \mathrm{a}_{12} & \dots & \mathrm{a}_{1n} \\ \mathrm{a}_{21} & \mathrm{a}_{22} - \lambda & \dots & \mathrm{a}_{2n} \\ \dots & \dots & \dots & \dots \\ \mathrm{a}_{n1} & \mathrm{a}_{n2} & \dots & \mathrm{a}_{nn} - \lambda \end{vmatrix}.$$

Прирівнюючи цей визначник до нуля, одержуємо характеристичне рівняння

$$\mathrm{D}(\lambda) = 0.$$

Якщо потрібно знайти всі корені характеристичного рівняння, то доцільно заздалегідь обчислити визначник.

Розгортаючи визначник, одержуємо поліном n-го степеня

$$\mathrm{D}(\lambda) = (-1)^{\mathrm{n}}[\lambda^{\mathrm{n}} - \sigma_1 \lambda^{\mathrm{n}-1} + \sigma_2 \lambda^{\mathrm{n}-2} - \ldots + (-1)^{\mathrm{n}} \sigma_{\mathrm{n}}], \tag{7.5}$$

де $\sigma_1 = \sum_{\alpha=1}^{\mathrm{n}} \mathrm{a}_{\alpha\alpha}$ - є сума усіх діагональних мінорів першого порядку матриці A;

$\sigma_2 = \sum_{\alpha<\beta}^{\mathrm{n}} \begin{vmatrix} \mathrm{a}_{\alpha\alpha} & \mathrm{a}_{\alpha\beta} \\ \mathrm{a}_{\beta\alpha} & \mathrm{a}_{\beta\beta} \end{vmatrix}$ - є сума всього діагонального мінору другого порядку матриці A;

$\sigma_3 = \sum_{\alpha<\beta<\gamma}^{\mathrm{n}} \begin{vmatrix} \mathrm{a}_{\alpha\alpha} & \mathrm{a}_{\alpha\beta} & \mathrm{a}_{\alpha\gamma} \\ \mathrm{a}_{\beta\alpha} & \mathrm{a}_{\beta\beta} & \mathrm{a}_{\beta\gamma} \\ \mathrm{a}_{\gamma\alpha} & \mathrm{a}_{\gamma\beta} & \mathrm{a}_{\gamma\gamma} \end{vmatrix}$ - сума всіх діагональних мінорів третього порядку матриці A і т.д. Нарешті

$$\sigma_{\mathrm{n}} = \det \mathrm{A}.$$

Легко переконатися, що число діагональних мінорів k-го порядку матриці A дорівнює

$$C_n^k = \frac{n(n-1)...(n-k+1)}{k!},\ k = 1,2,...,n.$$

Звідси одержуємо, що безпосереднє обчислення коефіцієнтів характеристичного полінома еквівалентно обчисленню

$$C_n^1 + C_n^2 + ... + C_n^n = 2^n - 1$$

визначників різних порядків. Остання задача технічно важко здійснена для великих значень n. Тому створені спеціальні методи розгортання характеристичних визначників (методи А. Н. Крилова, А. М. Данилевського, Левер'є, метод невизначених коефіцієнтів, метод інтерполяції та ін.).

7.2 Метод розгортання А. М. Данилевського

Суть методу А. М. Данилевського полягає в приведенні характеристичного визначника до так званого нормального виду Фробеніуса

$$D(\lambda) = \begin{vmatrix} p_1 - \lambda & p_2 & p_3 & \dots & p_1 \\ 1 & -\lambda & 0 & \dots & 0 \\ 0 & 1 & -\lambda & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\lambda \end{vmatrix} \qquad (7.6)$$

Якщо нам вдалося записати визначника у формі (7.6), то, розкладаючи його по елементах першого рядка, матимемо:

$$D(\lambda) = (p_1 - \lambda)(-\lambda)^{n-1} - p_2(-\lambda)^{n-2} + p_3(-\lambda)^{n-3} - ... + (-1)^{n-1} p_n$$

або

$$D(\lambda) = (-1)^n (\lambda^n - p_1\lambda^{n-1} - p_2\lambda^{n-2} - p_3\lambda^{n-3} - ... - p_n). \qquad (7.7)$$

Таким чином, розгортання характеристичного визначника, записаного в нормальній формі (7.6), не являє труднощів. Позначимо через

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$ - вихідну матрицю,

а через $P = \begin{bmatrix} p_1 & p_2 & \dots & p_{n-1} & p_n \\ 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$ - подібну їй матрицю Фробеніуса, тобто

$$P = S^{-1}AS,$$

де S - особлива матриця.

Оскільки подібні матриці володіють однаковими характеристичними поліномами, то маємо:

$$\det(A - \lambda E) = \det(P - \lambda E). \qquad (7.8)$$

Тому, для обґрунтування методу, досить показати яким чином, виходячи з матриці A, будується матриця P. Згідно методу Данилевського, перехід від матриці A до подібної їй матриці P здійснюється за допомогою $(n-1)$ перетворення подібності, що послідовно перетворюють рядки матриці A, починаючи з останньої, у відповідні рядки матриці P.

Покажемо початок процесу. Нам необхідно рядок

$$a_{n1}\ a_{n2}\ \dots\ a_{n,n-1}\ a_{nn}$$

перевести в рядок 0 0 ... 1 0. Припускаючи, що $a_{n,n-1} \neq 0$, розділимо всі елементи $(n-1)$- го стовпця матриці A на $a_{n,n-1}$. Тоді її n-й рядок прийме вигляд

$$a_{n1}\ a_{n2}\ \dots\ 1\ a_{nn}.$$

Потім віднімемо $(n-1)$-й стовпець перетвореної матриці, помножений відповідно на числа $a_{n1}, a_{n2}, \dots, a_{nn}$, зі всієї решти її стовпців.

В результаті одержимо матрицю, останній рядок якої має бажаний вигляд 0 0 ... 1 0. Вказані операції є елементарними перетвореннями, що здійснюються над стовпцями матриці A. Виконавши ці ж перетворення над одиничною матрицею, одержимо матрицю

$$M_{n-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ m_{n-1,1} & m_{n-1,2} & \dots & m_{n-1,n-1} & m_{n-1,n} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

де

$$m_{n-1,i} = -\frac{a_{ni}}{a_{n,n-1}}, \quad i \neq n-1 \tag{7.9}$$

і

$$m_{n-1,n-1} = \frac{1}{a_{n,n-1}}. \tag{7.10}$$

Звідси робимо висновок, що проведені операції рівносильні множенню справа матриці M_{n-1} на матрицю *A*, тобто після вказаних перетворень одержимо матрицю

$$AM_{n-1} = B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1,n-1} & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2,n-1} & b_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ b_{n-1,1} & b_{n-1,2} & \dots & b_{n-1,n-1} & b_{n-1,n} \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Використовуючи правило множення матриць, знаходимо, що елементи матриці *B* обчислюються за наступними формулами:

$$b_{ij} = a_{ij} + a_{i,n-1}m_{n-1,j} \qquad j \neq n-1$$
$$b_{i,n-1} = a_{i,n-1}m_{n-1,n-1}.$$

Проте побудована матриця $B = AM_{n-1}$ не буде подібна матриці *A*. Для того, щоб мати перетворення подібності, потрібно обернену матрицю $(M_{n-1})^{-1}$ зліва помножити на матрицю *B*:

$$(M_{n-1})^{-1}AM_{n-1} = (M_{n-1})^{-1}B.$$

Безпосередньою перевіркою легко переконатися, що обернена матриця $(M_{n-1})^{-1}$ має вигляд

$$(M_{n-1})^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ a_{n1} & a_{n2} & \dots & a_{n,n-1} & a_{nn} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Нехай

$$(M_{n-1})^{-1}AM_{n-1} = C.$$

Отже

$$C = (M_{n-1})^{-1}B.$$

Оскільки множення зліва матриці $(M_{n-1})^{-1}$ на матрицю B не змінює перетвореного рядка останньої, то матриця C має вигляд

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1,n-1} & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2,n-1} & c_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ c_{n-1,1} & c_{n-1,2} & \dots & c_{n-1,n-1} & c_{n-1,n} \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Таким чином, множення $(M_{n-1})^{-1}$ на матрицю B змінює лише (n-1)-й рядок матриці B. Одержана матриця C подібна матриці A і має один зведений рядок. Цим закінчується перший етап процесу.

Далі, якщо $c_{n-1,n-2} \neq 0$, то над матрицею C можна повторити аналогічні операції, узявши за основу (n-2)-й її рядок. В результаті одержимо матрицю

$$D = (M_{n-2})^{-1} C M_{n-2}$$

з двома зведеними рядками. Над останньою матрицею проробляємо ті ж операції. Продовжуючи цей процес одержимо матрицю Фробеніуса

$$P = (M_1)^{-1} \dots (M_{n-2})^{-1} (M_{n-1})^{-1} A M_{n-1} M_{n-2} \dots M_1,$$

якщо всі n-1 проміжних перетворень можливі. Весь процес може бути оформлений в зручну обчислювальну схему, складання якої покажемо на наступному прикладі.

Приклад. Привести до вигляду Фробеніуса матрицю

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}.$$

Розв’язання.

Обчислення розташовуємо в таблицю.

Номер	M^{-1}	Рядки матриці				Σ	Σ’
рядка		1	2	3	4		
1		1	2	3	4	10	
2		2	1	2	3	8	
3		3	2	1	2	8	
4		4	3	2	1	10	
I	M_3^{-1} M_3	–2	–1,5	0,5–1	–0,5	–5	
5	4	–5	–2,5	1,5	2,5	–3,5	–5
6	3	2	–2	1	2	–1	–2
7	2	1	0,5	0,5	1,5	3,5	3
8	1	0	0	1	0	1	0
7’		–24	–15	11	19	–9	
II	M_2^{-1} M_2	–1,600	–0,067	0,733	1,267	–0,600	
			–1				
9	–24	–1	0,167	–0,333	–0,667	–1,833	–2
10	–15	1,2	0,133	–0,467	–0,533	0,333	0,2
11	11	0	1	0	0	1	0
12	19	0	0	1	0	1	1
10’		6	5	34	24	69	
III	M_1^{-1} M_1	0,167–1	–0,833	–5,667	–4,000	–	
						11,500	
13	6	–0,167	1	5,333	3,333	9,500	9,667
14	5	1	0	0	0	1	0
15	34	0	1	0	0	1	1
16	24	0	0	1	0	1	1
13’		4	40	56	20	120	

У рядках 1-4 таблиці розміщуємо елементи $a_{ij}(i,\ j=1,2,3,4)$ даної матриці і контрольні суми $a_{i5}=\sum_{j=1}^{4}a_{ij}\ (i=1,2,3,4)\ (\sum)$. Відзначаємо елемент $a_{43}=2$, що належить третьому стовпцю (відмічений стовпець). У рядку 1 записуємо елементи третього рядка матриці $M_{n-1}=M_3$:

$$m_{31}=-\frac{a_{41}}{a_{42}}=-\frac{4}{2}=-2;$$

$$m_{32} = -\frac{a_{42}}{a_{43}} = -\frac{3}{2} = -1{,}5;$$

$$m_{33} = \frac{1}{a_{43}} = \frac{1}{2} = 0{,}5;$$

$$m_{34} = -\frac{a_{44}}{a_{43}} = -\frac{1}{2} = -0{,}5.$$

Сюди ж (рядок 1 таблиці) поміщаємо елемент

$$m_{35} = -\frac{a_{45}}{a_{43}} = -\frac{10}{2} = -5,$$

що одержується аналогічним прийомом з контрольного стовпця Σ. Число -5 повинно співпасти з сумою елементів рядка I, що не входять в контрольний стовпець (після заміни елементу m_{33} на -1). Для зручності число -1 записуємо поряд з елементом m_{33}, відокремлюючи від останнього межею.

У рядках 5-8 в графі M^{-1} виписуємо третій рядок матриці M^{-1}, яка співпадає з четвертим рядком початкової матриці А. У рядках 5-8 у відповідних стовпцях виписуємо елементи матриці

$$B = AM_3,$$

що обчислюються за двочленними формулами для невідмічених стовпців і по одночленній формулі для відміченого стовпця. Наприклад, для першого стовпця маємо:

$$b_{11} = 1 + 3(-2) = -5;$$
$$b_{21} = 2 + 2(-2) = -2;$$
$$b_{31} = 3 + 1(-2) = 1;$$
$$b_{41} = 4 + 2(-2) = 0$$

і т.д.

Перетворені елементи третього (відміченого) стовпця отримуються за допомогою множення початкових елементів на $m_{33} = 0{,}5$. Наприклад,

$$b_{13} = 3 \cdot 0{,}5 = 1{,}5;$$
$$b_{23} = 2 \cdot 0{,}5 = 1;$$
$$b_{33} = 1 \cdot 0{,}5 = 0{,}5;$$
$$b_{43} = 2 \cdot 0{,}5 = 1;$$

Відмітимо, що останній рядок матриці B повинен мати вигляд

0 0 1 0.

Для контролю поповнюємо матрицю B перетвореними по аналогічних двочленних формулах з $m_{35} = -5$ відповідними елементами стовпця Σ. Наприклад,

$$b_{16} = 10 + 3 \cdot (-5) = -5;$$

$$b_{26} = 8 + 2 \cdot (-5) = -2;$$

$$b_{36} = 8 + 1 \cdot (-5) = 3;$$

$$b_{46} = 10 + 2 \cdot (-5) = 0.$$

Отримані результати записуємо в стовпці Σ' у відповідних рядках. Додавши до них елементи третього стовпця, одержимо контрольні суми

$$b_{i5} = \sum_{j=1}^{4} b_{ij} \quad (i = 1, 2, 3, 4)$$

для рядків 5-8 (стовпець Σ).

Перетворення M_3^{-1},що проведене над матрицею і, що дає матрицю $C = M_3^{-1}B$, змінює лише третій рядок матриці B, тобто сьомий рядок таблиці. Елементи цього перетвореного рядка 7' є сумами парних добутків елементів стовпця M^{-1}, що знаходяться в рядках 5-8, на відповідні елементи кожного із стовпців матриці B. Наприклад

$$c_{31} = 4(-5) + 3(-2) + 2 \cdot 1 = -24$$

і т. д.

Такі ж перетворення проводимо над стовпцем Σ:

$$c_{35} = 4(-3{,}5) + 3(-1) + 2 \cdot 3{,}5 + 1 \cdot 1 = -9.$$

В результаті одержуємо матрицю C, що складається з рядків 5, 6, 7', 8 з контрольними сумами Σ, причому матриця C подібна матриці A і має один зведений рядок 8. Цим закінчується побудова першого подібного перетворення $C = M_3^{-1}AM_3$.

Далі, прийнявши матрицю C за вихідну і виділивши елемент $c_{32} = -15$(другий стовпець), продовжуємо процес аналогічним чином. В результаті одержуємо матрицю $D = M_2^{-1}CM_2$, елементи якої розташовані в рядках 9, 10', 11,

12, що містить два зведені рядки. Нарешті, відправляючись від елементу $d_{21}=6$ (перший стовпець) і перетворюючи матрицю D в подібну їй, одержуємо шукану матрицю Фробеніуса P, елементи якої записані в рядках 13', 14, 15, 16. На кожному етапі процесу контроль здійснюється за допомогою стовпців Σ і Σ'.

Таким чином, матриця Фробеніуса буде мати вигляд:

$$P=\begin{bmatrix} 4 & 40 & 56 & 20 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Звідси характеристичний визначник, приведений до нормального виду Фробеніуса, запишеться так:

$$D(\lambda)=\begin{vmatrix} 4-\lambda & 40 & 56 & 20 \\ 1 & -\lambda & 0 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 1 & -\lambda \end{vmatrix}$$

або

$$D(\lambda)=\lambda^4-4\lambda^3-40\lambda^2-56\lambda-20.$$

Процес Данилевського відбувається без жодних ускладнень, якщо всі елементи, що виділяються, відмінні від нуля.

Припустимо, що при перетворенні матриці A в матрицю Фробеніуса P після декількох кроків пришли до матриці вигляду

$$D=\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1k} & \dots & d_{1,n-1} & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2k} & \dots & d_{2,n-1} & d_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ d_{k1} & d_{k2} & \dots & d_{kk} & \dots & d_{k,n-1} & d_{kn} \\ 0 & 0 & \dots & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 1 & 0 \end{bmatrix},$$

причому виявилось, що $d_{k,k-1}=0$.

Тоді продовжувати перетворення по методу Данилевського не можна. Тут можливі два випадки.

1. Нехай якийсь елемент матриці D, що стоїть ліворуч нульового елемента $d_{k,\,k-1}$, відмінний від нуля, тобто $d_{k,\,l} \neq 0$, де $l > k-1$. Тоді цей елемент висуваємо на місце нульового елементу $d_{k,\,k-1}$, тобто переставляємо (k-1)-й і k-й стовпці матриці D і одночасно переставляємо її (k-1)-й і l-й рядки. Можна довести, що одержана нова матриця D' буде подібна колишній. До нової матриці застосовуємо метод Данилевського.

2. Нехай $d_{k,l} = 0\ (l = 1, 2, \ldots, k-1)$, тоді D має вигляд

$$\left[\begin{array}{cccc|ccc|c} & & (D_1) & & & (L) & & \\ c_{11} & c_{12} & \cdots & c_{1,\,k-1} & c_{1k} & \cdots & c_{1,\,n-1} & c_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{k-1,1} & c_{k-1,2} & \cdots & c_{k-1,\,k-1} & c_{k-1,\,k} & \cdots & c_{k-1,\,n-1} & c_{k-1,\,n} \\ \hline 0 & 0 & \cdots & 0 & c_{kk} & \cdots & c_{k,\,n-1} & c_{kn} \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \\ & & (0) & & & (D_2) & & \end{array}\right] = \left[\begin{array}{c|c} D_1 & L \\ \hline 0 & D_2 \end{array}\right].$$

У такому разі визначник $\det(\mathrm{D} - \lambda\mathrm{E})$ розпадається на два визначники

$$\det(\mathrm{D} - \lambda\mathrm{E}) = \det(\mathrm{D}_1 - \lambda\mathrm{E})\det(\mathrm{D}_2 - \lambda\mathrm{E}).$$

При цьому матриця D_2 вже приведена до канонічної форми Фробеніуса і тому $\det(\mathrm{D}_2 - \lambda\mathrm{E})$ обчислюється відразу. Залишається застосувати метод Данилевського до матриці D_1.

7.3 Обчислення власних векторів по методу Данилевського

Метод Данилевського дає можливість визначати власні вектори даної матриці A, якщо відомі її власні значення. Нехай λ — власне значення матриці A, а отже, і власне значення подібної їй матриці Фробеніуса P.

Знайдемо власний вектор $\mathrm{y} = (\mathrm{y}_1, \mathrm{y}_2, ..., \mathrm{y}_\mathrm{n})$ матриці P, відповідний даному значенню λ: $\mathrm{Py} = \lambda\mathrm{y}$. Звідси $(\mathrm{P} - \lambda\mathrm{E})\mathrm{y} = 0$ або

$$\begin{bmatrix} p_1-\lambda & p_2 & p_3 & \dots & p_n \\ 1 & -\lambda & 0 & \dots & 0 \\ 0 & 1 & -\lambda & \dots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \dots & -\lambda \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{bmatrix} = 0.$$

Перемножуючи матриці, одержимо систему для визначення координат $y_1, y_2, \dots, y_n$ власного вектора y:

$$\left.\begin{aligned} (p_1-\lambda)y_1 + p_2y_2 + \dots + p_ny_n &= 0, \\ y_1 - \lambda y_2 &= 0, \\ y_2 - \lambda y_3 &= 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ y_{n-1} - \lambda y_n &= 0. \end{aligned}\right\} \qquad (7.11)$$

Система (7.11) - однорідна. З точністю до коефіцієнта пропорційності розв'язки її можуть бути знайдені таким чином. Покладемо $y_n=1$. Тоді послідовно одержимо:

$$\left.\begin{aligned} y_{n-1} &= \lambda, \\ y_{n-2} &= \lambda^2, \\ \dots\dots\dots & \\ y_1 &= \lambda^{n-1}. \end{aligned}\right\}$$

Таким чином, шуканий власний вектор є

$$y = \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ . \\ . \\ . \\ 1 \end{bmatrix}.$$

Позначимо тепер через *x* власний вектор матриці *A*, що відповідає значенню λ. Тоді маємо:

$$x = M_{n-1}M_{n-2}\dots M_2M_1y.$$

Перетворення M_1, здійснене над *y*, дає:

$$M_1 y = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} m_{1k} y_k \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} m_{1k} y_k \\ \lambda^{n-2} \\ \vdots \\ 1 \end{bmatrix}.$$

Таким чином, перетворення M_1 змінює лише першу координату вектора. Аналогічно перетворення M_2 змінить лише другу координату вектора $M_1 y$ і т.д. Повторивши цей процес n-1 разів, одержимо шуканий власний вектор *x* матриці *A*.

7.4 Метод розгортання А. Н. Крилова

Приведемо метод розгортання характеристичного визначника, що належить А. Н. Крилову і заснований на істотно іншій ідеї, ніж метод А. М. Данилевського.

Нехай

$$D(\lambda) \equiv \det(\lambda E - A) = \lambda^n + p_1 \lambda^{n-1} + \dots p_n$$

- характеристичний поліном (з точністю до знаку) матриці *A*. Згідно тотожності Гамільтона-Келі, матриця *A* обертає в нуль свій характеристичний поліном; тому

$$A^n + p_1 A^{n-1} + \dots + p_n E = 0. \qquad (7.12)$$

Візьмемо тепер довільний ненульовий вектор

$$y^{(0)} = \begin{bmatrix} y_1^{(0)} \\ \vdots \\ y_n^{(0)} \end{bmatrix}.$$

Перемноживши обидві частини рівності (7.12) справа на $y^{(0)}$, одержимо:

$$A^n y^{(0)} + p_1 A^{n-1} y^{(0)} + \dots + p_n y^{(0)} = 0. \qquad (7.13)$$

Покладемо:

$$A^k y^{(0)} = y^{(k)} \ (k = 1, 2, \dots, n). \qquad (7.14)$$

Тоді рівність (7.13) набуває вигляду

$$y^{(n)} + p_1 y^{(n-1)} + \dots + p_n y^{(0)} = 0 \qquad (7.15)$$

або

$$\begin{bmatrix} y_1^{(n-1)} & y_1^{(n-2)} & \vdots & y_1^{(0)} \\ y_2^{(n-1)} & y_2^{(n-2)} & \vdots & y_2^{(0)} \\ \vdots & \vdots & \vdots & \\ y_n^{(n-1)} & y_n^{(n-2)} & \vdots & y_n^{(0)} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = -\begin{bmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \vdots \\ y_{1n}^{(n)} \end{bmatrix}. \qquad (7.16)$$

де

$$y^{(k)} = \begin{bmatrix} y_1^{(k)} \\ y_2^{(k)} \\ \vdots \\ y_{1n}^{(k)} \end{bmatrix} \quad (k = 0, 1, 2, \ldots, n).$$

Отже, векторна рівність (7.15) еквівалентна системі рівнянь

$$p_1 y_j^{(n-1)} + p_2 y_j^{(n-2)} + \ldots + p_n y_j^{(0)} = -y_j^{(n)} \qquad (j = 1, 2, \ldots, n), \quad (7.17)$$

з якої, можна визначити невідомі коефіцієнти $p_1, p_2, \ldots, p_n$.

Оскільки на підставі формули (7.14)

$$y^{(k)} = A y^{(k-1)} \ (k = 1, 2, \ldots, n),$$

то координати $y_1^{(k)}, y_2^{(k)}, \ldots, y_n^{(k)}$ вектора $y^{(k)}$ послідовно обчислюються за формулами

$$\left.\begin{aligned} y_i^{(1)} &= \sum_{i=1}^{n} a_{ij} y_i^{(0)}, \\ y_i^{(2)} &= \sum_{i=1}^{n} a_{ij} y_i^{(1)}, \\ &\ldots\ldots\ldots\ldots\ldots \\ y_i^{(n)} &= \sum_{i=1}^{n} a_{ij} y_i^{(n-1)} \quad (i = 1, 2, \ldots, n). \end{aligned}\right\} \qquad (7.18)$$

Таким чином, визначення коефіцієнтів p_j характеристичного полінома методом Крилова зводиться до розв'язання лінійної системи рівнянь (7.17), коефіцієнти якої обчислюються за формулами (7.18), причому координати початкового вектора

$$y^{(0)} = \begin{bmatrix} y_1^{(0)} \\ y_2^{(0)} \\ \vdots \\ y_{1n}^{(0)} \end{bmatrix}$$

довільні. Якщо система (7.17) має єдиний розв'язок, то її корені $p_1, p_2, \ldots, p_n$ є коефіцієнтами характеристичного полінома. Цей розв'язок може бути знайдено, наприклад, методом Гауса. Якщо система (7.17) не має єдиного розв'язку, то завдання ускладнюється. В цьому випадку рекомендується змінити початковий вектор.

Приклад. Методом А. Н. Крилова знайти характеристичний поліном матриці

$$A=\begin{bmatrix}1 & 2 & 3 & 4\\ 2 & 1 & 2 & 3\\ 3 & 2 & 1 & 2\\ 4 & 3 & 2 & 1\end{bmatrix}.$$

Розв'язання. Виберемо початковий вектор

$$y^{(0)}=\begin{bmatrix}1\\ 0\\ 0\\ 0\end{bmatrix}.$$

Користуючись формулами визначимо координати векторів

$y^{(k)}=A^k y^{(0)}\ (k=1,2,3,4)$.

Маємо:

$$y^{(1)}=Ay^{(0)}=\begin{bmatrix}1 & 2 & 3 & 4\\ 2 & 1 & 2 & 3\\ 3 & 2 & 1 & 2\\ 4 & 3 & 2 & 1\end{bmatrix}\begin{bmatrix}1\\ 0\\ 0\\ 0\end{bmatrix}=\begin{bmatrix}1\\ 2\\ 3\\ 4\end{bmatrix};$$

$$y^{(2)}=Ay^{(1)}=\begin{bmatrix}1 & 2 & 3 & 4\\ 2 & 1 & 2 & 3\\ 3 & 2 & 1 & 2\\ 4 & 3 & 2 & 1\end{bmatrix}\begin{bmatrix}1\\ 2\\ 3\\ 4\end{bmatrix}=\begin{bmatrix}30\\ 22\\ 18\\ 20\end{bmatrix};$$

$$y^{(3)}=Ay^{(2)}=\begin{bmatrix}1 & 2 & 3 & 4\\ 2 & 1 & 2 & 3\\ 3 & 2 & 1 & 2\\ 4 & 3 & 2 & 1\end{bmatrix}\begin{bmatrix}30\\ 22\\ 18\\ 20\end{bmatrix}=\begin{bmatrix}208\\ 178\\ 192\\ 242\end{bmatrix};$$

$$y^{(4)} = Ay^{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 208 \\ 178 \\ 192 \\ 242 \end{bmatrix} = \begin{bmatrix} 2108 \\ 1704 \\ 1656 \\ 1992 \end{bmatrix}.$$

Складемо систему:

$$\begin{bmatrix} y_1^{(3)} & y_1^{(2)} & y_1^{(1)} & y_1^{(0)} \\ y_2^{(3)} & y_2^{(2)} & y_2^{(1)} & y_2^{(0)} \\ y_3^{(3)} & y_3^{(2)} & y_3^{(1)} & y_3^{(0)} \\ y_4^{(3)} & y_4^{(2)} & y_4^{(1)} & y_4^{(0)} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = -\begin{bmatrix} y_1^{(4)} \\ y_2^{(4)} \\ y_3^{(4)} \\ y_4^{(4)} \end{bmatrix},$$ яка в нашому випадку має вигляд

$$\begin{bmatrix} 208 & 30 & 1 & 1 \\ 178 & 22 & 2 & 0 \\ 192 & 18 & 3 & 0 \\ 242 & 20 & 4 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = -\begin{bmatrix} 2108 \\ 1704 \\ 1656 \\ 1992 \end{bmatrix}.$$

Звідси

$$\left.\begin{aligned} 208p_1 + 30p_2 + p_3 + p_4 &= -2108, \\ 178p_1 + 22p_2 + 2p_3 &= -1704, \\ 192p_1 + 18p_2 + 3p_3 &= -1656, \\ 242p_1 + 20p_2 + 4p_3 &= -1991. \end{aligned}\right\}$$

Розв’язавши цю систему, одержимо:

$p_1 = -4;\ p_2 = -40;\ p_3 = -56;\ p_4 = -20$.

Отже

$\det(\lambda E - A) = \lambda^4 - 4\lambda^3 - 40\lambda^2 - 56\lambda - 20$, що співпадає з результатом, знайденим по методу Данилевського.

7.5 Обчислення власних векторів по методу Крилова

Метод А. Н. Крилова дає можливість просто знайти відповідні власні вектори.

Для простоти обмежимося випадком, коли характеристичний поліном

$$D(\lambda) \equiv \det(\lambda E - A) = \lambda^n + p_1\lambda^{n-1} + \ldots p_n$$

має різні корені $\lambda_1, \lambda_2, \ldots, \lambda_n$. Припустимо, що коефіцієнти полінома і його корені визначені. Потрібно знайти власні вектори $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$, що відповідають відповідно власним значенням $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Нехай $y^{(0)}, y^{(1)} = Ay^{(0)}, \ldots, y^{(n-1)} = A^{n-1}y^{(0)}$ - вектори, використані в методі Крилова для знаходження коефіцієнтів $p_i (i = 1, 2, \ldots, n)$. Розкладаючи вектор $y^{(0)}$ по власних векторах $x^{(i)} (i = 1, 2, \ldots, n)$, матимемо:

$$y^{(0)} = c_1 x^{(1)} + c_2 x^{(2)} + \ldots + c_n x^{(n)},$$

де $c_i (i = 1, 2, \ldots, n)$ - деякі числові коефіцієнти. Звідси, враховуючи, що

$$Ax^{(i)} = \lambda_i x^{(i)},$$
$$A^2 x^{(i)} = \lambda_i^2 x^{(i)}, (i = 1, 2, \ldots, n),$$
$$\ldots\ldots\ldots\ldots$$

тоді одержимо:

$$y^{(1)} = c_1 \lambda_1 x^{(1)} + c_2 \lambda_2 x^{(2)} + \ldots + c_n \lambda_n x^{(n)},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y^{(n-1)} = c_1 \lambda_1^{n-1} x^{(1)} + c_2 \lambda_2^{n-1} x^{(2)} + \ldots + c_n \lambda_n^{n-1} x^{(n)}.$$

Нехай

$$\varphi_i(\lambda) = \lambda^{n-1} + q_{1i}\lambda^{n-2} + \ldots + q_{n-1,\,i},$$

$(i = 1, 2, \ldots, n)$ - довільна система поліномів. Складаючи лінійну комбінацію векторів $y^{(n-1)}, y^{(n-2)}, \ldots, y^{(0)}$ з коефіцієнтами $1, q_{1-i}, \ldots, q_{n-1,i}$ знаходимо:

$$y^{(n-1)} + q_{1i} y^{(n-2)} + \ldots + q_{n-1,\,i} y^{(0)} = c_1 \varphi_i(\lambda_1) x^{(1)} + c_2 \varphi_i(\lambda_2) x^{(2)} + \ldots + c_n \varphi_i(\lambda_n) x^{(n)}$$

Якщо покласти

$$\varphi_i(\lambda) = \frac{D(\lambda)}{\lambda - \lambda_i} \ (i = 1, 2, \ldots, n),$$

то очевидно, що

$$\varphi_i(\lambda_j) = 0 \text{ при } i \neq j$$

і

$$\varphi_i(\lambda_j) = D'(\lambda_i) \neq 0.$$

Тоді отримаємо. що

$$c_i\varphi_i(\lambda_i)x^{(i)} = y^{(n-1)} + q_{1i}y^{(n-2)} + ... + q_{n-1,i}y^{(0)}\ (i = 1, 2, ..., n).$$

Таким чином, якщо $c_i \neq 0$, то одержана лінійна комбінація векторів $y^{(n-1)}, y^{(n-2)}, ..., y^{(0)}$ дає власний вектор $x^{(i)}$ з точністю до числового множника.

Коефіцієнти $q_{j,i}(j = 1,2,..., n-1)$ можуть бути легко визначені за схемою Горнера

$$\left.\begin{aligned} q_{0i} &= 1, \\ q_{ji} &= \lambda_i q_{j-1,i} + p_j. \end{aligned}\right\}$$

Лекція 8
Розв'язання систем алгебраїчних рівнянь

Інженеру часто доводиться вирішувати алгебраїчні рівняння і системи рівнянь, що можуть являти собою самостійну задачу або частину більш складних задач. В обох випадках практична цінність чисельного методу в значній мірі визначається швидкістю і ефективністю отримання розв'язку. Розглянемо найбільш відомі чисельні методи і ефективні алгоритми розв'язання систем лінійних алгебраїчних рівнянь.

8.1 Основні поняття та визначення

Системою лінійних алгебраїчних рівнянь (СЛАР) називають систему виду:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2m}x_m = b_2 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nm}x_m = b_n \end{cases}, \qquad (8.1)$$

де $x_i, (i = \overline{1,m})$ – невідомі; $b_i, (i = \overline{1,n})$ – вільні члени системи; $a_{ij}, (i = \overline{1,n}, j = \overline{1,m})$ – коефіцієнти системи.

В матричному вигляді рівняння (8.1) приймає наступний вигляд:

$$A \times \vec{X} = \vec{B},$$

де $\vec{X} = \{x_1, x_2, ..., x_n\}$ – вектор невідомих; $\vec{B} = \{b_1, b_2, ..., b_n\}$ – вектор вільних членів;

$A = \begin{Bmatrix} a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nm} \end{Bmatrix}$ – матриця коефіцієнтів СЛАР.

Розв'язком системи лінійних алгебраїчних рівнянь (5.1) називають вектор $\vec{X}$, координати якого $\{x_1, x_2, ..., x_n\}$ при підстановці у систему, що розв'язують, перетворюють кожне рівняння системи в тотожність .

Кількість невідомих *m* в системі називають ***порядком*** СЛАР.

Систему лінійних алгебраїчних рівнянь називають ***сумісною***, якщо вона має хоча б один ненульовий розв'язок. В протилежному випадку СЛАР називають ***несумісною***.

СЛАР називається ***визначеною,*** якщо вона має тільки один розв'язок (випадок, коли $m=n$). Систему називають ***невизначеною,*** якщо вона має безліч розв'язків ($m \neq n$).

Система називається ***виродженою,*** якщо головний визначник системи дорівнює нулю. Система називається ***невиродженою***, якщо головний визначник системи не дорівнює нулю.

Дві системи називаються ***еквівалентними,*** якщо ці системи сумісні, визначені і мають однаковий розв'язок.

СЛАР можна розв'язати на ЕОМ чисельними методами, якщо вона сумісна, визначена, невироджена.

8.2 Класифікація методів розв'язання СЛАР

Для розв'язання СЛАР на ЕОМ традиційно використовують дві групи чисельних методів, що представлені на рис. 8.1:

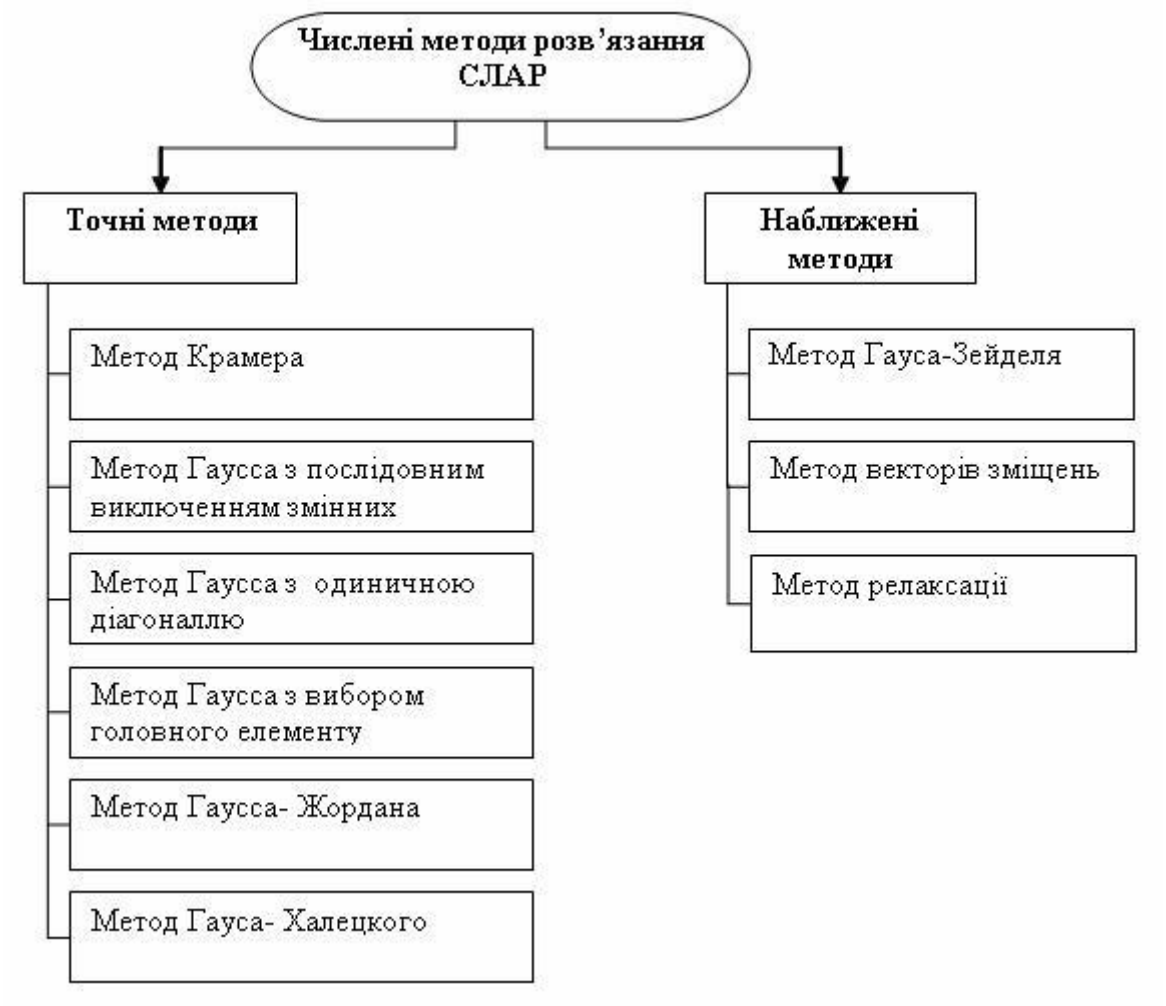

Рис. 8.1. Класифікація чисельних методів

До точних методів відносять методи, які дозволяють отримати точний розв'язок системи (8.1) за відповідну кількість операцій перетворення без урахування похибок заокруглення.

До наближених методів відносять методи, які дозволяють отримати розв'язок системи (5.1) у вигляді границі послідовності векторів $\lim\limits_{k\to\infty}\left\{\overline{X}^0,\overline{X}^1,\overline{X}^2,...,\overline{X}^n\right\}$, яка збігається до точного розв'язку системи, де:

$$\overline{X}^0=\begin{bmatrix}x_1^{(0)}\\x_2^{(0)}\\\vdots\\x_n^{(0)}\end{bmatrix},\quad \overline{X}^1=\begin{bmatrix}x_1^{(1)}\\x_2^{(1)}\\\vdots\\x_n^{(1)}\end{bmatrix},...,\overline{X}^n=\begin{bmatrix}x_1^{(n)}\\x_2^{(n)}\\\vdots\\x_n^{(n)}\end{bmatrix}$$

8.3 Особливості методів Гауса

Найбільш відомим з точних методів розв'язання системи лінійних алгебраїчних рівнянь (8.1) є методи Гауса, суть яких полягає в тому, що система

рівнянь, яка розв’язується, зводиться до еквівалентної системи з верхньою трикутною матрицею. Невідомі знаходяться послідовними підстановками, починаючи з останнього рівняння перетвореної системи. Алгоритми Гауса складаються із виконання однотипних операцій, які легко формалізуються. Однак, точність результату й витрачений на його отримання час у більшості випадків залежить від алгоритму формування трикутної матриці системи. У загальному випадку алгоритми Гауса складаються з двох етапів:

Прямий хід, в результаті якого СЛАР (8.1), що розв‘язується, перетворюється в еквівалентну систему з верхньою трикутною матрицею коефіцієнтів виду:

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ &0 \cdot x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ &\ldots \\ &0 \cdot x_1 + 0 \cdot x_2 + \ldots + a_{nn}x_n = b_n \end{aligned} . \qquad (8.2)$$

Зворотній хід дозволяє визначити вектор розв’язку починаючи з останнього рівняння системи (8.2) шляхом підстановки координат вектора невідомих, отриманих на попередньому кроці.

Відомо декілька різних алгоритмів отримання еквівалентної системи з верхньою трикутною матрицею. Розглянемо найбільш відомі з них.

Метод Гауса з послідовним виключенням невідомих

Метод Гауса з послідовним виключенням невідомих (базовий метод) засновано на алгоритмі, в основі якого лежить послідовне виключення невідомих вектора $\overline{X}$ з усіх рівнянь, починаючи з $(i+1)$–го, шляхом елементарних перетворень: перемноження обох частин рівняння на будь-яке число, крім нуля; додавання (віднімання) до обох частин одного рівняння відповідних частин другого рівняння, помножених на будь-яке число, крім нуля.

Суть алгоритму розглянемо на прикладі системи, яка складається з трьох лінійних алгебраїчних рівнянь з трьома невідомими:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases}. \quad (8.3)$$

1) Перевіряємо, щоб принаймні один із коефіцієнтів a_{11}, a_{21}, a_{31} не дорівнював нулю. Якщо, наприклад, $a_{11} = 0$, тоді необхідно переставити рівняння так, щоб коефіцієнт при x_1 у першому рівнянні не дорівнював нулю.

2) Обчислюємо множник:

$$M_2 = \frac{a_{21}}{a_{11}}. \quad (8.4)$$

3) Перше рівняння системи (8.3) множиться на M_2 і віднімається від другого рівняння системи, отриманої після перестановки рівнянь, якщо вона була необхідною. Результат обчислення має вигляд:

$$(a_{21} - M_2 a_{11})x_1 + (a_{22} - M_2 a_{12})x_2 + (a_{23} - M_2 a_{13})x_3 = b_2 - M_2 b_1, \quad (8.5)$$

але

$$a_{21} - M_2 a_{11} = a_{21} - \left(\frac{a_{21}}{a_{11}}\right) a_{11} = 0. \quad (8.6)$$

Тоді x_1 виключається із другого рівняння.

Позначимо нові коефіцієнти:

$$\begin{aligned} a'_{22} &= a_{22} - M_2 a_{12} \\ a'_{23} &= a_{23} - M_2 a_{13} \\ b'_2 &= b_2 - M_2 b_1 \end{aligned} \quad (8.7)$$

Тоді друге рівняння системи (8.3) набуває вигляду:

$$a'_{22}x_2 + a'_{23}x_3 = b_2. \quad (8.8)$$

Далі необхідно звільнитися від коефіцієнта a_{31} при x_1 в третьому рівнянні системи (8.3) за аналогічним алгоритмом.

4) Обчислюється множник для третього рівняння:

$$M_3 = \frac{a_{31}}{a_{11}}. \quad (8.9)$$

5) Перше рівняння системи (8.3) множиться на M_3 і віднімається від третього рівняння. Коефіцієнт при x_1 стає нулем, і третє рівняння набуває вигляду:

$$a'_{32}x_2 + a'_{33}x_3 = b_3, \tag{8.10}$$

де

$$a'_{32} = a_{32} - M_3 a_{12}, \tag{8.11}$$

$$a'_{33} = a_{33} - M_3 a_{13}, \tag{8.12}$$

$$b'_3 = b_3 - M_3 b_1, \tag{8.13}$$

Перетворена таким чином система рівнянь (8.3) набуває вигляду:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ 0 \cdot x_1 + a'_{22}x_2 + a'_{23}x_3 = b'_2 \\ 0 \cdot x_1 + a''_{32}x_2 + a''_{33}x_3 = b''_3 \end{cases}. \tag{8.14}$$

Ця система рівнянь еквівалентна початковій і має певні переваги, оскільки x_1 входить тільки до першого рівняння.

Тепер виключаємо x_2 з останнього рівняння. Якщо $a_{22} = 0$, а $a_{32} \neq 0$, тоді друге й третє рівняння переставляється так, щоб $a_{22} \neq 0$. Інакше система вироджена і має безліч розв'язків.

7) Обчислюємо множник $M''_3 = \dfrac{a_{32}}{a_{22}}$.

8) Друге рівняння системи множиться на M''_3 і віднімається від 3-го рівняння. При цьому коефіцієнт біля x_2 стає рівним нулю. Тоді отримуємо:

$$a''_{33}x_3 = b''_3 \tag{8.15}$$

Замінивши в системі (8.14) третє рівняння на (8.15), отримаємо систему рівнянь виду:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ 0 \cdot x_1 + a'_{22}x_2 + a'_{23}x_3 = b'_2 \\ 0 \cdot x_1 + 0 \cdot x_2 + a''_{33}x_3 = b''_3 \end{cases}. \tag{8.16}$$

Таку систему називають ***системою з трикутною матрицею коефіцієнтів,*** що еквівалентна СЛАР (8.3). Процес знаходження такої системи називається ***прямим ходом Гауса.*** Знайти розв'язок такої системи просто: із 3-го рівняння знайти x_3, підставити результат у друге і знайти x_2, підставити x_2 і x_3 в 1-е рівняння системи (8.16) і знайти x_1:

$$x_3 = \frac{b_3''}{a_{33}''}, \quad x_2 = \frac{b_2' - a_{23}' x_3}{a_{22}'}, \quad x_1 = \frac{b_1 - a_{12}x_2 - a_{13}x_3}{a_{11}}.$$

Процес знаходження вектора розв'язку системи (5.3) називають ***зворотнім ходом метода Гауса***.

Метод Гауса за схемою Халецького

Алгоритм метода включає також прямий і зворотній хід. Кінцевою метою прямого ходу є отримання СЛАР, яка еквівалентна заданій, з верхньою трикутною матрицею коефіцієнтів. Для цього матрицю коефіцієнтів початкової системи рівнянь *A* розбивають на дві трикутні:

$$A = C \cdot D, \tag{8.17}$$

де матриця *C* – нижня трикутна матриця; *D* – верхня трикутна матриця з одиничною головною діагоналлю:

$$C = \begin{bmatrix} c_{11} & 0 & \dots & 0 \\ c_{21} & c_{22} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}, \quad D = \begin{bmatrix} 1 & d_{12} & \dots & d_{1n} \\ 0 & 1 & \dots & d_{2n} \\ \dots & \dots & \dots & d_{3n} \\ 0 & 0 & \dots & d_{nn} \end{bmatrix}.$$

Алгоритм визначення коефіцієнтів матриць *C* і *D*.

1) Обчислюється перший стовпець матриці *C*, перший рядок матриці *D* і y_1 за формулами:

$$\begin{aligned} & c_{i1} = a_{i1}, \ i = \overline{1,n} \\ & d_{11} = 1, \quad d_{1i} = \frac{a_{1i}}{a_{11}}, \ i = \overline{2,n}. \\ & y_1 = \frac{b_1}{a_{11}} \end{aligned} \tag{8.18}$$

2) Обчислюються елементи другого стовпця матриці C, елементи другого рядка матриці D та елемент y_2:

$$\begin{aligned} & c_{i2} = a_{i2} - a_{i1}d_{12}, \ \ i = \overline{2,n} \\ & d_{21} = \frac{a_{2i} - c_{2i}d_{1i}}{a_{22}} \\ & y_2 = \frac{b_2 - c_{21}y_1}{a_{22}} \end{aligned} \quad . \tag{8.19}$$

3) Обчислюють елементи третього стовпця матриці C, елементи третього рядка матриці D та елемент y_3:

$$\begin{aligned} c_{i3} &= a_{i3} - (c_{i1}d_{31} - c_{i2}d_{23}),\ \overline{i=2,n} \\ d_{31} &= \frac{a_{31} - (c_{31}d_{1i} + c_{32}d_{2i})}{c_{33}} \\ y_3 &= \frac{b_3 - (c_{31}y_1 + c_{32}y_2)}{c_{33}} \end{aligned} \quad . \qquad (8.20)$$

Загальний вигляд формул для обчислення c_{ki}, d_{ki}, y_i елементів матриць *C*, *D* і *Y*:

$$\begin{aligned} & d_{ii} = 1,\ d_{1i} = \frac{a_{1i}}{a_{11}},\ c_{i1} = a_{i1},\ y_1 = \frac{b_1}{a_{11}}, \quad i = \overline{1,n} \\ & c_{ki} = a_{ki} - \sum_{j=1}^{i=1} c_{kj}d_{ji}, \quad k = \overline{i,n},\ i = \overline{2,n} \\ & d_{ik} = \frac{a_{ik} - \sum_{j=1}^{i=1} c_{ij}d_{jk}}{c_{ii}},\ k = \overline{i+1,n} \\ & y_i = \frac{b_i - \sum_{j=1}^{i=1} c_{ij}y_{ij}}{c_{ii}} \end{aligned} \quad . \qquad (8.21)$$

Метод Гауса з вибором головного елемента

Ідея цього методу виникла у зв'язку з тим, що коефіцієнти СЛАР є параметрами реальних інженерних систем та в більшості є наближеними значеннями тому, що отримані звичайно в результаті вимірювання або як статистичні дані. Для таких систем рівнянь при обчисленні масштабного множника $M = \frac{a_{ik}}{a_{kk}}$ можлива ситуація при визначені a_{kk}, що ділення наближеного числа a_{ik} на достатньо мале число a_{kk} веде до різкого збільшення похибки методу. Тому для того, щоб не збільшувати похибку результату, необхідно виконувати такі дії:

1) в системі (5.1) необхідно знайти з *k*-го стовпця найбільший за абсолютним значенням коефіцієнт a_{kj};

2) переставити *k*-те рівняння з рівнянням, у якому знаходиться цей максимальний коефіцієнт;

3) масштабний множник буде обчислюватись за формулою, де a_{kk} – максимальний коефіцієнт, а тому похибка розв'язання СЛАР у результаті арифметичних операцій не збільшується.

Метод Гауса з одиничними коефіцієнтами

В цьому методі зроблена спроба зменшити недоліки перших двох методів пов'язаних з багаторазовим діленням одного наближеного числа на інше. Для цього перед введенням масштабного множника *k*-те рівняння системи ділиться один раз на діагональний елемент a_{kk} так, щоб коефіцієнт при $x_k = 1$, а масштабний множник M_i буде дорівнювати a_{kj}. Результатом прямого ходу є система, еквівалентна СЛАР (5.1), з одиничними коефіцієнтами на головній діагоналі виду:

$$\begin{cases} 1 \cdot x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\ 0 \cdot x_1 + 1 \cdot x_2 + ... + a_{2n}x_n = b_2 \\ \vdots \\ 0 \cdot x_1 + 0 \cdot x_2 + ... + 1 \cdot x_n = b_n \end{cases}.$$

Дана система схожа на систему (8.2), яка отримується в результаті прямого ходу базового методу Гауса з послідовним вилученням невідомих і відрізняється від неї тільки діагональними коефіцієнтами. Для отримання такої системи необхідно використовувати алгоритм, який включає в себе наступні етапи:

1. Організація циклу по всім рівнянням від 1 до $(n-1)$ - $(k = 1,2,...,n-1)$.

2. В кожному *k*-му стовпці визначається номер *l*-го рівняння з головним елементом (тобто номер *l*-го рівняння, в якому знаходиться коефіцієнт при x_n зі всіх рівнянь починаючи з *k*-го до *n*-го).

3. Якщо номер цього рівняння *l* не дорівнює *k*, тоді необхідно переставити місцями *l*-е рівняння з *k*-м.

4.Нормування *k*-го рівняння, тобто ділення всіх коефіцієнтів *k*-го рівняння на a_{kk} (головний елемент при x_n), включаючи b_n.

5. Перетворення всіх *i*-х рівнянь, починаючи з $(k+1)$ до *n* у відповідності з базовим алгоритмом Гауса з метою отримати еквівалентну систему з верхньою трикутною матрицею коефіцієнтів.

6. Кінець циклу по *k*.

Метод Гауса-Жордана

Особливістю метода Гауса-Жордана є перетворення системи (8.1) (прямий хід) до еквівалентної з одиничною матрицею коефіцієнтів виду:

$$\begin{cases} 1\cdot x_1 + 0\cdot x_2 + ... + 0\cdot x_n = b_1 \\ 0\cdot x_1 + 1\cdot x_2 + ... + 0\cdot x_n = b_2 \\ \vdots \\ 0\cdot x_1 + 0\cdot x_2 + ... + 1\cdot x_n = b_n \end{cases}, \qquad (8.22)$$

тобто системи, яка містить тільки одиничну діагональ.

Для отримання такої системи в прямий хід алгоритму базового методу Гауса (з послідовним виключенням невідомих) додатково вводяться такі дії:

1. Організація циклу по *k* по всім рівнянням від 1 до $(n-1)$ - $(k = 1,2,...,n-1)$.

2. Процедура вибору головного елементу в кожному *k*-му стовпці при x_k;

3. Процедура нормування *k*-го рівняння системи, тобто в *k*-му рівнянні кожен коефіцієнт a_{kj} розділити на a_{kk}, включаючи b_k, так, щоб коефіцієнт $a_{kk} = 1$.

4. Перетворення всіх рівнянь системи, починаючи з 1-го до *n* у відповідності з базовим алгоритмом Гауса з метою отримати еквівалентну систему з одиничною діагоналлю. В даному випадку для розрахунку коефіцієнтів a_{ij} використовуються ті самі формули, що і в базовому алгоритмі Гауса:

$$M = a_{ik},\ a_{ij} = a_{ij} - M\cdot a_{kj},\ b_i = b_i - M\cdot b_k,$$

але використовуються вони для всіх рівнянь з 1-го до *n* крім *k*-го, в якому остається коефіцієнт a_{kk} рівний одиниці.

5. Кінець циклу по *k*.

8.4. Наближені методи розв'язання СЛАР

До наближених методів відносяться методи, які дозволяють розв'язок системи отримати як границю послідовних *k* розв'язків системи (8.1) при $k \to \infty$ виду:

$$\bar{x} = \lim\left\{\bar{x}^0, \bar{x}^1, \bar{x}^2, ..., \bar{x}^k\right\},$$

де $\bar{x}^0$ - вектор розв'язку 0-го наближення, $\bar{x}^1$ - вектор розв'язку 1-го наближення і т.д.

Для розв'язання СЛАР наближеними методами найбільшу цікавість представляють такі методи:

- метод послідовних наближень;
- метод Гауса-Зейделя;
- метод верхньої релаксації.

Розглянемо особливості загального підходу до розв'язання СЛАР наближеними методами.

Дано систему лінійних алгебраїчних рівнянь виду (8.2), розв'язання якої методами послідовних наближень необхідно виконати наступні кроки:

1) Кожне рівняння системи розділити на діагональний елемент a_{kk}, де $k = 1,2,...,n$ (n – кількість рівнянь в системі), і перетворити кожне рівняння системи відносно координат вектора, індекс якого співпадає з номером рівняння:

$$\begin{cases} x_1 = \dfrac{b_1}{a_{11}} - \left(\dfrac{a_{12}}{a_{11}} x_2 + \dfrac{a_{13}}{a_{11}} x_3 + ... + \dfrac{a_{1n}}{a_{11}} x_n \right) \\ x_2 = \dfrac{b_2}{a_{22}} - \left(\dfrac{a_{21}}{a_{22}} x_1 + \dfrac{a_{23}}{a_{22}} x_3 + ... + \dfrac{a_{2n}}{a_{22}} x_n \right) \\ x_3 = \dfrac{b_3}{a_{33}} - \left(\dfrac{a_{31}}{a_{33}} x_1 + \dfrac{a_{32}}{a_{33}} x_2 + ... + \dfrac{a_{3n}}{a_{33}} x_n \right) \\ \vdots \\ x_n = \dfrac{b_n}{a_{nn}} - \left(\dfrac{a_{n1}}{a_{nn}} x_1 + \dfrac{a_{n2}}{a_{nn}} x_2 + ... + \dfrac{a_{n(n-1)}}{a_{nn}} x_{n-1} \right) \end{cases} . \qquad (8.23)$$

2) Нехай $\dfrac{b_k}{a_{kk}} = \beta_k$, а $\left(-\dfrac{a_{ki}}{a_{kk}} \right) = \alpha_{ki}$, де $k,i = 1,2,...,n$. Тоді система (8.23) матиме вигляд:

$$\begin{cases} x_1 = \beta_1 + \alpha_{11}x_2 + \alpha_{12}x_2 + ... + \alpha_{1n}x_n \\ x_2 = \beta_2 + \alpha_{11}x_1 + \alpha_{22}x_2 + ... + \alpha_{2n}x_n \\ x_3 = \beta_3 + \alpha_{11}x_1 + \alpha_{22}x_2 + ... + \alpha_{3n}x_n \\ \vdots \\ x_n = \beta_n + \alpha_{11}x_1 + \alpha_{22}x_2 + ... + \alpha_{nn}x_n \end{cases}. \qquad (8.24)$$

Така система називається зведеною до нормального вигляду.

3) Представимо систему (8.24) в матричному вигляді:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \dots & \alpha_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \dots & \alpha_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix} \qquad (8.25)$$

або векторному

$$\vec{x} = \vec{\beta} + \vec{\alpha} \cdot \vec{x}. \qquad (8.26)$$

Якщо деяким чином визначити, так званий, вектор початкових значень $\vec{x}^{(0)}$, який знаходиться в правій частині (5.26), то можна отримати певні значення вектора $\vec{x}$.

В якості вектора початкових наближень $\vec{x}^{(0)}$ вибирають:

- вектор, в якого всі координати x_i дорівнюють 0;
- вектор, в якого всі координати x_i дорівнюють 1;
- вектор, координати x_i якого дорівнюють координатам вектора вільних членів β_i;
- координати вектору $\vec{x}$ вибирають в результаті аналізу особливостей об’єкту дослідження та задачі, яка розв’язується.

4) Якщо вектор початкових наближень $\vec{x}^{(0)}$ підставити в праву частину системи (8.25) або (8.26), то вона прийме вигляд:

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \\ \dots \\ x_n^{(1)} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \dots & \alpha_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \dots & \alpha_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \\ \dots \\ x_n^{(0)} \end{bmatrix}$$

або

$$\vec{x}^{(1)} = \vec{\beta} + \vec{\alpha} \cdot \vec{x}^{(0)}.$$

Отримана система легко розв'язується, тому що в правій частині містить всі визначені елементи, і дозволяє отримати розв'язок системи, який називається ***вектором першого наближення*** $\vec{x}^{(1)}$.

5) Перевіряється виконання умови закінчення ітераційного процесу пошуку розв'язку системи (8.2) виду:

$$\left|\vec{x}^{(1)} - \vec{x}^{(0)}\right| \le \varepsilon, \qquad (8.27)$$

де ε - задана похибка результатів розв'язання задачі.

Якщо умова (8.27) не виконується, то $x^{(1)}$ підставляється в праву частину (8.25) або (8.26) і знаходиться $x^{(2)}$ з системи виду:

$$\begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \\ \dots \\ x_n^{(2)} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \dots & \alpha_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \dots & \alpha_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \\ \dots \\ x_n^{(1)} \end{bmatrix}$$

або

$$\vec{x}^{(2)} = \vec{\beta} + \vec{\alpha} \cdot \vec{x}^{(1)}.$$

6) Знову перевіряється виконання умови закінчення ітераційного процесу пошуку розв'язку системи (8.2).

Якщо умова не виконується, то $x^{(2)}$ підставляється в праву частину (8.25) і знаходиться $x^{(3)}$ і т.д.

7) Етапи 4 та 5 повторюються доти, доки не виконується умова закінчення ітераційного процесу пошуку розв'язку системи (8.2).

Таким чином, процес пошуку розв'язку системи (8.2) наближеними методами з заданою похибкою ε є ***ітераційним***, а умовою виходу з цього процесу є умова (8.27).

Описаний вище алгоритм дозволяє отримати розв'язок системи (8.2) близький до точного (з заданою похибкою ε) тільки в тому випадку, коли ітераційний процес пошуку розв'язку СЛАР збігається.

Теорема про збіжність. Ітераційний процес пошуку розв'язку системи лінійних алгебраїчних рівнянь виду (8.25) наближеними методами збігається, якщо будь-яка канонічна норма матриці $\|\alpha\| < 1$.

Канонічною нормою матриці називається будь-яке дійсне додатне число, яке визначається за такими умовами:

<u>перша канонічна норма</u> – це максимальна з сум модулів елементів матриці коефіцієнтів α по стрічкам:

$$\|\alpha\|_1 = \max_i \sum_{j=1}^{n} |\alpha_{ij}|, \tag{8.28}$$

<u>друга канонічна норма</u> – це максимальна з сум модулів елементів матриці коефіцієнтів α по стовбцям:

$$\|\alpha\|_2 = \max_j \sum_{i=1}^{n} |\alpha_{ij}|, \tag{8.29}$$

<u>третя канонічна норма</u> – це корінь квадратний з сум квадратів модулів всіх елементів матриці коефіцієнтів α:

$$\|\alpha\|_3 = \sqrt{\sum_i \sum_j |\alpha_{ij}|}^2.$$

<u>Наслідок 1</u>: Ітераційний процес розв'язання системи (8.24) збігається, якщо сума модулів елементів стрічок матриці коефіцієнтів α або сума модулів елементів її стовбців менш одиниці, тобто виконуються умови

$$\max_i \sum_{j=1}^{n} |\alpha_{ij}| < 1 \quad \text{або} \quad \max_j \sum_{i=1}^{n} |\alpha_{ij}| < 1.$$

<u>Наслідок 2</u>: Ітераційний процес розв'язання системи (8.24) збігається, якщо елементи головної діагоналі більше суми модулів елементів відповідної стрічки крім діагонального елементу цієї стрічки, тобто виконуються умови:

$$|\alpha_{ii}| > \max_i \sum_{j=1}^{n} |\alpha_{ij}| \quad \text{або} \quad |\alpha_{jj}| > \max_j \sum_{i=1}^{n} |\alpha_{ij}|.$$

Розглянемо особливості алгоритмів наближених методів.

<u>Метод послідовних наближень (метод Якобі)</u>

Нехай задана система лінійних алгебраїчних рівнянь виду (8.2). Метод послідовних наближень (метод Якобі) відноситься до ітераційних методів, тому

потребує перетворити дану систему до нормального вигляду (8.25) та знайти канонічні норми матриці $\overline{\alpha}$ для того, щоб визначити умови збіжності ітераційного процесу пошуку розв'язку системи із заданою похибкою ε відповідно теоремі про збіжність. Якщо жодна з умов не виконується, то дану систему необхідно перетворити по певним правилам, та знову перевірити умови збіжності ітераційного процесу. Якщо жодна з умов знову не виконується, то метод послідовних наближень не має сенсу використовувати. Якщо хоча б одна з умов виконалась, то ітераційний процес пошуку розв'язку системи із заданою похибкою ε збігається і метод послідовних наближень можна використовувати.

Оцінка похибки метода Якобі

Якщо задана допустима похибка обчислень ε і x - вектор точного розв'язку системи лінійних рівнянь, а $x_j^{(k)}$ k-те наближення до вектору точного розв'язку, то для оцінки похибки метода Якобі послідовних наближень використовується формула:

$$\left\| x_j - x_j^k \right\| \le \frac{\|\alpha\|^{k+1}}{1-\|\alpha\|} \|\beta\|, \tag{8.30}$$

де $\|\alpha\|$ - одна з трьох норм матриці α; $\|\beta\|$ - аналогічна норма вектора β; k - кількість ітерацій, необхідна для досягнення потрібної точності ε.

Метод Гауса-Зейделя

Метод Зейделя являє собою модифікацію метода послідовних наближень, при чому у методі Зейделя при обчисленні *i*-ої координати вектора розв'язку $(k+1)$-го наближення використовуються значення всіх $(i-1)$ координат вектора $(k+1)$-го наближення, обчисленого раніше. Розглянемо метод більш детально.

Нехай початкова система лінійних алгебраїчних рівнянь приведена до нормального вигляду:

$$\begin{cases} x_1 = \beta_1 + \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n \\ x_2 = \beta_2 + \alpha_{21}x_1 + \alpha_{22}x_2 + \ldots + \alpha_{2n}x_n \\ \vdots \\ x_n = \beta_n + \alpha_{n1}x_1 + \alpha_{n2}x_2 + \ldots + \alpha_{nn}x_n \end{cases}. \tag{8.31}$$

Вибрати значення координат вектора початкових наближень $\bar{x} = \{x_1^{(0)}, ..., x_n^{(0)}\}$.

Визначити значення першої координати $x_1^{(1)}$ вектора першого наближення з першого рівняння системи:

$$x_1^{(1)} = \beta_1 + \alpha_{11}x_1^{(0)} + \alpha_{12}x_2^{(0)} + ... + \alpha_{1n}x_n^{(0)}.$$

Підставити в друге рівняння системи значення першої координати $x_1^{(1)}$, яке обчислене на попередньому кроці

$$x_2^{(1)} = \beta_1 + \alpha_{21}x_1^{(1)} + \alpha_{22}x_2^{(0)} + ... + \alpha_{2n}x_n^{(0)}.$$

Отримані значення координат першого наближення $x_1^{(1)}$, $x_2^{(1)}$ підставити у третє рівняння системи

$$x_3^{(1)} = \beta_1 + \alpha_{21}x_1^{(1)} + \alpha_{22}x_2^{(1)} + \alpha_{23}x_3^{(0)} + ... + \alpha_{2n}x_n^{(0)}$$

для знаходження третьої координати і т.д.

Для знаходження останньої координати вектора першого наближення $x_n^{(1)}$ в останнє рівняння системи треба підставити значення всіх (*n-1*) координат $(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, ..., x_{n-1}^{(1)})$, які отримані на попередніх кроках та значення координати $x_n^{(0)}$

$$x_n^{(1)} = \beta_n + \alpha_{n1}x_1^{(1)} + \alpha_{n2}x_2^{(1)} + ... + \alpha_{nn-1}x_{n-1}^{(1)} + \alpha_{nn}x_n^{(0)}.$$

Аналогічно будують друге, третє та інші наближення.

Умови збіжності ітераційного процесу Зейделя

Даний процес розв'язання СЛАР - ітераційний, тому важливим є аналіз умов збіжності ітераційного процесу. Процес Зейделя для системи лінійних рівнянь $\bar{x} = \bar{\beta} + \alpha\bar{x}$ збігається до точного розв'язку із заданою похибкою при будь-якому виборі вектора початкових наближенньь, якщо будь яка норма матриці α менша 1.

Відомо, що процес Зейделя сходиться до точного розв'язку СЛАР швидше, ніж метод послідовних наближень.

Оцінка похибки методу Гауса-Зейделя

Якщо $\bar{x}$ - точне значення вектора розв'язку системи лінійних рівнянь; а $\bar{x}^{(k)}$ - k-е наближення, обчислене за методом Гауса-Зейделя, то для оцінки похибки цього метода використовується формула:

$$\left\|\bar{x} - \bar{x}^k\right\|_1 \le \frac{\left\|\alpha\right\|_1^k}{1 - \left\|\alpha\right\|_1}\left\|\bar{x}^{(1)} - \bar{x}^{(0)}\right\|_1.$$

Метод верхньої релаксації

В основі метода верхньої релаксації використовується алгоритм та обчислювальна схема метода Гауса-Зейделя, але на відміну від нього нові значення координат вектора k-го наближення визначаються за формулами:

$$x_i^{k+1} = x_i^k + \varpi(\bar{x}_i^{k+1} - x_i^k),$$

де $\bar{x}_i^{k+1}$ - уточнене значення змінної по методу Гауса-Зейделя, ϖ - параметр релаксації, значення якого визначається з інтервалу $1 \le \varpi \le 2$. При $\varpi = 1$ метод тотожний методу Гауса-Зейделя. Швидкість збіжності ітераційного процесу залежить від значення ϖ.

Лекція 9
Одновимірні методи нелінійної оптимізації

9.1 Метод рівномірного пошуку

Постановка задачі. Необхідно знайти абсолютний мінімум функції $f(x)$ однієї змінної, тобто таку точку $x^* \in R$, що $f(x^*) = \min_{x \in R} f(x)$.

Стратегія пошуку. Метод відноситься до пасивних стратегій. Задається початковий інтервал невизначеності $L_0 = [a_0, b_0]$ і кількість обрахунків функції N. Обрахунки проводяться в N рівнорозміщених одна від одної точках (при цьому інтервал L_0 ділиться на $N+1$ рівних інтервалів). Шляхом порівняння величин $f(x_i), i = 1,..., N$ знаходиться точка x_k, в котрій значення функції найменше. Шукана точка мінімуму x^* вважається заключною в інтервалі $[x_{k-1}, x_{k+1}]$ (рис. 9.1)

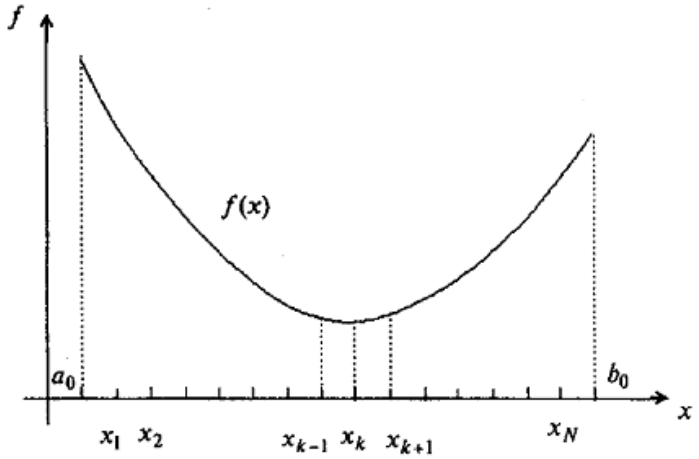

Рис. 9.1 Ілюстрація методу рівномірного пошуку

Алгоритм

Крок 1. Задати початковий інтервал невизначеності $L_0 = [a_0, b_0]$, N – кількість обрахованих функцій.

Крок 2. Обрахувати точки $x_i = a_0 + i \cdot \frac{(b_0 - a_0)}{N+1}, i = 1,..., N$, рівнорозміщенні одна від одної.

Крок 3. Обрахувати значення функції в N знайдених точках: $f(x_i), i = 1,..., N$

Крок 4. Серед точок $x_i, i = 1,..., N$, знайти таку, в котрій функція приймає найменше значення: $f(x_k) = \min\limits_{1 \le i \le N} f(x_i)$

Крок 5. Точка мінімуму x^* належить інтервалу: $x^* \in [x_{k-1}, x_{k+1}] = L_N$, на котрому в якості наближеного розв'язку може бути обрана точка $x^* = x_k$.

9.2 Метод половинного поділу

Постановка задачі. Необхідно знайти абсолютний мінімум функції $f(x)$ однієї змінної, тобто таку точку $x^* \in R$, що $f(x^*) = \min\limits_{x \in R} f(x)$.

Стратегія пошуку. Метод відноситься до послідовних стратегій і дозволяє виключити з подальшого розгляду на кожній ітерації в точності половину поточного інтервалу невизначенності. Задається початковий інтервал невизначеності, а алгоритм зменшення інтервалу, котрий являється в загальному випадку, «гарантуючим», оснований на аналізі величин функції в трьох точках, рівномірно розміщених на поточному інтервалі (ділять його на чотири рівні частини). Умови закінчення процесу пошуку стандартні: пошук закінчується, коли довжина поточного інтервалу невизначеності стає меншим встановленої величини.

Алгоритм:

Крок 1. Задати початковий інтервал невизначеності $L_0 = [a_0, b_0]$ та $l > 0$ - необхідну точність.

Крок 2. Допустимо, що $k = 0$.

Крок 3. Вирахувати середню точку $x_k^c = \frac{a_k + b_k}{2}, |L_{2k}| = b_k - a_k, f(x_k^c)$.

Крок 4. Вирахувати точки: $y_k = a_k + \frac{|L_{2k}|}{4}, z_k = b_k - \frac{|L_{2k}|}{4}$ і $f(y_k), f(z_k)$. Треба відмітити, що точки y_k, x_k^c, z_k ділять інтервал $[a_k, b_k]$ на чотири рівні частини.

Крок 5. Порівняти вирази $f(y_k)$ і $f(x_k^c)$:

а) якщо $f(y_k) < f(x_k^c)$, виключити інтервал $(x_k^c, b_k]$, присвоївши $b_{k+1} = x_k^c, a_{k+1} = a_k$. Середньою точкою нового інтервалу стає точка $y_k : x_{k+1}^c = y_k$ (рис. 9.2, а). Перейти до кроку 7;

б) якщо $f(y_k) \geq f(x_k^c)$, перейти до кроку 6.

Крок 6. Порівняти $f(z_k)$ з $f(x_k^c)$:

а) якщо $f(z_k) < f(x_k^c)$, виключити інтервал $[a_k, x_k^c)$, присвоївши $a_{k+1} = x_k^c, b_{k+1} = b_k$. Середньою точкою нового інтервалу стає точка $z_k : x_{k+1}^c = z_k$ (рис. 9.2, б). Перейти до кроку 7;

б) якщо $f(z_k) \geq f(x_k^c)$, виключити інтервали $[a_k, y_k), (z_k, b_k]$, присвоївши $a_{k+1} = y_k, b_{k+1} = z_k$. Середньою точкою нового інтервалу залишається $x_k^c : x_{k+1}^c = x_k^c$ (рис. 9.2,в)

Крок 7. Обрахувати $|L_{2(k+1)}| = |b_{k+1} - a_{k+1}|$ і перевірити умову закінчення:

а) якщо $|L_{2(k+1)}| \leq l$, процес пошуку закінчується і $x^* \in L_{2(k+1)} = [a_{k+1}, b_{k+1}]$. В якості наближеного рішення можна взяти середину останнього інтервалу: $x^* = x_{k+1}^c$;

б) якщо $|L_{2(k+1)}| > l$, то присвоїти k=k+1 і перейти до кроку 4.

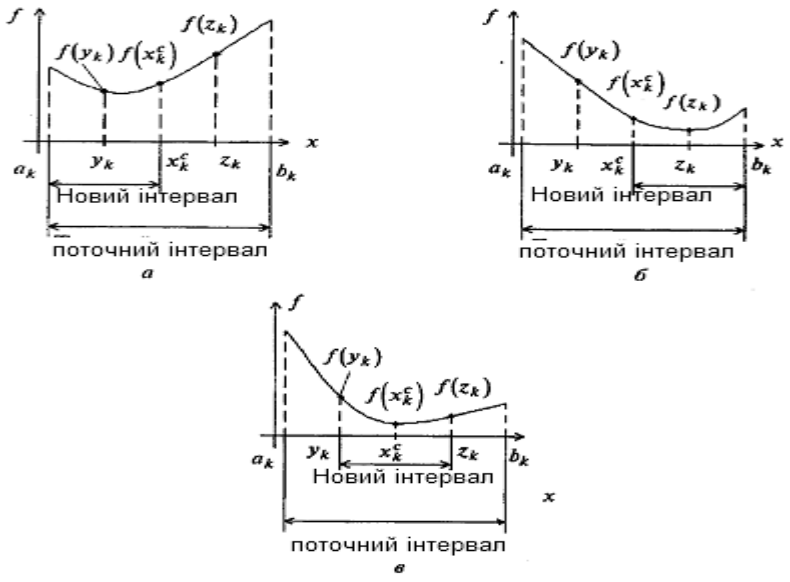

Рис.9.2 Ілюстрація методу ділення інтеравлу навпіл

9.3 Метод дихотомії

Постановка задачі. Потрібно знайти абсолютний мінімум функції $f(x)$ однієї змінної, тобто таку точку $x^* \in R$, що $f(x^*) = \min\limits_{x \in R} f(x)$.

Стратегія пошуку. Метод відноситься до послідовних стратегій. Задається початковий інтервал невизначеності і потрібна точність. Алгоритм опирається на аналіз значень функції в двух точках. Для їх знаходження поточний інтервал невизначеності ділиться навпіл і в обидві сторони від середини відкладається по $\frac{\varepsilon}{2}$, де ε - мале позитичне число. Умови закінчення процесу пошуку стандартні: пошук закінчується тоді, коли довжина поточного інтервала невизначеності стає менше встановленої величини.

Алгоритм:

Крок 1. Задати початковий інтервал невизначеності $L_0 = [a_0, b_0], \varepsilon > 0$-мале число, $l > 0$- точність.

Крок 2. Допустимо, що k=0.

Крок 3. Обрахувати $y_k = \frac{a_k + b_k - \varepsilon}{2}, f(y_k), z_k = \frac{a_k + b_k + \varepsilon}{2}, f(z_k)$.

Крок 4. Порівняти $f(y_k)$ з $f(z_k)$:

а) якщо $f(y_k) \le f(z_k)$, присвоїти $a_{k+1} = a_k, b_{k+1} = z_k$ (рис. 9.3,а) і перейти до кроку 5;

б) якщо $f(y_k) > f(z_k)$, присвоїти $a_{k+1} = y_k, b_{k+1} = b_k$ (рис. 9.3,б).

Крок 5. Обрахувати $|L_{2(k+1)}| = |b_{k+1} - a_{k+1}|$ і перевірити умови закінчення:

а) якщо $|L_{2(k+1)}| \le l$, процес пошуку закінчується і $x^* \in L_{2(k+1)} = [a_{k+1}, b_{k+1}]$. В якості наближеного рішення можна взяти середину останнього інтервалу: $x^* = \frac{a_{k+1} + b_{k+1}}{2}$;

б) якщо $|L_{2(k+1)}| > l$, присвоїти k=k+1і перейти до кроку 3.

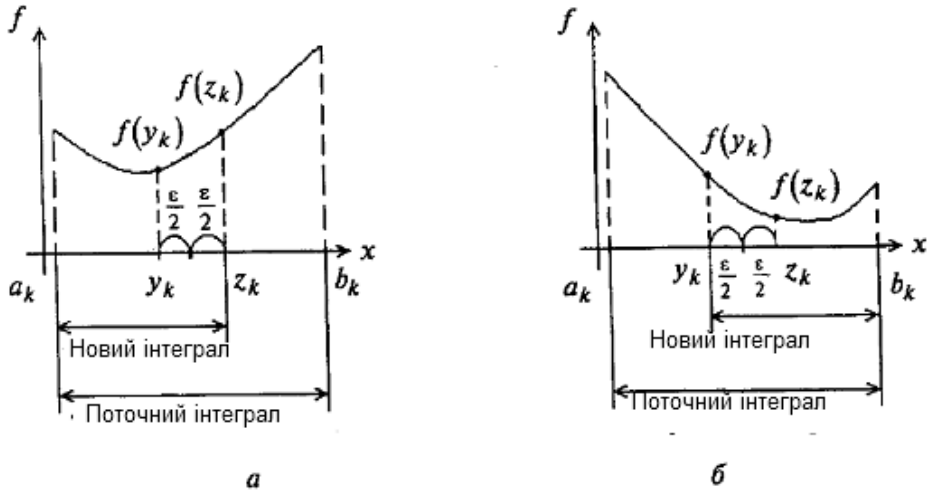

Рис. 9.3 Ілюстрація методу дихотомії

9.4 Метод золотого січення

Постановка задачі. Потрібно знайти абсолютний мінімум функції $f(x)$ однієї змінної, тобто таку точку $x^* \in R$, що $f(x^*) = \min_{x \in R} f(x)$.

Для побудови конкретного методу одномірної мінімізації, який працює по принципу послідовного зменшення інтервалу невизначеності, потрібно задати правило вибору на кожному кроці дві внутрішні точки. Звісно, бажано, щоб одна з них завжди використовувалась в якості внутрішньої і для наступного інтервалу. Тоді число обрахунків функції зменшиться вдвоє і одна ітерація потребує розрахунку лише одного нового значення функції. В методу золотого січення в якості двох внутрішніх точок вибираються точки золотого січення.

Визначення. Точка робить «золоте січення» відрізку, якщо відношення довжини всього відрізка до більшої частини рівне відношенню більшої частини до меншої.

На відрізку $[a_0, b_0]$ є дві симетричні відносно його кінців точки y_0 та z_0:

$$\frac{b_0 - a_0}{b_0 - y_0} = \frac{b_0 - y_0}{y_0 - a_0} = \frac{b_0 - a_0}{z_0 - a_0} = \frac{z_0 - a_0}{b_0 - z_0} = \frac{1+\sqrt{5}}{2} \cong 1{,}618.$$

Крім того, точка y_0 робить золоте січення відрізка $[a_0, z_0]$, а точка z_0 - відрізка $[y_0, b_0]$(рис .9.4).

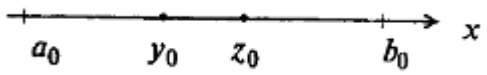

Рис. 9.4 Ілюстрація «золотого січення»

Стратегія пошуку. Метод відноситься до послідовних стратегій. Задається початковий інтервал невизначеності і шукана точність. Алгоритм зменшення інтервалу спирається на аналіз значень функції в двух точках. В якості точок обрахунку функції вибираються точки золотого січення. Тоді з врахуванням властивостей золотого січення на кожній ітерації, крім першої, потрібний лише один новий розрахунок функції. Умови закінчення процесу пошуку стандартні: пошук закінчується, коли довжина поточного інтервалу невизначеності стає меншим встановленої величини.

Алгоритм:

Крок 1. Задати початковий інтервал невизначеності $L_0 = [a_0, b_0]$, точність $l > 0$.

Крок 2. Присвоїти $k = 0$.

Крок 3. Обрахувати:

$$y_0 = a_0 + \frac{3-\sqrt{5}}{2}(b_0 - a_0); \quad z_0 = a_0 + b_0 - y_0, \quad \frac{3-\sqrt{5}}{2} = 0{,}38196.$$

Крок 4. Обрахувати $f(y_k), f(z_k)$.

Крок 5. Порівняти $f(y_k)$ з $f(z_k)$:

а) якщо $f(y_k) \le f(z_k)$, то присвоїти $a_{k+1} = a_k, b_{k+1} = z_k, y_{k+1} = a_{k+1} + b_{k+1} - y_k, z_{k+1} = y_k$. Перейти до кроку 6;

б) якщо $f(y_k) > f(z_k)$, то присвоїти $a_{k+1} = y_k, b_{k+1} = b_k, y_{k+1} = z_k, z_{k+1} = a_{k+1} + b_{k+1} - z_k$.

Крок 6. Обрахувати $\Delta = |a_{k+1} - b_{k+1}|$ та перевірити умову закінчення:

а) якщо $\Delta \le l$, процес пошуку завершується і $x^* \in [a_{k+1}, b_{k+1}]$. В якості наближеного рішення можна взяти середину останнього інтервалу: $x^* = \frac{a_{k+1} + b_{k+1}}{2}$;

б) якщо $\Delta > l$, присвоїти $k = k + 1$ та перейти до кроку 4.

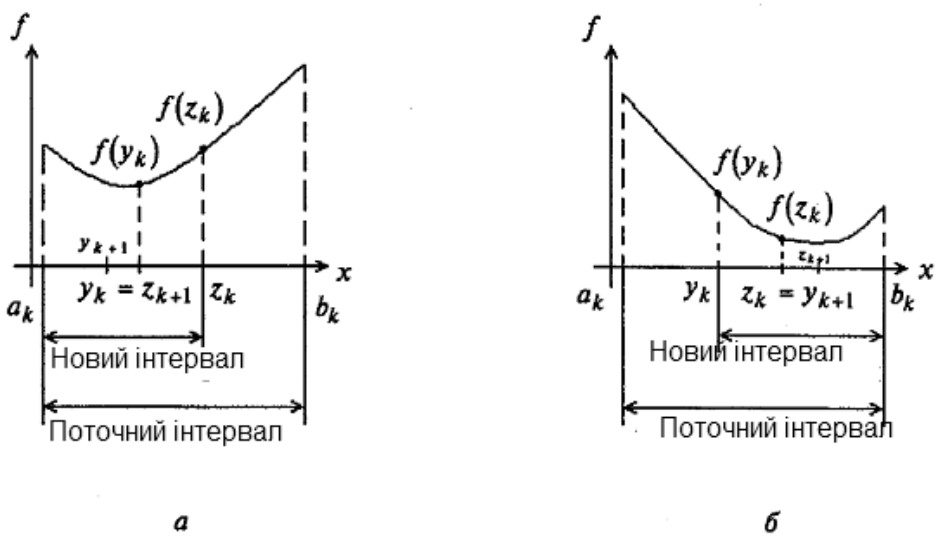

Рис.9.5 Ілюстрація методу «золотого січення»

Лекція 10
Наближені методи розв'язання звичайних диференціальних рівнянь

Часто доводиться зіштовхуватись з диференційними рівняннями і системами диференційних рівнянь при розробці нових виробів чи технологічних процесів, так як більша частина законів фізики формалізується саме у вигляді диференційних рівнянь. Будь-яка задача проектування в кінцевому рахунку зводиться до розв'язку диференційних рівнянь. Нажаль, лише дуже малу частину з них можливо вирішити без допомоги обчислювальних машин. Тому чисельні методи розв'язку

диференційних рівнянь відіграють важливу роль у практиці інженерних розрахунків.

10.1 Основні визначення та поняття

Рівняння, в якому невідома функція входить під знаком похідної чи диференціала, називається ***диференціальним рівнянням***. Наприклад,

$$\frac{dy}{dx} = 2(y-3); \quad \frac{d^2y}{dt^2} = t+1; \quad \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0,$$
$$y' = x^2; \quad xdy = y^3 dx$$

Якщо невідома функція, що входить у диференціальне рівняння, залежить тільки від однієї незалежної змінної, то диференціальне рівняння називається ***звичайним диференціальним рівнянням***. Наприклад, диференціальні рівняння

$$x^2 \cdot \frac{d^2y}{dx^2} = 2; \qquad 2sdt = tds\,.$$

відносяться до звичайних.

Якщо ж невідома функція, що входить у диференціальне рівняння, є функцією двох чи більшого числа незалежних змінних, то таке рівняння називається ***диференціальним рівнянням у частинних похідних***. Наприклад, диференціальне рівняння

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$$

відноситься до рівняння в частинних похідних.

Порядком диференціального рівняння називається найвищий порядок похідної (чи диференціала), що входить у рівняння.

Розглянемо звичайні диференціальні рівняння.

Звичайне диференціальне рівняння n-го порядку в загальному випадку містить незалежну змінну, невідому функцію і її похідні чи диференціали до n-го порядку включно і має вид

$$F(x, y, y', y'', ..., y^{(n)}) = 0\,. \qquad (10.1)$$

У цьому рівнянні x - незалежна змінна, y - невідома функція, $(y', y'', ..., y^{(n)})$ - похідні цієї функції.

Розв'язком (чи ***інтегралом***) рівняння (10.1) називається будь-яка диференціюєма функція $y = \varphi(x)$, що задовольняє цьому рівнянню, тобто така, після підстановки, якої у рівняння (10.1) воно перетворюється в тотожність.

Графік розв'язку звичайного диференціального рівняння називається ***інтегральною кривою*** цього ***рівняння***.

Розв'язок диференціального рівняння, що містить стільки незалежних довільних (постійних) параметрів, який його порядок, називається ***загальним розв'язком*** (чи ***загальним інтегралом***) цього ***рівняння***.

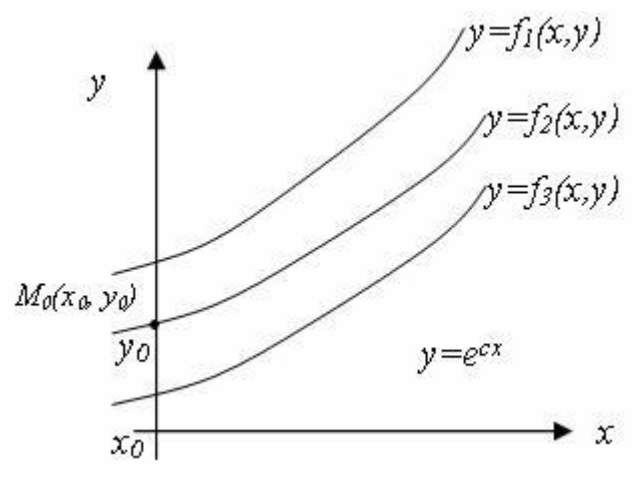

Рис. 10.1. Сімейство інтегральних кривих диференціального рівняння

Геометрично, загальний розв'язок диференціального рівняння являє собою сімейство інтегральних кривих рівняння (10.1) (рис. 10.1).

Частинним розв'язком диференціального рівняння називається будь-який розв'язок, що може бути отриманий з загального при визначених числових значеннях довільних постійних (рис. 10.1). Довільні постійні, що входять в загальний розв'язок, визначаються з початкових або крайових умов.

Задача з початковими умовами ставиться так: знайти розв'язок $y = \varphi(x)$ рівняння $y^{(n)} = f(x, y, y', y'', ..., y^{(n-1)})$, що задовольняє додатковим умовам, які складаються з того, що розв'язок $y = \varphi(x)$, повинний приймати разом зі своїми похідними до (n-1)-го порядку задані числові значення $y_0, y_0', y_0'', ..., y_0^{(n-1)}$ при заданому числовому значенні $x = x_0$ незалежної змінної x.

Такі умови називаються ***початковими умовами***, а задача відшукання розв‘язку $y = \varphi(x)$ диференціального рівняння (10.1), що задовольняє заданим початковим умовам - ***задачею з початковими умовами***, або ***задачею Коші***.

Задача з крайовими умовами ставиться так: знайти розв'язок $y = \varphi(x)$ рівняння $y^{(n)} = f(x, y, y', y'', ..., y^{(n-1)})$, що задовольняє додатковим умовам, які складаються з того, що розв'язок $y = \varphi(x)$, повинний приймати разом зі своїми похідними до (n-1)-го порядку задані числові значення $y_0, y_0', y_0'', ..., y_0^{(n-1)}$ при заданому числовому значенні $x = x_0$ та $y_n, y_n', y_n'', ..., y_n^{(n-1)}$ при заданому числовому значені $x = x_n$ незалежної змінної x.

Такі умови називаються ***крайовими умовами***, а задача відшукання розв‘язку $y = \varphi(x)$ диференціального рівняння (10.1), що задовольняє заданим крайовим умовам – ***крайовою задачею***.

У випадку рівняння першого порядку, тобто при n=1, одержуємо задачу Коші для рівняння $y' = f(x, y)$ з початковою умовою $x = x_0$, $y = y_0$.

Геометрично задача Коші для рівняння першого порядку полягає в тому, що з усіх інтегральних кривих, що представляють собою загальний розв‘язок, потрібно знайти ту інтегральну криву, що проходить через точку M_0 з координатами $x = x_0$, $y = y_0$ (рис.10.1).

Часто в задачі Коші у ролі незалежної змінної виступає час t. Прикладом може бути задача про вільні коливання тіла, яке підвішене на пружині. Рухи такого тіла описуються диференційним рівнянням, в якому незалежною змінною є час t. Якщо додаткові умови задані у вигляді значень переміщень чи швидкості при t=0, то це також задача Коші.

Задача Коші має єдиний розв‘язок, що задовольняє умові в $(x_0) = y_0$, якщо функція $f(x, y)$ неперервна в деякій області $R_{[a,b]} = \{|x - x_0| < a, \ |y - y_0| < b\}$ і задовольняє в цій області умові Ліпшица:

$$|f(x, \bar{y}) - f(x, y)| \le N|\bar{y} - y|,$$

де N - постійна Ліпшица, що залежить від a і b (a і b - межі області).

Методи точного інтегрування диференціальних рівнянь придатні лише для порівняно невеликої частини рівнянь, що зустрічаються на практиці.

Тому в задачах моделювання та дослідження складних технічних систем, наприклад, систем автоматичного управління, великого значення набувають методи наближеного розв’язання диференційних рівнянь, що в залежності від форми представлення розв’язку можна розділити на дві групи:

1) ***аналітичні методи***, що дають наближений розв’язок диференційного рівняння у виді аналітичного виразу;

2) ***чисельні методи***, що дають наближений розв’язок у вигляді таблиці.

Похибки

Перед тим, як перейти до розглядання методів чисельного розв'язання диференційних рівнянь, зупинимося на джерелах похибок, пов'язаних з чисельною апроксимацією. Таких джерел три:

1. ***Похибка заокруглення*** зумовлена обмеженнями на представлення чисел в ЕОМ, тому що число значущих цифр, що запам'ятовується і використовується в обчисленнях, обмежене.

2. ***Похибка відсічення*** пов'язана з тим, що для апроксимації функції замість

$$y = y_0,\ y' = y'_0,\ y'' = y''_0,\ldots, y^{(n-1)} = y_0^{(n-1)} \quad \text{при} \quad x = x_0 \qquad (10.2)$$

нескінчених рядів часто використовується лише декілька перших їх членів.

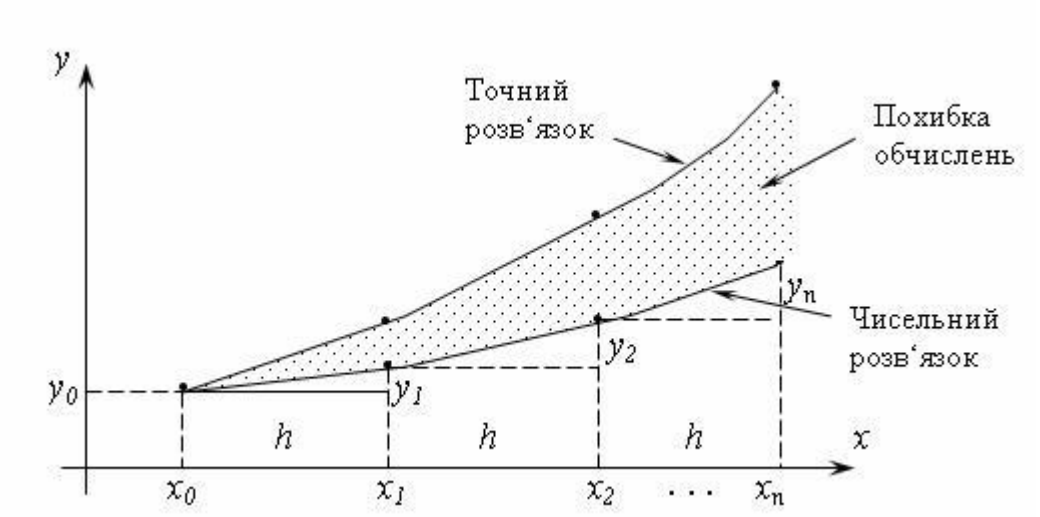

Рис. 10.2. Геометричне представлення накопичування похибки в процесі обчислень

3. ***Похибка поширення*** являється результатом накопичення похибок, що з'явились у попередніх результатах розрахунку. Так як ні один з наближених методів не може дати зовсім точних результатів, то будь-яка похибка, яка виникла в процесі обчислень, зберігається і на наступних стадіях розрахунку (рис. 10.2).

Вказані три джерела похибок є причиною помилок двох типів:

Локальна помилка – сума похибок, що вносяться у розрахунковий процес на кожному етапі обчислення.

Глобальна помилка – різниця між розрахованим та точним значеннями величини на кожному етапі реалізації чисельного алгоритму, що визначає сумарну похибку, що накопичується з моменту початку розрахунку.

10.2. Класифікація методів розв'язання задачі Коші

На протязі багатьох років чисельний розв‘язок задачі Коші був об‘єктом пильної уваги науковців, оскільки він широко застосовується в різних галузях науки і техніки. Тому і кількість розроблених для нього методів дуже велика.

Чисельні методи розв'язання задачі Коші розділяються на 3 групи:

- одноточкові;
- багатоточкові (методи прогнозу та корекції);
- методи з автоматичним вибором кроку інтегрування.

На рис. 10.3 представлена класифікація найбільш відомих чисельних методів розв‘язання диференційних рівнянь (ДР) на ЕОМ.

Рис. 10.3. Класифікація чисельних методів розв'язання задачі Коші

До одноточкових методів відносять методи, які мають певні загальні риси, такі як:

1. В основі усіх одноточкових методів лежить розклад функції в ряд Тейлора, в якому зберігаються члени, що мають h в степені до k включно. Ціле число k називається ***порядком метода***. Похибка на кроці має порядок k+1.

2. Всі одноточкові методи не потребують дійсного обчислення похідних, тому що обчислюється лише сама функція, однак можуть потребуватися її значення в деяких проміжних точках. Це тягне за собою, звичайно, додаткові затрати часу і зусиль.

3. Для отримання інформації у новій точці, потрібно мати дані лише в попередній точці. Цю властивість можна назвати „самостартуваням". Властивість „самостартуваня" дозволяє легко змінювати величину кроку h.

В порівнянні з одноточковими методами методи прогнозу і корекції мають ряд особливостей:

1. Для реалізації методів прогнозу і корекції необхідно мати інформацію про декілька попередніх точок (вони не відносяться до „самостартуючих" методів), тому для отримання додаткової інформації доводиться застосовувати одноточковий метод.

2. Одноточкові методи і методи прогнозу і корекції забезпечують приблизно однакову точність результатів. Однак другі на відміну від перших дозволяють лише оцінити похибку на кроці. З цієї причини, користуючись одноточковими методами, величину кроку h звичайно обирають трохи менше, ніж це необхідно, тому методи прогнозу і корекції виявляються найбільш ефективними.

Використовуючи метод Рунге-Кутта четвертого порядку точності, на кожному кроці доводиться обчислювати чотири значення функції, але для збіжності методу прогнозу і корекції того ж порядку точності часто достатньо двох значень функції. Тому методи прогнозу і корекції вимагають майже вдвічі менше машинного часу, ніж методи Рунге-Кутта порівнюваної точності.

10.3 Одноточкові методи розв'язання задачі Коші

Розв'язати диференційне рівняння $y' = f(x, y)$ чисельним методом - це значить для заданої послідовності аргументів $x_0, x_1, ..., x_n$ і y_0 знайти такі значення $y_0, y_1, ..., y_n$, що $y_i = F(x_i),\ i = 1,2,...,n$ та $F(x_0) = y_0$. Таким чином, чисельні методи дозволяють замість функції $y = F(x)$ одержати таблицю значень цієї функції для заданої послідовності аргументів. Величина $h = x_k - x_{k-1}$ називається ***кроком інтегрування***.

Графічно чисельний розв‘язок являє собою послідовність коротких прямолінійних відрізків, якими апроксимується аналітичний розв’язок $y = F(x)$ рівняння (кусково-лінійна апроксимація).

Розглянемо алгоритми найбільш відомих чисельних методів.

Метод Ейлера

Метод Ейлера є порівняно грубим і застосовується в основному для орієнтованих розрахунків. Однак ідеї, покладені в основу методу Ейлера, є базовими для інших методів.

Нехай дано диференційне рівняння першого порядку

$$y' = f(x, y) \tag{10.3}$$

з початковими умовами

$$x = x_0, \quad y(x_0) = y_0. \tag{10.4}$$

Необхідно знайти розв’язок рівняння на відрізку $[x_0, x_n]$.

Розіб'ємо відрізок $[x_0, x_n]$ на n рівних частин і одержимо послідовність $x_0, x_1, x_2, ..., x_n$, де $x_i = x_0 + ih\ (i = 0,1,2,...,n)$, а $h = \dfrac{x_n - x_0}{n}$ - крок інтегрування.

Виберемо k-й відрізок $[x_k, x_{k+1}]$ і проінтегруємо рівняння (10.3):

$$\int_{x_k}^{x_{k+1}} f(x, y)dx = \int_{x_k}^{x_{k+1}} y'dx = y(x)\Big|_{x_k}^{x_{k+1}} = y(x_{k+1}) - y(x_k) = y_{k+1} - y_k$$

або

$$y_{k+1} = y_k + \int_{x_k}^{x_{k+1}} f(x, y)dx. \tag{10.5}$$

Якщо в останньому інтегралі підінтегральну функцію на відрізку $[x_k, x_{k+1}]$ прийняти постійною і рівною початковому значенню в точці $x = x_k$, то одержимо

$$\int_{x_k}^{x_{k+1}} f(x_k, y_k)dx = f(x_k, y_k) \cdot x \Big|_{x_k}^{x_{k+1}} = f(x_k, y_k)(x_{k+1} - x_k) = y'_k h .$$

Тоді формула (10.5) прийме вигляд

$$y_{k+1} = y_k + y'_k h . \tag{10.6}$$

Позначивши $y_{k+1} - y_k = \Delta y_k$, отримаємо:

$$y'_k h = \Delta y_k . \tag{10.7}$$

Продовжуючи цей процес, і щоразу приймаючи, що на відрізку $[x_k, x_{k+1}]$ інтегральна крива $y = F(x)$ приблизно заміняється прямолінійним відрізком, що виходить із точки $M_k = (x_k, y_k)$ кутовим коефіцієнтом $f(x_k, y_k)$. Тому в якості наближення шуканої інтегральної кривої одержуємо ламану лінію з вершинами в точках $M_o = (x_o, y_o), M_1 = (x_1, y_1), ..., M_n = (x_n, y_n)$ (рис. 10.4).

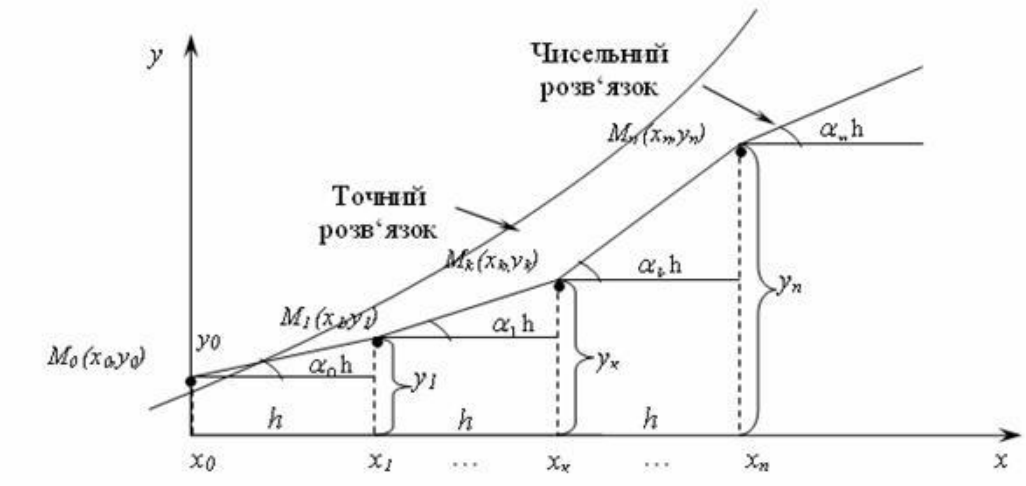

Рис. 10.4. Геометрична інтерпретація методу Ейлера

Якщо функція $f(x, y)$ у деякій прямокутній області $R\{|x - x_0| \le a, |y - y_0| \le b\}$ задовольняє умові:

$$|f(x, \bar{y}) - f(x, y)| \le N|\bar{y} - y| , \tag{10.8}$$

і, крім того,

$$\left|\frac{df}{dx}\right| = \left|\frac{df}{dx} + f\,\frac{df}{dy}\right| < M, \quad M = \text{const} , \tag{10.9}$$

то має місце наступна оцінка похибки:

$$\left|y(x_n)-y_n\right| \le \frac{hM}{2N}[(1+hN)^n-1], \qquad (10.10)$$

де $y(x_n)$ - значення точного розв'язку рівняння при $x = x_n$, а y_n - наближене значення, отримане на n-у кроці.

Формула (10.10) має в основному теоретичне застосування. На практиці, як правило, застосовують "подвійний прорахунок". Спочатку чисельне розв'язання рівняння ведеться з кроком h, потім крок дроблять і повторний розрахунок ведеться з кроком h/2. Похибка більш точного значення оцінюється формулою

$$\left|y_n^* - y(x_n)\right| \approx \left|y_n^* - y_n\right|. \qquad (10.11)$$

Метод Ейлера може бути застосований до розв'язку систем диференційних рівнянь вищих порядків. Однак в останньому випадку диференційні рівняння повинні бути приведені до системи диференційних рівнянь першого порядку.

Нехай задана система двох рівнянь першого порядку

$$\begin{cases} y' = f_1(x,y,z) \\ z' = f_2(x,y,z) \end{cases}, \qquad (10.12)$$

з початковими умовами

$$y(x_0) = y_0, \quad z(x_0) = z_0. \qquad (10.13)$$

Наближені значення $y(x_i) = y_i$ та $z(x_i) = z_i$ знаходяться по формулах:

$$\begin{cases} y_{i+1} = y_i + \Delta y_i \\ z_{i+1} = z_i + \Delta z_i \end{cases}, \qquad (10.14)$$
$$\Delta y_i = hf_1(x_i, y_i, z_i), \quad \Delta z_i = hf_2(x_i, y_i, z_i), \quad i = 0,1,2,...$$

Модифікації методу Ейлера

З метою підвищення точності методу Ейлера використовують різні його модифікації.

Суть **удосконаленого методу Ейлера** полягає в використанні ітераційної формули виду:

$$y_{i+1}^{(0)} = y_i + hf\left(x_{i+\frac{1}{2}}, y_{i+\frac{1}{2}}\right), \qquad (10.15)$$

де $x_{i+\frac{1}{2}}$ - значення аргументу x в точці $(x_i + \frac{h}{2})$, а $y_{i+\frac{1}{2}}$ - значення функції в точці $(x_i + \frac{h}{2})$.

Розглянемо диференціальне рівняння $y' = f(x, y)$ з початковою умовою $y(x_0) = y_0$. Необхідно знайти розв'язок рівняння на відрізку $[a, b]$.

Розіб'ємо відрізок $[a, b]$ на n рівних частин точками $x_i = x_0 + ih$ $(i = 0,1,2,..., n)$, де $h = \frac{b-a}{n}$.

Алгоритм методу складається з:

1. визначення похідної y'_0 в точці (x_0, y_0): $y'_0 = f(x_0, y_0)$;

2. змінна незалежної змінної x за формулою: $x_{0+\frac{1}{2}} = x_0 + \frac{h}{2}$;

3. визначення значення $y_{0+\frac{1}{2}}$ при $x_{0+\frac{1}{2}}$: $y_{0+\frac{1}{2}} = y_0 + \frac{h}{2} y'_0$;

4. визначення похідної в точці $(x_{0+\frac{1}{2}}, y_{0+\frac{1}{2}})$: $y'_{0+\frac{1}{2}} = f(x_{0+\frac{1}{2}}, y_{0+\frac{1}{2}})$;

5. використовуємо отримане значення $y'_{0+\frac{1}{2}}$ для визначення y_1 за формулою:

$y_1 = y_0 + hy'_{0+\frac{1}{2}} = y_0 + hf(x_{0+\frac{1}{2}}, y_{0+\frac{1}{2}})$;

6. змінюємо $x_1 = x_{0+\frac{1}{2}} + \frac{h}{2}$;

7. повторюємо всі кроки алгоритму, починаючи з першого.

Графічна інтерпретація методу представлена на рис. 10.5.

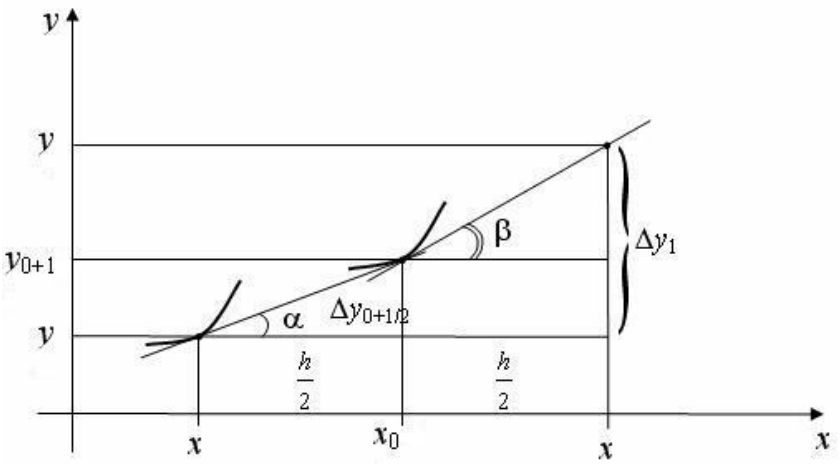

Рис. 10.5. Графічна інтерпретація удосконаленого методу Ейлера

Зауваження. Оцінка похибки в точці x_i може бути отримана за допомогою "подвійного прорахунку": розрахунок повторюють із кроком h/2. Похибку більш точного значення (при кроці h/2) оцінюють в такий спосіб:

$$\left|y_i^* - y(x_i)\right| \approx \frac{1}{3}\left|y_i^* - y_i\right|,$$

де $y(x_i)$ - точний розв'язок диференціального рівняння. Удосконалений метод Ейлера є більш точним у порівнянні з методом Ейлера та відноситься до методів 3-го порядку точності.

Модифікований метод Ейлера заснований на використанні ітераційної формули виду:

$$y_{i+1} = y_i + \frac{h}{2}[f(x_i, y_i) + f(x_{i+1}, y_{i+1})]. \qquad (10.16)$$

Геометрична інтерпретація представлена на рис. 10.6.

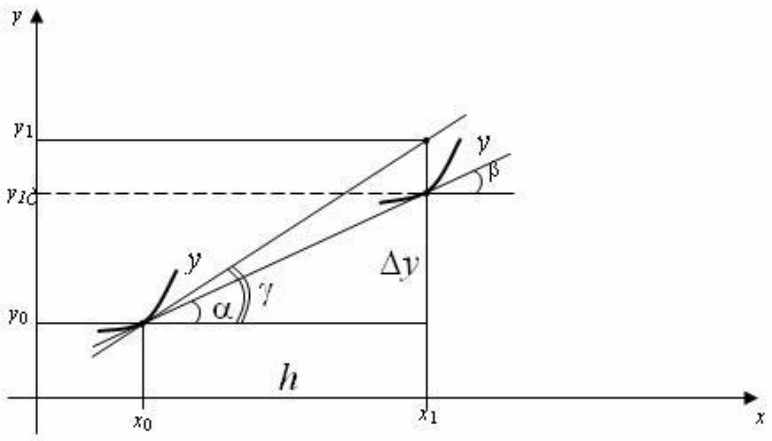

Рис.10.6. Графічна інтерпретація модифікованого методу Ейлера

Алгоритм методу включає наступні кроки:

1. визначення похідної y_0' в точці (x_0, y_0): $y_0' = f(x_0, y_0)$;

2. зміна незалежної змінної x за формулою: $x_1 = x_0 + h$;

3. визначення допоміжного значення y_{1d} за формулою методу Ейлера: $y_{1d} = y_0 + hy_0'$;

4. визначення допоміжної похідної в точці (x_1, y_{1d}): $y_{1d}' = f(x_1, y_{1d})$;

5. визначення середньо арифметичного значення двох похідних: $\Delta y = \frac{y_0' + y_{1d}'}{2}$;

6. визначення y_1 за формулою: $y_1 = y_0 + h\Delta y = \frac{h}{2}(y'_0 + y'_{1d})$.

7. ітераційний процес повторюється, починаючи з першого кроку.

Метод Рунге–Кутта

Метод Рунге-Кутта є одним з методів підвищеної точності, але має багато загального з методом Ейлера.

Нехай на відрізку $[a,b]$ необхідно знайти чисельний розв'язок рівняння $y' = f(x,y)$ з початковою умовою $y(x_0) = y_0$.

В методі Рунге-Кутта, аналогічно методу Ейлера, послідовні значення y_i шуканої функції визначаються за формулою

$$y_{i+1} = y_i + \Delta y_i .$$

Якщо розкласти функцію у ряд Тейлора й обмежитися членами до h^4 включно, то збільшення функції (Δy) можна представити у вигляді:

$$\Delta y = y(x+h) - y(x) = hy'(x) + \frac{h^2}{2}y''(x) + \frac{h^3}{6}y'''(x) + \frac{h^4}{24}y^{(4)}(x), \quad (10.17)$$

де похідні $y''(x), y'''(x), y^{(4)}(x)$ визначаються послідовним диференціюванням.

Замість безпосередніх обчислень по формулі (10.17) у методі Рунге-Кутта визначаються чотири числа:

$$\begin{aligned} k_1 &= hf(x,y), \\ k_2 &= hf\left(x+\frac{h}{2}, y+\frac{k_1}{2}\right), \\ k_3 &= hf\left(x+\frac{h}{2}, y+\frac{k_2}{2}\right), \\ k_4 &= hf(x+h, y+k_3). \end{aligned} \quad (10.18)$$

Можна довести, що якщо числам k_1, k_2, k_3, k_4 додати відповідно вагу $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}$, то середньозважене цих чисел, тобто $\frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4$ з точністю до четвертих ступенів дорівнює значенню y, обчисленому по формулі (10.17):

$$\Delta y = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \quad (10.19)$$

Таким чином, для кожної пари поточних значень x_i та y_i по формулах (10.18) визначаються значення:

$$
\begin{aligned}
k_i^{(i)} &= hf(x_i, y_i),\\
k_2^{(i)} &= hf(x_i + \frac{h}{2}, y_i + \frac{k_i^{(i)}}{2}),\\
k_3^{(i)} &= hf(x_i + \frac{h}{2}, y_i + \frac{k_2^{(i)}}{2}),\\
k_4^{(i)} &= hf(x_i + h, y_i + k_3^{(i)}).
\end{aligned}
\qquad (10.20)
$$

і потім

$$y_{i+1} = y_i + \Delta y_i .$$

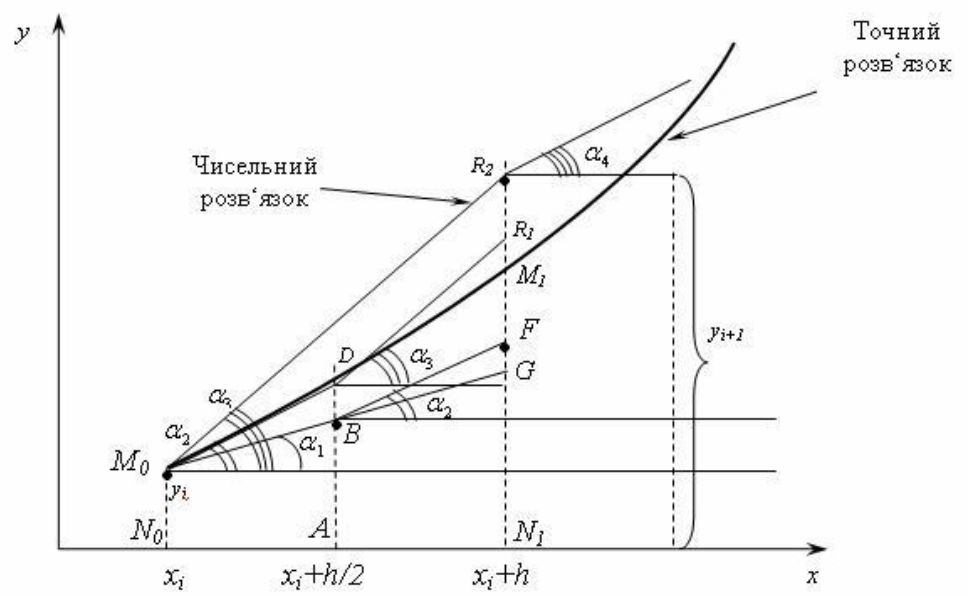

Рис.10.7. Геометрична інтерпретація метода Рунге-Кутта

Обчислення по методу Рунге-Кутта зручно розташовувати за алгоритмом:

1. значення x_0 та y_0 підставляють у праву частину диференціального рівняння, визначають $f(x, y)$;

2. отримане значення $f(x, y)$ множать на крок інтегрування h, обчислюють $k_1 = hf(x, y)$;

3. змінюють значення x_0: $x_{0+\frac{1}{2}} = x_0 + \frac{h}{2}$;

4. визначають допоміжне значення y_{0d}: $y_{0d} = y_0 + \frac{k_1}{2}$;

5. визначення похідної в точці $(x_{0+\frac{1}{2}}, y_{0d})$: $y'_{0d} = f(x_{0+\frac{1}{2}}, y_{0d})$;

6. визначають значення k_2: $k_2 = hf(x_{0+\frac{1}{2}}, y_{0d})$;

7. визначають нове допоміжне значення $\dot{y}_{0d}$: $\dot{y}_{0d} = y_0 + \frac{k_2}{2}$;

8. визначення похідної в точці $(x_{0+\frac{1}{2}}, \dot{y}_{0d})$: $\dot{y}'_{0d} = f(x_{0+\frac{1}{2}}, \dot{y}_{0d})$;

9. визначають значення k_3: $k_3 = hf(x_{0+\frac{1}{2}}, \dot{y}_{0d})$;

10. визначають нове значення допоміжного y_{1d}: $y_{1d} = y_0 + k_3$;

11. змінюють значення $x_{0+\frac{1}{2}}$: $x_1 = x_{0+\frac{1}{2}} + \frac{h}{2}$;

12. визначають допоміжну похідну в точці (x_1, y_{1d}): $y'_{1d} = f(x_1, y_{1d})$;

13. визначають значення k_4: $k_4 = h \cdot y'_{1d} = h \cdot f(x_0 + h, y_0 + k_3)$

14. визначають нове значення y_1 за формулою: $y_1 = y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$.

Для визначення $y_2, y_3, ..., y_n$ повторюють ітераційний процес, починаючи з першого кроку, поки не буде пройдений весь відрізок $[a, b]$.

Метод Рунге - Кутта має порядок точності h^4 на усьому відрізку $[a, b]$. Оцінка точності цього методу дуже складна. Грубу оцінку погрішності можна одержати за допомогою "подвійного прорахунку" по формулі:

$$\left| y_i^* - y(x_i) \right| \approx \frac{y_i^* - y_i}{15},$$

де $y(x_i)$ - значення точного розв'язку рівняння у точці x_i, а y_i^* та y_i - наближені значення, отримані з кроком h/2 і h.

Нехай задана система диференціальних рівнянь першого порядку:

$$\begin{cases} y' = f(x, y, z) \\ z' = g(x, y, z) \end{cases}.$$

У цьому випадку паралельно визначаються числа Δy_i та Δz_i:

$$\Delta y_i = \frac{1}{6}(k_1^{(i)} + 2k_2^{(i)} + 2k_3^{(i)} + k_4^{(i)}), \qquad \Delta z_i = \frac{1}{6}(l_1^{(i)} + 2l_2^{(i)} + 2l_3^{(i)} + l_4^{(i)}),$$

$$k_i^{(i)} = hf(x_i, y_i, z_i),$$
$$l_i^{(i)} = hg(x_i, y_i, z_i);$$

$$k_2^{(i)} = hf(x_i + \frac{h}{2}, y_i + \frac{k_i^{(i)}}{2}, z_i + \frac{l_1^{(i)}}{2}),$$
$$l_2^{(i)} = hg(x_i + \frac{h}{2}, y_i + \frac{k_i^{(i)}}{2}, z_i + \frac{l_1^{(i)}}{2});$$

$$k_3^{(i)} = hf(x_i + \frac{h}{2}, y_i + \frac{k_2^{(i)}}{2}, z_i + \frac{l_2^{(i)}}{2}),$$
$$l_3^{(i)} = hg(x_i + \frac{h}{2}, y_i + \frac{k_2^{(i)}}{2}, z_i + \frac{l_2^{(i)}}{2});$$

$$k_4^{(i)} = hf(x_i + h, y_i + k_3^{(i)}, z_i + l_3^{(i)}),$$
$$l_4^{(i)} = hg(x_i + h, y_i + k_3^{(i)}, z_i + l_3^{(i)}).$$

10.4 Методи прогнозу і корекції (багатоточкові методи)

В методах прогнозу і корекції для обчислення значення нової точки розв‘язку ДР використовується інформація про декілька раніше отриманих точок відрізку дослідження. Для цього використовуються дві формули, що називаються відповідно формулами прогнозу і корекції. Розглянемо особливості алгоритмів методів прогнозу і корекції.

Так як в методах, що розглядаються використовується інформація про декілька раніше отриманих точок, то на відміну від однокрокових методів вони не володіють властивістю „самостартування”. Тому, перед тим як застосовувати метод прогнозу і корекції, необхідно обчислювати вихідні данні за допомогою будь-якого однокрокового методу. Часто для цього використовують метод Рунге-Кутта. Обчислення проводять наступним чином. Спочатку по формулі прогнозу і початковим значенням змінних знаходять значення $y_{n+1}^{(0)}$. Верхній індекс означає, що прогнозоване значення є одним з послідовності значень y_{n+1}, що розташовані в порядку зростання точності. По прогнозованому значенню $y_{n+1}^{(0)}$ за допомогою приведеного вище диференціального рівняння знаходять похідну $y'^{(0)}_{n+1} = f(x_{n+1}, y_{n+1}^{(0)})$, яка потім підставляється у формулу корекції для обчислення уточненого значення $y_{n+1}^{(j+1)}$.

В свою чергу $y_{n+1}^{(j+1)}$ використовується для отримання більш точного значення похідної за допомогою диференціального рівняння

$$y'^{(j+1)}_{n+1} = f(x_{n+1}, y^{(j+1)}_{n+1}).$$

Якщо це значення похідної недостатньо близьке до попереднього, то воно вводиться у формулу корекції й ітераційний процес продовжується. Якщо ж похідна змінюється в допустимих границях, то значення $y'^{(j+1)}_{n+1}$ використовується для обчислення остаточного значення y_{n+1}. Після цього процес повторюється – здійснюється наступний крок, на якому обчислюється y_{n+2}.

Звичайно при вводі формул прогнозу і корекції розв'язок рівняння розглядають як процес наближеного інтегрування, а самі формули отримують за допомогою кінцево-різницевих методів.

Якщо диференціальне рівняння $y'(x) = f(x, y)$ проінтегровано в інтервалі $[x_n, x_{n+k}]$, то результат прийме вигляд:

$$y(x_{n+k}) - y(x_n) = \int_{x_n}^{x_{n+k}} f(x, y)dx.$$

Цей інтеграл не можна обчислити безпосередньо, так як залежність $y(x)$ невідома. Наближене значення інтегралу можна знайти за допомогою одного з кінцево-різницевих методів. Вибір методу і буде визначати метод розв'язку диференціальних рівнянь. На етапі прогнозу можна використовувати будь-яку формулу чисельного інтегрування, якщо до неї не входить попереднє значення $y'(x_{n+1})$.

Метод Мілна

В цьому методі на етапі прогнозу використовується формула Мілна

$$y_{n+1} = y_{n-3} + \frac{4}{3}h(2y'_n - y'_{n-1} + 2y'_{n-2}) + \frac{28}{90}h^5 y^{(5)},$$

а на етапі корекції - формула Сімпсона

$$y_{n+1} = y_{n-1} + \frac{1}{3}h(y'_{n+1} + 4y'_n + y'_{n-1}) - \frac{1}{90}h^5 y^{(5)}.$$

Останні члени в обох формулах в дійсності в ітераційному процесі не використовуються і слугують лише для оцінки помилки відсічення. Метод Мілна

відносять до методів четвертого порядку точності, так як в ньому відкидаються члени, які містять h в п'ятій степені і більш високих степенях. Похибка відсічення при корекції в 28 разів менше і тому представляє великий інтерес. Незважаючи на те що формула Мілна містить менший числовий коефіцієнт (1/90) перед членом, що відкидається, її використовують рідше, ніж інші (з більшими відкидуваними членами), так як їй притаманна нестійкість. Це означає, що похибка поширення може рости експоненціально, при чому цей висновок справедливий для всіх формул корекції, основаних на правилі Сімпсона.

Метод Адамса - Башфорта

Цей метод також має четвертий порядок точності. Формула, що використовується в ньому отримана інтегруванням оберненої інтерполяційної формули Ньютона і має вид:

$$y_{n+1} = y_n + \frac{1}{24}h(55y'_n - 59y'_{n-1} + 37y'_{n-2} - 9y'_{n-3}) + \frac{251}{720}h^5 y^{(5)},$$

а на етапі корекції використовується формула:

$$y_{n+1} = y_n + \frac{1}{24}h(9y'_{n+1} - 19y'_n - 37y'_{n-1} - y'_{n-2}) + \frac{19}{720}h^5 y^{(5)}.$$

Розрахунки по методу Адамса - Башфорта виконуються так, як і по методу Мілна, але на відміну від останнього похибка, внесена на якому-небудь кроці, не має тенденції до експоненціального росту.

Метод Хемінга

В методі Хемінга використовуються наступні формули уточнення прогнозу:

$$y^{(0)}_{n+1} = y_{n-3} + \frac{4}{3}h(2y'_n - y'_{n-1} + 2y'_{n-2})$$

$$\bar{y}^{(0)}_{n+1} = y^{(0)}_{n+1} + \frac{112}{121}(y_n - y^{(0)}_n),$$

$$[\bar{y}^{(0)}_{n+1}] = f(x_{n+1}, \bar{y}^{(0)}_{n+1})$$

та корекції

$$y^{(j+1)}_{n+1} = \frac{1}{8}(9y_n - y_{n-2}) + \frac{3}{8}h([\bar{y}^{j+1}_{n+1}]' + 2y'_n - y'_{n-1}).$$

Це стійкий метод четвертого порядку точності, в основі якого лежать наступні формули прогнозу:

$$y_{n+1} = y_{n-3} + \frac{4}{3}h(2y'_n - y'_{n-1} + 2y'_{n-2}) + \frac{28}{90}h^5 y^{(5)}$$

і корекції

$$y_{n+1} = \frac{1}{8}[9y_n - y_{n-2} + 3h(y'_{n+1} + 2y'_n - y'_{n-1})] - \frac{1}{40}h^5 y^{(5)}.$$

Особливістю методу Хемінга є те, що він дозволяє оцінювати похибки, що вносяться на стадіях прогнозу і корекції і усувати їх. Завдяки простоті та стійкості цей метод є одним з найбільш поширених методів прогнозу і корекції.