МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет будівництва і архітектури

О.В. Горда

ЧИСЕЛЬНІ МЕТОДИ

РОЗВ'ЯЗАННЯ НЕЛІНІЙНИХ РІВНЯНЬ ТА СИСТЕМ РІВНЯНЬ

Конспект лекцій

для студентів, які навчаються за напрямом підготовки 6.050101
"Комп'ютерні науки"

Київ 2010

УДК 51:004
ББК 32.811
Г68

Рецензент В.М. Михайленко, доктор техн. наук, професор

Затверджено на засіданні вченої ради факультету автоматизації та інформаційних технологій, протокол № 1 від 9 вересня 2009 року.

Горда О.В.

Г68 Чисельні методи. Розв’язання нелінійних рівнянь та систем рівнянь: конспект лекцій / О.В. Горда. – К.: КНУБА, 2010. – 72 с.

Розглянуто основні формули чисельного диференціювання та інтегрування та особливості їх застосування, а також методи розв’язання нелінійних рівнянь та систем лінійних та нелінійних рівнянь та умови їх використання та адаптації до конкретних практичних задач. Теоретичний матеріал проілюстрований детально опрацьованими прикладами.

Призначено для студентів, які навчаються за напрямом підготовки 6.050101 "Комп’ютерні науки".

УДК 51:004
ББК 32.811

ВСТУП

В процесі дослідження реальних явищ та об’єктів при побудові математичних моделей та їх дослідженні необхідно обчислювати похідні від складних функцій, що потребує знання апарату чисельного диференціювання. Вміння обчислювати значення похідних є необхідним при розв’язанні диференційних рівнянь, за допомогою яких описується більшість динамічних систем. Основний метод розв’язку диференційних рівнянь – інтегрування, для якого також застосовують спеціальні чисельні методи.

Також значна кількість задач, що виникають під час проведення науково-дослідних робіт, пов’язана з необхідністю розв’язку нелінійних рівнянь та систем лінійних та нелінійних рівнянь.

В другій частині конспекту лекцій з дисципліни «Чисельні методи» розглядаються питання, що пов’язані з вищезазначеними задачами і відповідають матеріалу другого модуля. В результаті вивчення чисельних методів зазначеного розділу студенти повинні вміти класифікувати рівняння та функції, знати переваги, недоліки та умови застосування чисельних методів, правильно та обґрунтовано підбирати їх для конкретних практичних задач, при необхідності вміти адаптувати задачі для можливості застосування того чи іншого методу.

Для програмної реалізації того чи іншого методу студенти повинні вміти будувати їх алгоритми та розуміти особливості комп’ютерної реалізації.

Лекція 9. Чисельне диференціювання

Сіткові функції

Часто в чисельних методах функція від неперервного аргументу заміщується функцією від дискретного аргументу – сітковою функцією.

Розіб‘ємо відрізок $[a;b]$, на якому розглядається функція $f(x)$, на n відрізків довжиною h, де $h=(b-a)/n$ (покриємо сіткою). Точки розбиття пронумеруємо зліва направо, причому $y_0 = f(a)$ і

$y_n = f(b)$ (рис. 1). Будемо задавати нашу неперервну функцію $f(x)$ множиною точок $(x_i; y_i)$.

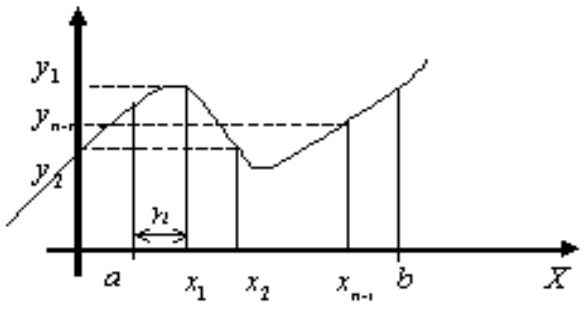


Рис. 1. Покриття відрізку диференціювання сіткою

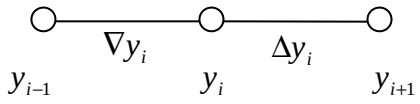
Сіткову функцію можна розглядати як функцію від цілочислового аргументу:

$$y(i) = y_i,\ i \in Z.$$

Для сіткової функції вводиться поняття різниць. Так, різниці першого порядку задаються наступним чином:

$\Delta y_i = y_{i+1} - y_i$ – права різниця;

$\nabla y_i = y_i - y_{i-1}$ – ліва різниця;



$$\delta y_i = \frac{1}{2}(\Delta y_i + \nabla y_i) = \frac{1}{2}(y_{i+1} - y_{i-1}) \text{ – центральна різниця.}$$

Формули чисельного диференціювання

Поліноміальні формули

Задачу чисельного диференціювання можна поставити наступним чином: значення функції $y = f(x)$, виміряні у рівновіддалених точках $x_k = x_0 + kh$, $y_k = f(x_k)$, де $k = 0, \pm 1, \pm 2, \ldots$ (функція задана своєю сіткою), необхідно обчислити значення похідної $y' = f'(x)$ в тих самих точках. Іншими словами, за табличними значеннями функції з постійним кроком h необхідно скласти таблицю її похідних з тим самим кроком.

Нехай в деякій точці x існує похідна функції $f(x)$

$$y' = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Взявши в ролі Δx мале значення, можна наближено обчислити значення цієї похідної:

$$y' \approx \frac{f(x+\Delta x)-f(x)}{\Delta x}.$$

Для чисельного визначення похідної заданої функції $f(x)$ її можна наблизити іншою функцією $\varphi(x)$, яка легко обчислюється. В ролі такої функції $\varphi(x)$ можна взяти інтерполяційні поліноми. Так, найпростішим наближенням буде інтерполяційний поліном Ньютона:

$$\varphi(x) = y(x_0)+(x-x_0)y(x_0,x_1)+(x-x_0)(x-x_1)y(x_0,x_1,x_2)+\ldots =$$

$$= y_0 + p_0\Delta y_0 + p_0 p_1 \Delta^2 y_0 + \ldots ,$$

де $p_i = x - x_i$, $i = 0,1,2,\ldots$.

Візьмемо похідну від цього поліному і отримаємо:

$$\varphi'(x) = \Delta y_0 + (p_0 + p_1)\Delta^2 y_0 + (p_0 p_1 + p_0 p_2 + p_1 p_2)\Delta^3 y_0 + \ldots .$$

Загальна формула k-ї похідної буде мати вигляд:

$$\varphi^{(k)}(x) = k!\left(\Delta^k y + \left(\sum_{i=0}^{k} p_i \right)\Delta^{k+1} y + \left(\sum_{i>j\geq 0}^{k+1} p_i p_j \right)\Delta^{k+2} y + \left(\sum_{i>j>l\geq 00}^{k+2} p_i p_j p_l \right)\Delta^{k+3} y + \ldots \right).$$

Оцінка точності формули за сіткою з $n+1$ вузлами:

$$R_n^{(k)} = \frac{M_{n+1}}{(n\_1-k)!}\max_i \xi_i^{n+1-k},\ M_{n+1} = \max\left| y^{(n+1)} \right|.$$

Якщо сітка рівномірна (має постійний крок), то

$$R_n^{(k)} < M_{n+1}\left(\frac{en}{n+1-k} \right)^{n+1-k} = O\left(h^{n+1-k}\right).$$

Найпростіші формули чисельного диференціювання

В задачах чисельного диференціювання завжди необхідно брати до уваги, що незначні помилки у таблиці початкових значень функції можуть суттєво спотворити значення похідної. Тому, перш ніж застосовувати формули чисельного диференціювання, початкові данні необхідно згладити. Тому для обчислення похідних застосовують формули, які згладжують дані за методом найменших квадратів.

Найпростіша і найуживаніша формула похідної є:

$$y_0' = \frac{1}{h}\left(\frac{y_1 - y_{-1}}{2}\right) = \frac{1}{h}\left(\frac{\Delta y_0 + \Delta y_{-1}}{2}\right) = \frac{1}{h}\left(\frac{\Delta y_0 + \nabla y_0}{2}\right),$$

яка отримується при лінійному згладжуванні за трьома точками (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1) і представляє собою скінченну різницю першого порядку. Ця формула може застосовуватись для всіх точок, крім першої і останньої. Для визначення похідної в першій і останній точках користуються наступними відповідними формулами:

$$y_0' = \frac{1}{2h}(-3y_0 + 4y_1 - y_2) = \frac{1}{h}\left(\Delta y_0 - \frac{1}{2}\Delta^2 y_0\right),$$

$$y_0' = \frac{1}{2h}(3y_0 - 4y_{-1} + y_{-2}) = \frac{1}{h}\left(\Delta y_{-2} + \frac{3}{2}\Delta^2 y_{-2}\right),$$

але ці формули менш точні, тому у крайніх точках переважно похідні не обчислюють. Для рівновіддалених вузлів формули можна переписати наступним чином:

$$y'(x_0 + ph) = \frac{1}{h}(p - 0{,}5)y_{-1} - 2py_0 + (p + 0{,}5)y_1,$$

де $p = (x - x_0)/h$, $x = x_0 + ph$.

Формули чисельного диференціювання за трьома точками є дуже наближеними, і більш точний результат можна отримати при збільшенні кількості точок, особливо це необхідно при диференціюванні функцій з вищими показниками степеня. Вузли диференціювання нарощують по обидва боки від точки, в якій відшукується значення похідної. Таке нарощування точок приводить до формул, які називаються формулами центральної різниці. Так формула центральної різниці для першої похідної може бути записана наступним чином:

$$y'(x) \approx \frac{f(x+h) - f(x-h)}{2h} = \frac{y_1 - y_{-1}}{2h}.$$

При виборі кількості точок для обчислення похідної необхідно керуватись наступними положеннями:

1) точністю з якої обчислюється похідна,
2) степенем функції, яка диференціюється,
3) розташуванням точки диференціювання на заданому відрізку значень.

Необхідно пам‘ятати, що нарощування кількості точок приводить до ускладнення обчислень. Так, формули чисельного диференціювання за сімома точками майже не застосовуються.

Наведемо основні центральні формули чисельного диференціювання у таблиці:

к-ть вузлів	формула похідної	похибка
3	$y' = (y_1 - y_{-1})/2h$	$-h^2 y^{(3)}/6$
	$y'' = (y_1 - 2y_0 + y_{-1})/h^2$	$-h^2 y^{(4)}/12$
5	$y' = (-y_2 + 8y_1 - 8y_{-1} + y_{-2})/12h$	$h^4 y^{(5)}/30$
	$y'' = (-y_2 + 16y_1 - 30y_0 + 16y_{-1} - y_{-2})/12h^2$	$-h^4 y^{(5)}/90$
	$y''' = (y_2 - 2y_1 + 2y_{-1} - y_{-2})/2h^3$	$-h^2 y^{(5)}/4$
7	$y' = (y_3 - 9y_2 + 45y_1 - 45y_{-1} + 9y_{-2} - y_{-3})/60h$	$-h^6 y^{(7)}/140$
	$y'' = (2y_3 - 27y_2 + 270y_1 - 490y_0 + 270y_{-1} - 27y_{-2} + 2y_{-3})/180h^2$	$-h^6 y^{(8)}/560$
	$y' = (-y_3 + 8y_2 - 13y_1 + 13y_{-1} - 8y_{-2} + y_{-3})/8h^3$	$7h^4 y^{(7)}/120$

Приклад. Обчислити похідну функції $y = \sin x$ в точці $x = \pi/6$.

Візьмемо $h = 0.1$.

1. $y' = \cos(\pi/6) = 0{,}866025$ – точне значення;
2. $y' = \dfrac{y(x+h) - y(x)}{h} = \dfrac{\sin(\pi/6 + 0.1) - \sin(\pi/6)}{0.1} = 0{,}839604.$ – права різниця, абсолютна похибка $\Delta = 0{,}026$, відносна похибка $\delta = 0{,}031$;
3. $y' = \dfrac{y(x+h) - y(x-h)}{2h} = \dfrac{\sin(\pi/6 + 0.1) - \sin(\pi/6 - 0.1)}{0.2} = 0{,}864583$ – центральна різниця (за трьома точками), абсолютна похибка $\Delta = 0{,}00142$, відносна похибка $\delta = 0{,}001665$;
4. $y' = \dfrac{(-y_2 + 8y_1 - 8y_{-1} + y_{-2})}{12h} =$ $= \dfrac{-\sin(\pi/6 + 0{,}2) + 8\sin(\pi/6 + 0{,}1)18\sin(\pi/6 - 0{,}1) + \sin(\pi/6 - 0{,}2)}{12h} = 0{,}866023$ – за п‘ятьма точками, абсолютна похибка $\Delta = 2 \cdot 10^{-6}$, відносна похибка $\delta = 2{,}31 \cdot 10^{-6}$.

Похідні вищих порядків обчислюються як похідні від похідних нижчих порядків. Так, похідна другого порядку обчислюється як похідна від похідної першого порядку $f''(x) = (f'(x))'$.

Наведені у таблиці формули чисельного диференціювання можна застосовувати для обчислення часткових похідних функцій багатьох змінних, якщо задавати приріст однієї змінної і лишати постійними інші змінні.

Уточнення похідної за Річардсоном

Уточнення формули диференціювання виконують виходячи з основного оператора диференціювання D_h:

$$D_{h,0}y_0 = \frac{y_1 - y_{-1}}{2h},$$

де h – крок.

Цей оператор дає значення похідної $y_0' \approx f'(x_0)$ з похибкою порядку h^2. Для вилучення головного члену помилки цього оператора вводиться оператор $D_{h,1}y_0 = \frac{4D_h y_0 - D_{2h}y_0}{3} = D_h y_0 + \frac{D_h y_0 - D_{2h}y_0}{3}$, де $D_{2h}y_0 = (y_2 - y_{-2})/4h$ – теж основний оператор чисельного диференціювання. Похибка оператора $D_{h,1}y_0$ має вже порядок h^4. Якщо таке уточнення є недостатнім, то вводиться наступний оператор:

$$D_{h,2}y_0 = \frac{2^4 D_{h,1} - D_{2h,1}}{2^4 - 1} = D_{h,1} + \frac{D_{h,1} - D_{2h,1}}{15},$$

який вилучає головну частину похибки оператора $D_{h,1}y_0$. Для процесу уточнення можна вивести рекурентну формулу:

$$D_{h,n+1} = \frac{2^{2n+2} D_{h,n} - D_{2h,n}}{2^{2n+2} - 1}, \quad m = 0,1,2,\ldots.$$

Враховуючи той факт, що початкові дані містять похибки, процес уточнення недоцільно продовжувати досить далеко.

Приклад. Нехай задана функція $y = e^x$ з кроком $h = 0,1$ з точністю до четвертого знака. Знайти похідні і оцінити їхню похибку.

x	y	$D_h y$	$D_{2h} y$	$D_{h,1} y$	похибка $D_{h,1} y$	похибка $D_h y$
1,0	2,7183					
1,1	3,0042	3,0090				0,0048
1,2	3,3201	3,3255	3,3422	3,3199	-0,0002	0,0054
1,3	3,6693	3,6755	3,6937	3,6694	0,0001	0,0062
1,4	4,0552	4,0620	4,0822	4,0553	0,0001	0,0068
1,5	4,4817	4,4890	4,5115	4,4815	-0,0002	0,0073
1,6	4,9530	4,9610	4,9860	4,9527	-0,0003	0,0080
1,7	5,4739	5,4830				0,0091
1,8	6,0496					

Похибка чисельного диференціювання складається з двох частин: безпосередньо похибки самої формули чисельного диференціювання (похибка усікання) і похибки початкових даних (похибка округлення). При зміні кроку h чисельного диференціювання ці похибки змінюються в протилежних напрямках, що дозволяє підібрати оптимальний крок, який дасть найменшу сумарну похибку, яка оцінюється значенням $\varepsilon / h + M_3 h^2 / 6$, де ε – гранична похибка зміни значень y_i, $M_3 = \max\left| f'''(x) \right|$.

Ця сумарна оцінка досягає мінімуму при значенні $\approx \varepsilon^{2/3} M_3^{1/3}$, коли крок

$h = h_{\text{опт}} = \sqrt[3]{3\varepsilon / M_3}$ (рис.2).

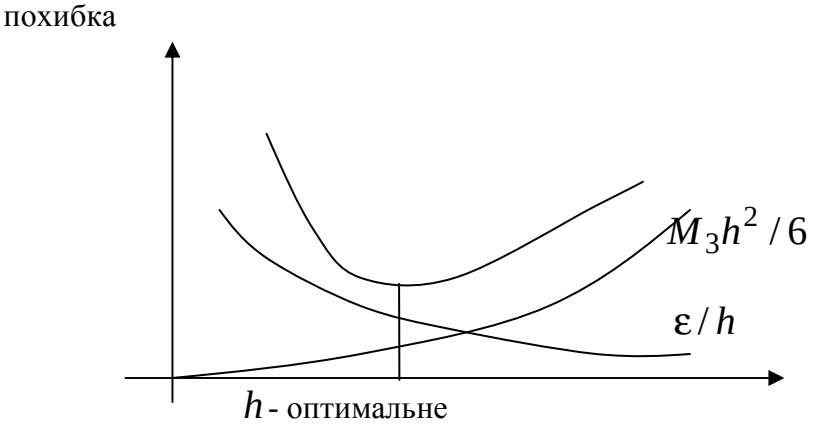


Рис. 2. Графік сумарної похибки чисельного диференціювання

При цьому похибка усікання складає 1/3 сумарної похибки, а похибка округлення –2/3. Для практичного вибору оптимального кроку рекомендують наступне правило: вибирати крок таким чином, щоб абсолютна величина різниць 3-го порядку $\Delta^3 y_i$ знаходилась в межах $(3\varepsilon;6\varepsilon)$. Іншими словами, значення $\left(\Delta^3 y_{-1} + \Delta^3 y_{-2}\right)/12$ за абсолютною величиною повинно лежати між $\varepsilon/2$ і ε.

Якщо за умови оптимального кроку точність обчислення похідної є недостатньою, то беруть більш точну формулу з нарощуванням кількості опорних точок. Оптимальний крок за п‘ятьма точками можна визначити як $h = h_{\text{опт}} = \sqrt[5]{45\varepsilon / 4M_5}$ з найменшою сумарною похибкою $15\varepsilon/8h \approx 1{,}2\varepsilon^{4/5} M_5^{1/5}$.

Контрольні запитання

1. Яка функція називається сітковою?
2. Що називається скінченною різницею (правою, лівою, центральною)?
3. За рахунок чого можна збільшити точність обчислення значення похідної у точці?
4. Як можна визначити оптимальне значення кроку диференціювання?
5. У чому полягає ідея операції уточнення значення похідної до заданої точності?
6. Яким критерієм можна керуватись для визначення оптимального кроку чисельного диференціювання?

Лекція 10. Чисельне інтегрування

Формули чисельного інтегрування спираються на геометричний зміст визначеного інтеграла як площі криволінійної трапеції, обмеженої інтервалом інтегрування, віссю OX і підінтегральною функцією.

Поняття квадратури

Нехай ставиться задача наближено обчислити значення визначеного інтеграла від функції $f(x)$ на інтервалі $[a;b]$. Якщо важко знайти значення інтегралу або виразити результат у зручній формі через елементарні функції, тоді підінтегральну функцію $f(x)$ заміщують деякою апроксимуючою функцією $\varphi(x,a)$, що $f(x) \approx \varphi(x)$. За функцію $\varphi(x)$ найчастіше беруть інтерполяційний многочлен або ряд (приклад, розглянутий у розділі 1 $\int_0^{1/2} e^{x^2} dx$). Оскільки таке наближення є лінійним відносно параметрів, то функцію заміщують деяким лінійним виразом, коефіцієнтами якого є значення функції $f(x)$ у вузлах інтерполяції $\{x_i\}_{i=0}^n$:

$$f(x) = \sum_{i=0}^{n} f(x_i)\varphi_i(x) + E(f).$$

Визначення. Нехай $a = x_0 < x_1 < ... < x_n = b$. Формула вигляду:

$$Q(f) = \sum_{i=0}^{n} c_i f(x_i) = c_0 f(x_0) + c_1 f(x_1) + ... + c_n f(x_n),$$

яка має наступну властивість: $\int_a^b f(x)dx = Q(f) + E(f)$, називається формулою чисельного інтегрування або формулою квадратури (підінтегральна функція заміщується сумою). Доданок $E(f)$ – залишковий член, який становить похибку усікання для чисельного інтегрування. Множина точок $\{x_i\}$ називається вузлами квадратури, а $\{c_i\}$ – вагою квадратури.

Вузли квадратури можна вибирати різним чином: вони повинні бути з постійним кроком для формул прямокутників, трапецій, Сімпсона, або вибиратись за спеціальними правилами, наприклад, бути нулями певних поліномів Лежандра для формули Гаусса-Лежандра.

Зауважимо, що на практиці формула прямокутників (наближення кривої, що задається функцією $f(x)$ ламаною лінією) не

застосовується, оскільки дає значну похибку. Ця формула дуже просто виводиться з геометричного змісту первісної, і тому ми її розглядати не будемо.

Отримання квадратурних формул базується на інтерполяційному многочлені (існує єдиний поліном степені $\le n$, який проходить через $n+1$ точку). Коли цей поліном застосовується для наближення підінтегральної функції $f(x)$, а далі інтеграл наближається за допомогою полінома $P_n(x)$, то отримана формула інтеграла називається формулою Ньютона-Котса. Якщо для інтегрування застосовуються лише точки $x_0 = a$ і $x_n = b$, то формула називається замкненою формулою Ньютона-Котса.

Якщо функція $f(x)$ має достатню кількість похідних, тоді похибка $E(f)$ для квадратури Ньютона-Котса містить відповідні похідні вищого порядку.

Якщо відрізок інтегрування $[a,b]$ розбити на n рівних частин, довжина яких $h = (b-a)/n$ і $x_k = a + ih,\ k = 0,1,2,...,n$, обчислити площу криволінійної трапеції на кожному частковому відрізку за допомогою певної формули квадратури, то інтеграл буде наближено дорівнювати сумі отриманих значень, а узагальнена формула для n відрізків називається складеною або узагальненою.

Виведемо формули для різних квадратур.

Формула трапецій

Проапроксимуємо задану підінтегральну функцію на відрізку $[a,b]$ лінійною функцією (крива замінюється січною (рис. 3)).

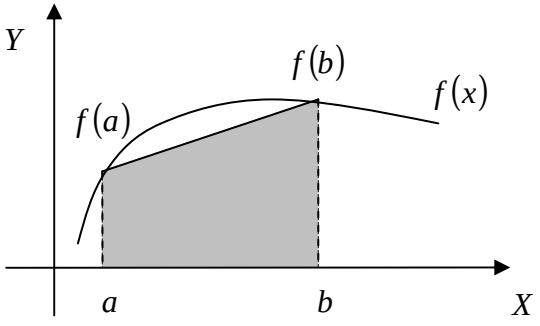


Рис. 3. Наближення функції прямою

Тоді значення визначеного інтеграла на відрізку $[a,b]$ наближено буде дорівнювати площі утвореної трапеції, де $f(a)$, $f(b)$ – довжини основ цієї трапеціє, а $(b-a)$ – висота. Тоді: $S_{\text{трапеції}} = \frac{f(a)+f(b)}{2}(b-a) \approx \int_a^b f(x)dx$ – основна замкнена квадратура Ньютона-Котса для формули трапецій.

Для оцінки похибки формули трапецій розкладемо функцію $f(x)$ в ряд Тейлора в точці середини відрізка інтегрування $€ = (b-a)/2$:

$$f(x) = f(€) + f'(€)(x-€) + \frac{f''(€)}{2}(x-€)^2 + \ldots$$

і підставимо цей вираз у формулу абсолютної похибки, отримаємо:

$$E(x) = \int_a^b f(x)dx - \frac{f(a)+f(b)}{2}(b-a) \approx -\frac{f''(€)}{12}(b-a)^3 .$$

Зрозуміло, що чим більше буде відрізок інтегрування тим більшою буде похибка формули трапецій.

Для зменшення похибки відрізок $[a,b]$ розіб'ємо на n рівних частин і виведемо складену формулу трапецій (рис. 4).

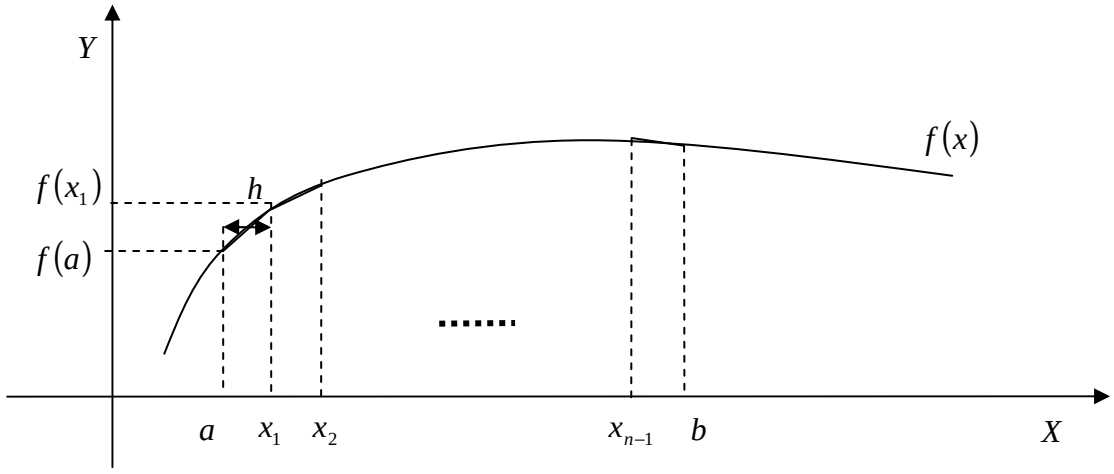


Рис. 4. Розбиття відрізку інтегрування на часткові інтервали

Запишемо формули обчислення площі для кожної трапеції, де $h = (b-a)/n$:

$$S_1 = \frac{f(a)+f(x_1)}{2}h = \frac{f(a)+f(a+h)}{2}h$$

$$S_2 = \frac{f(x_1)+f(x_2)}{2}h = \frac{f(a+h)+f(a+2h)}{2}h$$

.....

$$S_n = \frac{f(x_{n-1})+f(b)}{2}h = \frac{f(a+(n-1)h)+f(b)}{2}h.$$

Тоді інтеграл буде наближено дорівнювати сумі площ утворених трапецій, і узагальнену формулу можна записати двома способами:

1. $I = \int_a^b f(x)dx \approx \sum_{i=1}^{n} S_i = h\left(\frac{f(a)+f(b)}{2} + \sum_{i=1}^{n-1} f(a+ih)\right)$

2. $I = \int_a^b f(x)dx \approx \frac{h}{2}\sum_{k=1}^{n}(f(x_{k-1})+f(x_k))$

з похибкою $E(x) \approx -\frac{1}{12}h^3\sum_i f''(\xi_i) \approx -\frac{1}{12}h^2\int_a^b f''(x)dx$ при $h \to 0$. Щоб оцінити похибку за наведеною формулою, необхідно існування другої похідної (неперервність функції), якщо функція на відрізку $[a,b]$ кусково-неперервна, то оцінити похибку можна за допомогою мажоранти:

$$E(x) \le \frac{b-a}{12}h^2M_2, \text{ де } M_2 = \max_{[a,b]}|f''(x)|.$$

Для довільного розбиття відрізку $[a,b]$ (сітки з довільним кроком) формулу трапецій можна представити наступним чином:

$$I = \int_a^b f(x)dx \approx \frac{}{2}\sum_{k=1}^{n}(f(x_{k-1})+f(x_k))(x_k - x_{k-1}).$$

Причому, формула трапецій має другий порядок точності відносно кроку сітки (як видно з випадку рівномірної сітки).

Приклад. Використовуючи формулу трапецій, обчислити $\int_0^{\pi} \sin x dx$.

Розв'язок. Знайдемо точне значення інтеграла:

$$\int_0^{\pi/2} \sin x dx = -\cos x\Big|_0^{\pi/2} = \cos 0 - \cos\frac{\pi}{2} = 1.$$

1. Застосуємо формулу основної квадратури:

$$\int_0^{\pi/2} \sin x dx \approx \frac{\sin 0{,}5\pi + \sin 0}{2}\frac{\pi}{2} = 0{,}785398,\ \Delta = 0{,}214602.$$

2. Застосуємо складену формулу для $n = 10$, тоді

$$h = \left(\frac{\pi}{2} - 0\right)/10 = 0{,}1570796$$

$$\int_0^{\pi/2} \sin x dx \approx h\left(\frac{\sin 0{,}5\pi + \sin 0}{2} + \sum_{i=1}^{9} \sin(0 + ih)\right) = 0{,}997943\ \ \Delta = 0{,}002057.$$

Наближення поліномом (формули Сімпсона)

Нехай на відрізку інтегрування $[a,b]$ задано n точок $\{x_i\}_{i=0}^n$ так, що $x_0 = a$, $x_n = b$ і $y_i = f(x_i)$. Побудуємо за цими точками інтерполяційний многочлен Лагранжа:

$$P_n(x) = \sum_{i=0}^{n} y_i \frac{\varpi_{n+1}(x)}{\varpi'_{n+1}(x_i)(x - x_i)},$$

де $y_i = P_n(x_i)$, $\varpi_{n+1}(x) = (x - x_0)(x - x_1)...(x - x_n)$,
і підставимо його замість функції $f(x)$.

$$I = \int_a^b f(x)dx \approx \int_a^b P_n(x)dx = \sum_{i=0}^{n} y_i \int_a^b \frac{\varpi_{n+1}(x)dx}{\varpi'_{n+1}(x_i)(x - x_i)} = \sum_{i=0}^{n} y_i A_i .$$

Поклавши, що $f(x) = x^k$, $k = 0,1,2,...$, отримаємо систему лінійних рівнянь:

$$I_0 = \sum_{i=0}^{n} A_i$$

$$I_1 = \sum_{i=0}^{n} A_i x_i$$

.....

$$I_n = \sum_{i=0}^{n} A_i x_i^n ,$$

де $I_k = \int_a^b x^k dx = \frac{b^{k+1} - a^{k+1}}{k+1}$. При наближенні підінтегральної функції многочленом, як правило, вибирають рівномірну сітку ($h = (b-a)/n - \text{const}$, $q = (x - x_0)/h$), тоді коефіцієнти системи знаходять за формулою:

$$A_i = (b-a)\frac{(-1)^{n-i}}{n \cdot i! \cdot (n-i)!}\int_0^n \frac{q^{[n+1]}}{q-i} dq = (b-a)H_i$$

де $q^{[n+1]} = q(q-1)(q-2)...(q-n)$.

Квадратурна формула буде мати вигляд:

$$\int_a^b f(x)dx \approx (b-a)\sum_{i=0}^n H_i y_i,\ y_i = f(a+ih).$$

Зауваження. Для контролю корисно користуватись властивістю:

$$\sum_{i=0}^n H_i = 1,\ H_i = H_{n-i}.$$

Якщо підінтегральну функції наблизити квадратичною параболою $n = 2$ (рис. 5), обчислимо коефіцієнти

$$H_0 = \frac{1}{4}\int_0^2 (q-1)(q-2)dq = \frac{1}{6}$$

$$H_1 = \frac{1}{2}\int_0^2 q(q-2)dq = \frac{2}{3}$$

$$H_2 = \frac{1}{4}\int_0^2 q(q-1)dq = \frac{1}{6},$$

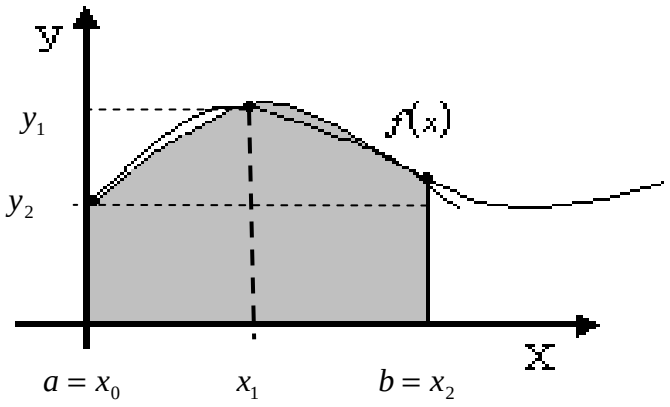


Рис. 5. Наближення підінтегральної функції параболою

та отримаємо формулу квадратури Сімпсона:

$$\int_a^b f(x)dx \approx (b-a)\left(\frac{1}{6}y_0 + \frac{2}{3}y_1 + \frac{1}{6}y_2\right) = \frac{h}{3}(y_0 + 4y_1 + y_2),\ h = (b-a)/2.$$

Для випадку наближення кубічним поліномом ($n = 3$) отримаємо

$\int_{x_0}^{x_1} f(x)dx \approx \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + y_3)$ – формулу Сімпсона 3/8.

Для $n = 4$ – $\int_{x_0}^{x_1} f(x)dx \approx \frac{2h}{45}(7y_0 + 32y_1 + 12y_2 + 32y_3 + 7y_4)$ – формулу Буля.

Для виводу складеної формули Сімпсона необхідно, щоб відрізок був розбитий на $2n$ підінтервали однакової довжини і загальна кількість точок становила $2n+1$ (це пов‘язано з тим, що для побудови полінома другого степеня необхідно мати на кожному частковому відрізку три точки). Складену формулу Сімпсона можна записати наступними еквівалентними способами:

1. $I = \int_a^b f(x)dx \approx \frac{h}{3}\sum_{k=1}^{n}(f(x_{2k-2}) + 4f(x_{2k-1}) + f(x_{2k}))$;
2. $I = \int_a^b f(x)dx \approx \frac{h}{3}\left(f(a) + f(b) + 2\sum_{k=1}^{n-1} f(x_{2k}) + 4\sum_{k=1}^{n} f(x_{2k-1})\right)$.

Залишковий член для складеної формули Сімпсона оцінюється наступною формулою: $E(f,h) = \frac{-(b-a)f^{(4)}(c)h^4}{180} = O(h^4),\ c \in [a,b]$.

Приклад. Обчислити $\int_0^{\pi/2} \sin x dx$, використовуючи формулу Сімпсона.

Розв‘язок.

1. За формулою основної квадратури:

$$\int_0^{\pi/2} \sin x dx \approx \frac{\pi}{6}\left(\sin 0 + 4\sin\frac{\pi}{4} + \sin\frac{\pi}{2}\right) = 0{,}893839,\ \Delta = 0{,}106161;$$

2. При $n = 10$ $i = 2*n+1 = 21,\ h = \pi/42 = 0{,}076624$

$$\int_0^{\pi/2} \sin x dx \approx \frac{\pi}{6}\left(\sin 0 + \sin\frac{\pi}{2} + 2\sum_{i=1}^{9}\sin(2ih) + 4\sum_{i=1}^{10}\sin(h(2i+1))\right) = 1{,}055986,\ \Delta = 0{,}055986.$$

Порівнявши результати, отримані за допомогою формули трапецій та Сімпсона, бачимо, що для формул основних квадратур формула Сімпсона вдвічі точніша, а для складених формул більш точний результат дала формула трапецій. Цей результат обумовлений накопиченням похибки округлення (за формулою Сімпсона було виконано значно більше операцій).

Якщо задана допустима похибка обчислень ε, то крок для застосування формули Сімпсона знаходять з умови:

$$h < \sqrt[4]{\frac{180\varepsilon}{(b-a)M_4}},\ \text{де}\ M_4 = \max_{[a;b]}\left|y^{(4)}(x)\right|.$$

Якщо похибка заздалегідь невідома, то виконують обчислення інтеграла з кроком h і $H = 2h$, і за наближене значення інтеграла приймають величину:

$$I = \sum{}_h + \frac{\sum{}_h - \sum{}_H}{15}.$$

Для функцій без особливостей формула трапецій і Сімпсона дають майже однакову точність

Для випадку, коли підінтегральна функція містить особливості, ці формули можуть давати значну похибку, і для таких випадків застосовують спеціальні методи.

Схема Ромберга

Метод Ромберга полягає в послідовному уточненні наближеного значення інтеграла. Цей метод застосовують, коли необхідно обчислити значення визначеного інтеграла із заданою або досить високою точністю. Зручність методу обумовлена використанням рекурентних формул.

В основі методу Ромберга лежить обчислення інтегралу за формулою трапецій на регулярній сітці з кроком h і уточнення результату за тією ж формулою з кроком $2h$.

Введемо наступні позначення:

$T_i = h_i\left(\frac{y_0 + y_n}{2} + \sum_{i=1}^{n-1} y_i\right)$ – формула трапецій для кроку h_i;

$h_0 = (b-a)/n$ – початковий крок;

$h_{i+1} = h_i / 2,\ i = 0,1,2,...k-1.$

Обчислимо початкове наближення інтеграла з кроком h_0:

$$T_0 = h_0\left(\frac{f(a)+f(b)}{2} + \sum_{i=1}^{n-1} f(a+ih_0)\right).$$

Далі, обчислимо наступне значення кроку $h_1 = h_0/2 = (b-a)/2n$ і наступне наближення:

$$T_1 = h_1\left(\frac{f(a)+f(b)}{2} + \sum_{i=1}^{2n-1} f(a+ih_1)\right).$$

Щоб не виконувати зайвих операцій і повторно не обчислювати ті значення, які були вже обчислені на попередньому кроці, для визначення T_1 достатньо тільки знайти суму значень функції $f(x)$ у середніх точках $a+h_1,\ a+3h_1,...,\ a+(2n-1)h_1$. Запишемо цю суму як:

$$V_0 = h_0\sum_{i=1}^{n} f\left(a+\left(i-\frac{1}{2}\right)h_0\right).$$

Тоді, користуючись введеними позначеннями,

$$T_1 = \frac{1}{2}(T_0 + V_0).$$

За двома знайденими наближеннями обчислюють найкраще наближення інтеграла за формулою, що визначає правило Сімпсона:

$$S_1 = T_1 + \frac{1}{3}(T_1 - T_0) = \frac{1}{3}(4T_1 - T_0).$$

Далі кількість інтервалів ділять на 4 (новий крок $h_2 = h_1/2 = h_0/4$), обчислюють новий поправочний член:

$$V_1 = h_1\sum_{i=1}^{2n} f\left(a+\left(i-\frac{1}{2}\right)h_1\right)$$

і нове наближення за формулою трапецій:

$$T_2 = \frac{1}{2}(T_1 + V_1)$$

і за правилом Сімпсона:

$$S_2 = T_2 + \frac{1}{3}(T_2 - T_1) = \frac{1}{3}(4T_2 - T_1).$$

За двома знайденими наближеннями інтеграла (S_1 з кроком $h_1 = 2h_2$ і S_2 з кроком h_2) обчислимо найкраще наближення:

$$E_2 = S_2 + \frac{1}{15}(S_2 - S_1) = \frac{1}{15}(16S_2 - S_1).$$

Якщо покласти $I \approx E_2$, то похибка усікання вже має шостий порядок $O\left(h_2^6\right)$, а похибка округлення за правилом 3-х сігм ($\approx 3\frac{1}{6}\frac{(b-a)\sigma}{\sqrt{2n}}$) перевищує похибку округленнь у формулі Сімпсона приблизно на 6%.

Продовження процесу уточнення для $k \geq 3$ називається схемою Ромберга і може бути представлено наступним чином:

h	T	S	R	Q
h_0	T			
	$\downarrow$			
$h_1 = 1/3\, h_0$	T_1	$\rightarrow S_1$		
	$\downarrow$			
$h_2 = 1/2\, h_1 = 1/4\, h_0$	T_2	$\rightarrow S_2$	$\rightarrow R_2$	
	$\downarrow$			
$h_3 = 1/2\, h_2 = 1/8\, h_0$	T_3	$\rightarrow S_3$	$\rightarrow R_3$	$\rightarrow Q_3$
	...	...	...	...

де в кожному рядку обчислення виконуються за наступними формулами:

$$h_i = \frac{b-a}{2^i n},\ T_{i+1} = \frac{1}{2}(T_i + V_i),\ V_i = h_i \sum_{j=1}^{2^i n} f\left(a + \left(j - \frac{1}{2}\right)h_i\right),$$

$$S_{i+1} = \frac{2^2 T_{i+1} - T_i}{2^2 - 1},\ R_{i+1} = \frac{2^4 S_{i+1} - S_i}{2^4 - 1},\ Q_{i+1} = \frac{2^6 R_{i+1} - R_i}{2^6 - 1},\ \ldots$$

Процес обчислень завершується, коли різниця двох послідовних наближень буде менше або дорівнювати заданій точності.

Для програмної реалізації схеми можна записати наступну рекурентну формулу: $\int_a^b f(x)dx \approx R(i,i)$,

де $R(i,k) = R(i,k-1) + \dfrac{R(i,k-1) - R(i-1,k-1)}{4^k - 1}$ – елементи рядка таблиці наближень $R(i,k)$, $i \geq k$. Значення $R(i,0)$ обчислюються шляхом послідовного застосування формули трапецій на 2^i підінтервалах відрізку $[a;b]$. Процес завершується, коли виконається умова: $|R(i,i) - R(i+1,i+1)| \leq \varepsilon$, де ε – задана точність.

Зауваження. Якщо підінтегральна функція задана таблично, то для обчислення її інтеграла за схемою Ромберга не потребує попереднього згладжування початкових даних, бо саме інтегрування виконує часткове згладжування та „очищує” дані від „шуму”.

Формула Ейлера

Формула Ейлера отримується при наближенні підінтегральної функції поліномом Ерміта. Ця формула залежить від значень похідних у вузлах інтерполювання і в загальному випадку досить громіздка і незручна. Запишемо формулу Ейлера з використанням першої похідної. Для цього скористаємось формулою трапецій і запишемо її залишковий член через першу похідну:

$$E(x) \approx -\frac{1}{12}(b-a)^3 f''(x) \approx \frac{1}{12}(b-a)^2 (f'(a) - f'(b)),$$

тоді отримаємо:

$$\int_a^b f(x)dx \approx \frac{1}{2}(b-a)(f(a) + f(b)) + \frac{1}{12}(b-a)^2 (f'(a) - f'(b)).$$

Дати оцінку точності формули Ейлера можна, обчисливши залишковий член за допомогою розкладення підінтегральної функції у ряд Тейлора:

$$E(x)_{\text{Ейлера}} = \frac{1}{720}(b-a)^5 f^{(4)}(x).$$

З вигляду залишкового члена видно, що формула Ейлера має порядок точності $O\left(h^4\right)$, що на два порядки точніша за формулу трапецій, з якої вона була отримана.

Якщо залишковий член формули трапецій виражати через вищі похідні, то отримані формули називають формулами Ейлера-Маклорена. Спрощений вигляд має узагальнена формула Ейлера-Маклорена для рівномірної сітки:

$$\int_a^b f(x)dx \approx h\left(\frac{f(a)+f(b)}{2}+\sum_{i=1}^{n-1} f(a+ih)\right)+\frac{1}{12}h^2\left(f'(a)-f'(b)\right),$$

де $h=(b-a)/n-\text{const}$.

Формули Гаусса-Крістоффеля

Розглянемо узагальнену задачу чисельного інтегрування. Нехай необхідно обчислити

$$I=\int_a^b f(x)\rho(x)dx,\ \rho(x)>0, \qquad (1)$$

де функція $f(x)$ неперервна на відрізку $[a;b]$, а вагова функція $\rho(x)$ неперервна на інтервалі $(a;b)$. Необхідно отримати квадратурну формулу:

$$I\approx\sum_{i=0}^{n} c_i f(x_i)+E(x),$$

де x_i – вузли, c_i – вагові коефіцієнти є її параметрами. Для n вузлів формула квадратури містить $2n$ праметрів, таку ж саму кількість параметрів містить многочлен степені $2n-1$. Отже квадратурну формулу можна підібрати так, щоб вона була точною для будь-якого многочлена степені не вище $2n-1$.

Будемо вважати, що вага $\rho(x)>0$ і неперервна на інтервалі (a,b), а на кінцях відрізку може приймати нульові значення так, щоб існував інтеграл $\int_a^b \rho(x)dx$. Тоді існує повна система алгебраїчних многочленів $P_m(x)$, ортогональних на $[a,b]$ із заданою вагою

$$\int_a^b P_k(x)P_m(x)\rho(x)dx = \delta_{km}\|P_k(x)\|_{L_2}^2 .$$

Всі корені цих многочленів є дійсними числами, що розташовані на інтервалі (a,b) і являються вузлами формули Гасса-Кристоффеля з вагою $\rho(x)$, яку можна визначити, якщо вузли вже відомі. Функція

$$\psi_m(x) = \prod_{k=1,k\neq m}^{n} (x - x_k)/(x_m - x_k)$$

є многочленом степеня $n-1$, отже для неї функція Гаусса-Кристоффеля є точною. Враховуючи, що ця функція дорівнює нулю в усіх вузлах, крім m–го, вагові коефіцієнти формули Гаусса-Кристоффеля можна представити наступним чином:

$$c_m = \int_a^b \rho(x) \prod_{k=1,k\neq m}^{n} (x - x_k)/(x_m - x_k)dx.$$

Підставляючи у формулу Гаусса $f(x)=1$, отримаємо формулу:

$$\sum_{k=1}^{n} c_k = \int_a^b \rho(x)dx,$$

з якої випливає рівномірна обмеженість вагових коефіцієнтів.

Формули Гаусса-Крістоффеля називають формулами найвищої алгебраїчної точності.

Окремі випадки формули Гаусса-Крістоффеля

1. Безпосередня формула Гаусса відповідає випадку $\rho(x)=1$. На відрізку $[-1,1]$ ортогональними з одиничною вагою є многочлени Лежандра. Наведемо їх формули для перших значень індекса n, корені $\xi_i^{(n)}$ та відповідні ваги $\gamma_i^{(n)} \cdot L_0(x)=1$

Таблиця многочленів Лежандра

$L_n(x)$	ξ_n	γ_n
$L_1(x)=x$	$\xi_1=0$	$\gamma_1=2$
$L_2(x)=\frac{3x^2-1}{2}$	$-\xi_1=\xi_2=\sqrt{1/3}$	$\gamma_1=\gamma_2=1$
$L_3(x)=\frac{5x^3-3x}{2}$	$-\xi_1=\xi_3=\sqrt{3/5}, \xi_2=0$	$\gamma_1=\gamma_3=5/9, \gamma_2=8/9$

$L_4(x)=\dfrac{35x^4-30x^2+3}{8}$	$-\xi_1=\xi_4=\sqrt{\left(15+2\sqrt{30}\right)/35},$ $-\xi_2=\xi_3=\sqrt{\left(15-2\sqrt{30}\right)/35}$	$-\gamma_1=\gamma_4=\left(18-\sqrt{30}\right)/36$ $\gamma_2=\gamma_3=\left(18+\sqrt{30}\right)/36$
$L_5(x)=\dfrac{63x^5-70x^3+15x}{8}$	$-\xi_1=\xi_5=\sqrt{\left(35+2\sqrt{70}\right)/63},$ $-\xi_2=\xi_4=\sqrt{\left(35-2\sqrt{70}\right)/63},$ $\xi_3=0$	$\gamma_1=\gamma_5=\left(322-13\sqrt{70}\right)/900$ $\gamma_2=\gamma_4=\left(322+13\sqrt{70}\right)/900,$ $\gamma_3=128/225$

Значення вузлів та вагових коефіцієнтів для довільного відрізку $[a,b]$ можна отримати за допомогою лінійних перетворень:

$$x_k=\frac{a+b}{2}+\frac{b-a}{2}\xi_k,\ c_k=\frac{b-a}{2}\gamma_k,\ 1\le k\le n.$$

Похибка формули Гаусса пропорційна похідній, яка відповідає найменшому неврахованому степеню аргументу. Верхня границя похибки становить:

$$\max|R|=\frac{(b-a)^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3}M_{2n}\approx\frac{b-a}{2{,}5\sqrt{n}}\left(\frac{b-a}{3n}\right)^{2n}M_{2n},\ \text{де}\ M_{2n}=\max_{[a,b]}\left|f^{(2n)}(x)\right|.$$

Формула Гаусса застосовується до гладких функцій, які мають високі похідні, але не приймають досить великі значення за абсолютною величиною.

2. Для формули Ерміта на відрізку $[-1{,}1]$ з вагою $\rho(x)=1/\sqrt{1-x^2}$ ортогональними є многочлени Чебишова першого роду $T_n(x)$. Відповідні вузли та вагові коефіцієнти (вони для всіх вузлів є однаковими) обчислюються за наступними формулами:

$$\xi_k-\cos[\pi(k-0{,}5)n],\gamma_k=\pi/n,\ 1\le k\le n.$$

Для переходу на довільний відрізок застосовуються ті ж самі лінійні перетворення, що і для формули Гаусса.

Похибка формули Ерміта не перевищує $\max|E|=\pi\dfrac{M_{2n}}{2^{n-1}(2n)!}$.

Лекція 11. Інтегрування функцій, що містять особливості

Інтегрування розривних функцій

Якщо застосовувати квадратурні формули до функцій, що містять особливості, не виділяючи особливих точок, то зменшення

кроку сітки не дасть суттєвого прискорення збіжності інтегралу. Наприклад,

$$\int_{-1}^{2} x|x|dx = 7/3$$

– підінтегральна функція є неперервною та гладкою, але друга похідна має розрив у точці $x = 0$. Якщо для такої функції виділити відрізки неперервності $[-1,0]$ і $[0,2]$, то квадратурні формули будуть давати досить точну відповідь. У протилежному випадку при інтегруванні за всім проміжком точка $x = 0$ не буде вузловою, і навіть згущення сітки не дасть хорошої збіжності.

Якщо підінтегральна функція та її похідні є кусково-неперервними, то відрізок інтегрування $[a,b]$ можна розбити на скінченну кількість відрізків таким чином, щоб на кожному частковому відрізку функція та її похідні були неперервними. Тоді інтеграл від такої функції можна представити як суму інтегралів за частковими відрізками. Якщо для кожного відрізку застосувати квадратурну формулу порядку $q \le p$ і одночасно та однаково зменшувати крок сітки на них, то порядок точності всього інтегралу також буде рівний q, і методом Рунге-Ромберга його можна підвищити до p.

Нелінійні методи

Для підвищення точності розрахунків часто застосовують нелінійну апроксимацію. У випадку обчислення інтегралу необхідно підбирати таку апроксимуючу функцію, щоб її інтеграл можна було точно обчислити, інакше наближення не буде мати сенсу. Тому при застосуванні апроксимації намагаються відшукати вирівнюючі змінні, у яких вже два вільні параметри забезпечували б задовільне наближення. Для досягнення цієї мети відрізок інтегрування $[a,b]$ розбивають на підінтервали (вводять сітку), і на кожному частковому інтервалі функцію заміщують її нелінійним наближенням, параметри якого виражаються через табличні значення функції.

У випадку функцій, які близькі до експоненти, застосувавши для вирівнювання інтерполяційний многочлен Ньютона, будемо мати:

$$f(x) \approx f_{i-1} \exp[(x - x_{i-1}) \ln(f_i / f_{i-1}) / (x_i x_{i-1})] \quad x \in [x_{i-1}, x_i].$$

Проінтегрувавши на кожному частковому відрізку замість функції $f(x)$ її наближене значення, отримаємо квадратурну формулу:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} (x_i - x_{i-1})(f_i - f_{i-1}) / \ln(f_i / f_{i-1}).$$

Зрозуміло, що наведена квадратурна формула дає хороший результат лише для експоненціальних функцій.

Для побудови наближення часто застосовують інтерполяційний многочлен Лагранжа, який будується для кожного часткового інтервалу окремо, і для виведення квадратурних формул застосовують методи типу середніх (формули Сімпсона або Гаусса не використовують, тому що вони дають досить складний результат).

Інтегрування на змінному проміжку

При обчисленні інтегралу $F(x) = \int_a^x f(\xi)\rho(\xi)d\xi$ для кожного конкретного значення змінної x його можна розглядати як інтеграл з постійними межами і застосовувати описані вище методи. Якщо інтеграл необхідно визначити для великої кількості значень x, то доцільно вибрати сітку і за допомогою високоточних методів інтегрування скласти таблицю значень інтеграла на цій сітці, тоді

$$F(x) = F(x_n) + \int_{x_n}^{x} f(\xi)\rho(\xi)d\xi \quad x_n \le x < x_{n+1},$$

де інтеграл можна обчислювати за простими формулами.

Інтегрування функцій на нескінченному інтервалі

Якщо ваговий коефіцієнт $\rho(x) = e^{-x}$ у формулі (1), то для таких випадків існують спеціальні формули інтегрування на нескінченному інтервалі.

Для інтервалу $(0, \infty)$ справедливою є формула:

$$\int_0^{\infty} e^{-x} f(x)dx \approx \sum_{i=1}^{n} H_i f(a_i), \qquad (2)$$

де коефіцієнти a_i і H_i та похибки відповідних усікань визначені у табл. 1.

Таблиця 1

Коефіцієнти a_i і H_i квадратурної формули (2)

a_i	H_i	похибка
0,585786 3,414214	0,853553 0,146447	$\frac{1}{6} f^{(4)}(c)$
0,415775 2,294280 6,289945	0,711093 0,278518 0,010389	$\frac{1}{20} f^{(6)}(c)$
0,322548 1,745761 4,536620 9,395071	0,603154 0,357419 0,038888 0,000539	$\frac{1}{70} f^{(8)}(c)$
0,263560 1,413403 3,596426 7,085810 12,640801	0,521759 0,398667 0,075942 0,003612 0,000023	$\frac{1}{252} f^{(10)}(c)$

Для вагових коефіцієнтів $\rho(x) = e^{-x^2}$ існує формула чисельного інтегрування на нескінченному проміжку $(-\infty, \infty)$:

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx \approx \sum_{i=1}^{n} H_i' f(a_i'). \qquad (3)$$

Відповідні коефіцієнти a_i' і H_i' та похибки відповідних усікань визначені у таблиці 2 (c – деяке дійсне число).

Таблиця 2

Коефіцієнти a_i і H_i квадратурної формули (3)

	H_i	похибка
±0,707107	0,886227	$\frac{\sqrt{\pi}}{48} f^{(4)}(c)$
0 ±1,224745	1,181636 0,295409	$\frac{\sqrt{\pi}}{960} f^{(6)}(c)$
±0,524648 ±1,650680	0,804914 0,081313	$\frac{\sqrt{\pi}}{26880} f^{(8)}(c)$
	0,945309 0,393619 0,019953	$1?832 \cdot 10^{-6} f^{(10)}(c)$

Наприклад, для $n = 3$

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx \approx 1{,}181636\, f(0) + 0{,}295409\left(f(1{,}224745) + f(-1{,}224745)\right).$$

Інтегрування функцій з розривами на кінцях інтервалу

Якщо підінтегральна функція на кінцях інтервалу інтегрування має розриви (набуває нескінченно великих значень порядку 0,5 відносно величин $\frac{1}{|x-a|}$ і $\frac{1}{|x-b|}$), то за допомогою лінійного перетворення переходять до інтервалу $(-1,1)$ і застосовують формулу:

$$\int_{-1}^{1} f(x)\frac{dx}{\sqrt{1-x^2}} \approx \frac{\pi}{n}\sum_{i=1}^{n} f\left(\cos\frac{(2i-1)\pi}{2n}\right),$$

похибка якої складає $\frac{2\pi}{2^{2n}(2n)!} f^{(2n)}(c)$, де $c \in (-1,1)$.

Якщо функція має особливість лише відносно одного краю інтервалу інтегрування $\frac{1}{|x-a|}$, то доцільно спочатку позбутися цієї

особливості за допомогою відповідних перетворень, а потім виконувати чисельне інтегрування:

$$\int_a^b f(x)\frac{dx}{\sqrt{x-a}} = 2\int_0^{\sqrt{b-a}} f\left(a+t^2\right)dt,$$

або, поклавши $\frac{f(x)}{\sqrt{x-a}} = g(x)$, отримаємо:

$$\int_a^b g(x)dx = 2\int_0^{\sqrt{b-a}} g\left(a+t^2\right)tdt.$$

Кратні інтеграли

Метод прямокутників

В основі методу прямокутників лежить геометричний зміст визначеного інтеграла, як і в аналогічному методі для звичайного інтеграла.

Розглянемо подвійний інтеграл, заданий на прямокутнику $G(a \le x \le b, c \le y \le d)$. Розіб'ємо область інтегрування на однакові прямокутнички (покриємо підінтегральну функцію сіткою). В ролі значення функції можна взяти її значення в середині прямокутничка, і тоді інтеграл можна наближено обчислити за формулою:

$$\iint_G f(x,y)dxdy \approx \sum_i S_i f(\bar{x}_i, \bar{y}_i), \qquad (4)$$

де S_i – площа i-го прямокутничка, $(\bar{x}_i, \bar{y}_i)$ – координати точки, що знаходиться всередині відповідного прямокутничка $\bar{x}_i = 0{,}5(x_{i+1} - x_i)$, $\bar{y}_i = 0{,}5(y_{i+1} - y_i)$ (рис. 6).

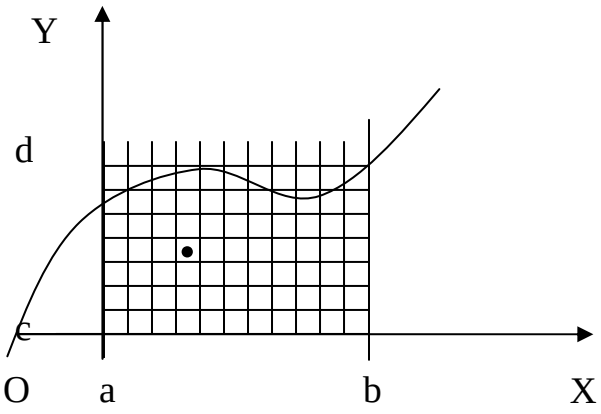


Рис. 6. Покриття площі криволінійної трапеції сіткою

Формула (4) є точною для будь-якої лінійної функції (є апроксимаційною формулою поверхні площиною). Якщо підінтегральну функцію розкласти в ряд Тейлора, отримаємо:

$$f(x,y)=f(\bar{x},\bar{y})+\xi f'_x+\eta f'_y+\frac{\xi^2 f''_{xx}}{2}+\xi\eta f''_{xy}+\frac{\eta^2 f''_{yy}}{2}+\ldots,$$

де $\xi=x-\bar{x}$, $\eta=y-\bar{y}$, а всі похідні беруться в точці центра прямокутничка. Підставивши розкладення у квадратурну формулу, отримаємо оцінку похибки:

$$E=\iint\limits_G f(x,y)dxdy-Sf(\bar{x},\bar{y})\approx\frac{1}{24S\left[(b-a)^2 f''_{xx}+(d-c)^2 f''_{yy}\right]}$$

(всі непарні члени ряду відносно центру симетрії скорочуються). Якщо область розбита на m та n частин відповідно по осі X та Y, то узагальнена формула похибки обчислення інтеграла буде мати вигляд:

$$E\approx\frac{1}{24}\left[\left(\frac{b-a}{m}\right)^2\iint\limits_G f''_{xx}dxdy+\left(\frac{d-c}{n}\right)^2 f''_{yy}dxdy\right]=O\left(m^{-2}+n^{-2}\right).$$

Видно, що формула похибки має другий порядок точності, отже є точною для будь-якої лінійної функції (апроксимує поверхню деякою площиною).

Послідовне інтегрування

Прямокутна область. Розглянемо інтеграл, заданий на прямокутній області, що розбита сіткою на окремі прямокутнички. Такий інтеграл можна обчислити, послідовно інтегруючи його за кожною змінною:

$$I=\int\limits_c^d\int\limits_a^b f(x,y)dxdy=\int\limits_c^d F(y)dy\text{, де }F(y)==\int\limits_a^b f(x,y)dx.$$

Для обчислення кожного простого інтеграла можна застосовувати квадратурні формули. Послідовне інтегрування у двох напрямках приводить до кубатурних формул, які є прямим добутком одномірних квадратурних формул:

$$F(y_i)\approx\sum_i c_i f(x_i,y_i),\ I\approx\sum_j c_j F(y_j),$$

або

$$I \approx \sum_{i,j} c_{ij} f(x_i, y_j), \text{ де } c_{ij} = c_i c_j .$$

У загальному випадку для різних напрямків можна використовувати квадратурні формули різних порядків точності. Тоді головний член похибки можна представити як $T = O\left(h_x^p + h_y^q\right)$, де p і q – порядки точності відповідних квадратурних формул. Цей факт необхідно враховувати у разі застосування методу Рунге-Ромберга: при зменшенні кроку сітки необхідно, щоб відношення h_x^p / h_y^q лишалось постійним. Якщо $p \neq q$, то дотриматись виконання цієї умови непросто, тому бажано для всіх напрямків використовувати квадратурні формули однакового порядку точності.

Якщо за кожним напрямком вибрати квадратурну формулу трапецій і рівномірну сітку, то вагові коефіцієнти $c_{ij} / \left(h_x h_y\right)$ будуть дорівнювати 1, $1/2$, $1/4$ відповідно для внутрішніх, зовнішніх та кутових вузлів сітки; для двічі неперервно диференційовних функцій ця формула буде мати другий порядок точності, і до неї можна застосовувати процедуру Рунге-Ромберга.

Можна підібрати вагові коефіцієнти та сітку таким чином, щоб кожна одномірна квадратурна формула була точною для многочлена максимальної степені, іншими словами – була формулою Гаусса, де

$c_{ij} = \frac{1}{4}(b-a)(c-d)\gamma_i\gamma_j$, $x_i = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)\xi_i$, $y_j = \frac{1}{2}(c+d) + \frac{1}{2}(d-c)\xi_j$, $1 \le i, j \le n$, ξ, γ – нулі многочлена Лежандра і відповідні ваги. Ці формули розраховані на досить гладкі функції і дають для них високий степінь точності для невеликої кількості вузлів.

Довільна область. Розглянемо послідовне інтегрування на довільній області. Для цього покриємо область прямими лініями і розставимо на них вузли (рис. 6). Інтеграл представимо як:

$$I = \iint\limits_G f(x,y)dxdy = \int\limits_c^d F(y)dy, \text{ де } F(y) = \int\limits_{\varphi_i(y)}^{\varphi_2(y)} f(x,y)dx$$

Спочатку обчислимо інтеграл вздовж осі OX для кожної лінії, застосовуючи деяку вибрану квадратурну формулу. Потім будемо обчислювати інтеграл по y, і за вузли будемо брати проекції наших ліній на вісь OY.

При обчисленні інтеграла по y необхідно враховувати, що якщо область обмежена гладкою кривою, то при $y \to c$ ($y \to d$) довжина лінії не лінійно (як $\sqrt{y-c}$) буде прямувати до нуля, отже в околі цієї точки $F(y) \sim \sqrt{y-c}$, а отже використовувати для інтегрування $F(y)$ формули високої точності недоцільно. На практиці з $F(y)$ виділяють основну особливість у вигляді вагового коефіцієнта $\rho(y) = \sqrt{(d-y)(y-c)}$, якому відповідають ортогональні многочлени Чебишова другого роду, і друге інтегрування виконується за формулами Гаусса-Крістоффеля:

Контрольні запитання

1. В чому полягає геометричний зміст визначеного інтегралу?
2. Що називається основною квадратурою?
3. Апроксимація якою функцією лежить в основі формули :
   - трапеції;
   - Сімпсона;
   - Буля?
4. За рахунок чого можна збільшити точність обчислення значення визначеного інтегралу?
5. У чому полягає основна ідея формул Крістоффеля-Гаусса?
6. Який прийом застосовується для інтегрування розривних функцій?
7. В чому полягає ідея виведення формул для кратних інтегралів?

Лекція 12. Нелінійні рівняння ($f(x)=0$)

У багатьох наукових та інженерних задачах математична модель процесу або об‘єкта може бути представлена рівнянням вигляду: $f(x, p_1, p_2, ..., p_n)=0$, де f – задана функція; x – невідома величина; $p_i, i=\overline{1,n}$ – параметри задачі.

Тільки для найпростіших рівнянь можна відшукати розв‘язок аналітичними методами, але частіше для відшукання рішення застосовуються чисельні методи. Класифікація математичних рівнянь відносно складності і можливості аналітичного розв‘язку наведена у таблиці:

Таблиця класифікації рівнянь

Тип	Лінійні рівняння			Нелінійні рівняння		
Рівняння	Одне	Декілька	Багато	Одне	Декіль-ка	Багато
Алгебраїч-не	Тривіаль-но	Легко	Частіше немож-ливо	Дуже важко	Дуже важко	Немож-ливо
Звичайне диферен-ційне	Легко	Важко	Частіше немож-ливо	Дуже важко	Немож-ливо	Немож-ливо
В частин-них похідних	Важко	Частіше немож-ливо	Немож-ливо	Немож-ливо	Немож-ливо	Немож-ливо

Також чисельні методи для розв‘язання нелінійних рівнянь класифікують відносно використання порядку похідної в ітераційному процесі. Так, методи, які не вимагають обчислення похідної, називають методами нульового порядку, які використовують першу похідну – методами першого порядку, другу похідну – методами другого порядку і так далі.

Найпростішим методом відшукання кореня або представлення поведінки функції є побудова таблиці $x, f(x)$. Але такий спосіб є

ненадійним, бо на відрізку табулювання корінь може не існувати, тому для розв'язання задачі розроблені спеціальні методи.

Критерієм вибору методу служить відповідність функції умовам його застосування та кількість операцій, яку необхідно виконати для відшукання кореня. Для різних типів рівнянь різні методи дають різну збіжність, і не існує точної універсальної методики вибору методу для того чи іншого рівняння.

Розглянемо деяку фізичну задачу. Нехай куля радіуса $r = 10$ занурена у рідину на глибину d. Куля зроблена із деревини, яка має щільність $\rho = 0{,}638$. Ставиться питання: яка частина кулі буде у рідині?

Маса витисненої рідини $M_p = \int_0^d \pi\left(r^2 - (x-r)^2\right)dx = \frac{\pi d^2(3r-d)}{3}$,

маса кулі: $M_k = 4\pi r^3\rho/3$. За законом Архімеда $M_p = M_k$, в результаті чого отримаємо наступне рівняння, яке необхідно розв'язати:

$$\frac{\pi\left(d^3 - 3d^2r + 4r^3\rho\right)}{3} = 0.$$

Для нашого випадку при $r = 10$ і $\rho = 0{,}638$: $\frac{\pi\left(2552 - 30d^2 + d^3\right)}{3} = 0$ – поліном 3-го порядку, графік якого наведений на рис. 7.

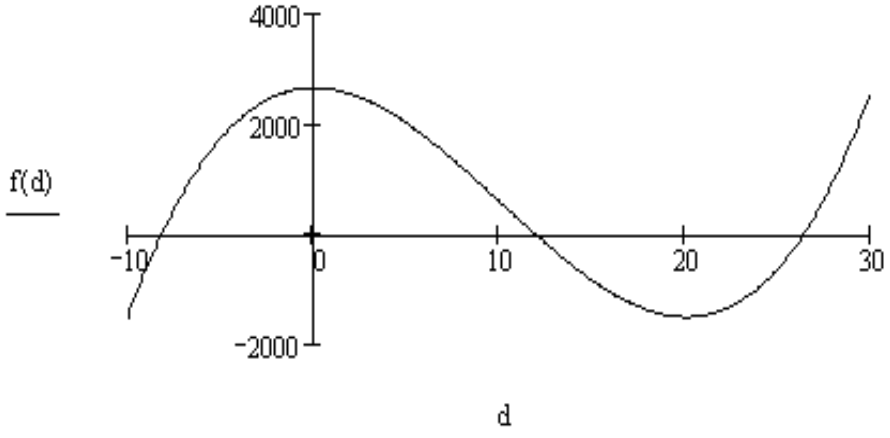


Рис. 7. Графік функції $f(d) = \pi\left(4r^3\rho - 3rd^2 + d^3\right)/3$

З графіка видно, що розв‘язок рівняння наближено $d = 12$. Але ми знаємо, що кубічне рівняння може мати 3 корені, для нашого рівняння це $d_1 = -8{,}1760$, $d_2 = 11{,}8615$ і $d_3 = 26{,}3146$. Зрозуміло, що перший корінь нам не підходить, бо кулі не можуть бути від‘ємними, 3-й корінь також не підходить, бо отриманий розмір кулі більший за діаметр. Отже єдиною правильною відповіддю буде 2-й корінь.

З наведеного прикладу видно, що розв‘язуючи конкретні реальні задачі, необхідно не тільки механічно відшукувати корені, застосовуючи ті чи інші методи, а й вміти аналізувати отримані результати.

Поняття ітерацій та нерухомої точки

Основним засобом розв‘язання задач за допомогою комп‘ютера є ітерація – процес, який циклічно продовжується, поки не буде отриманий результат. Ітерації застосовуються для розв‘язання нелінійних рівнянь, систем рівнянь та розв‘язку диференційних рівнянь. Для застосування ітерацій необхідно вивести рекурентну формулу – правило виконання ітерацій, задати, початкові дані, наприклад, початкове наближення. Так, наприклад, для розв‘язання рівнянь типу $x = g(x)$, задавши початкове наближення x_0, ми отримаємо наступну послідовність:

$$x_0, \quad x_1 = g(x_0), \quad x_2 = g(x_1), \ldots, x_k = g(x_{k-1}), \quad x_{k+1} = g(x_k) \ldots$$

Отже, ми можемо отримати нескінченну послідовність, яка може прямувати до певної границі, або бути розбіжною, або періодичною. Зрозуміло, що для існування розв‘язку і можливості його відшукання необхідно, щоб послідовність ітерацій збігалась. Ця вимога пов‘язана з поняттям нерухомої точки.

Визначення 1. Нерухомою точкою функції $g(x)$ називається таке дійсне число x^*, що $x^* = g(x^*)$. Геометрична інтерпретація нерухомої точки – це точка перетину графіка функції $y = g(x)$ і $y = x$.

Визначення 2. Ітерація $x_{k+1} = g(x_k)$ для $k = 1,2,\ldots$ називається ітерацією нерухомої точки.

Теорема 1. Нехай $g(x)$ – неперервна функція і $\{x_n\}_{n=0}^{\infty}$ – послідовність ітерацій нерухомої точки. Якщо $\lim\limits_{n\to\infty} x_n = x^*$, то x є нерухомою точкою $g(x)$.

Доведення. Якщо $\lim\limits_{n\to\infty} x_n = x^*$, то $\lim\limits_{n\to\infty} x_{n+1} = x^*$. З неперервності $g(x)$ і $x_{k+1} = g(x_k)$ випливає, що $g(x^*) = g\left(\lim\limits_{n\to\infty} x_n\right) = \lim\limits_{n\to\infty} g(x_n) =$ $= \lim\limits_{n\to\infty} x_{n+1} = x^* \Rightarrow$, що x^* – нерухома точка.

Приклад. Розглянемо наступні ітерації $x_0 = 0{,}5$ і $x_{k+1} = e^{-x_k}$, $k = 1,2,\ldots$.

$$x_1 = e^{-0,5} = 0{,}606531$$
$$x_2 = e^{-0,606531} = 0{,}545239$$
$$x_3 = e^{-0,545239} = 0{,}579703$$
$$\ldots$$
$$x_{10} = e^{-0,567560} = 0{,}566907$$

Очевидно, що послідовність збігається, і подальші обчислення показують, що $\lim\limits_{n\to\infty} x_n = 0{,}567143$. Це значення і є наближенням нерухомої точки.

Наступна теорема визначає умову існування нерухомої точки та збіжність ітераційного процесу:

Теорема 2. Нехай $g(x) \in C[a,b]$.

a) якщо область значень $y = g(x)$ задовольняє умові $y \in [a,b]$ $\forall x \in [a,b]$, то функція $g(x)$ має нерухому точку на відрізку $[a,b]$;

b) нехай $g'(x) \in (a,b)$ і існує додатна константа $K < 1$, така що $g'(x) \le K < 1 \quad \forall x \in (a,b)$, тоді $g(x)$ має єдину нерухому точку $x^* \in [a,b]$.

Приклад. Доведемо, що функція $g(x) = \cos(x)$ має строго одну нерухому точку.

Очевидно, що $g \in C[0,1]$ і на цьому інтервалі є спадною функцією, отже, виконується умова теореми а) і функція має нерухому точку на $[0,1]$. Якщо $x \in (0,1)$, то $|g'(x)| = |-\sin(x)| = \sin(x) \le \sin(1) < 0{,}8415 < 1$ – виконується друга умова теореми. Отже, $g(x)$ має єдину нерухому точку на відрізку $[0,1]$.

Теорема 3. Нехай

1) $g(x), g'(x) \in C[a,b]$;
2) $x_0 \in (a,b)$;
3) $K > 0$ – додатна константа;
4) $g(x) \in [a,b] \; \forall x \in [a,b]$.

Тоді:

a) якщо $|g'(x)| \le K < 1 \; \forall x \in [a,b]$ то ітерації $x_{k+1} = g(x_k)$ збігаються до єдиної нерухомої точки $x^* \in [a,b]$. В цьому випадку говорять, що x^* є точкою притягання;

b) якщо $|g'(x)| > 1 \; \forall x \in [a,b]$ то ітерації $x_{k+1} = g(x_k)$ не збігаються до x^*. В цьому випадку говорять, що x^* – нерухома точка відштовхування і ітерації проявляють локальну розбіжність.

Зауваження 1. Вважається , що $x_0 \ne x^*$.

Зауваження 2. Оскільки $g(x)$ неперервна на інтервалі, то допустимо використовувати більш прості критерії $|g'(x)| \le K < 1$ і $|g'(x)| > 1$.

Наслідок. Нехай $g(x)$ задовольняє умові а) теореми 3. Грані похибки, яка виникає при застосуванні наближень x_k для нерухомої точки x^*, задаються формулами:

$$|x^* - x_n| \le K^n |x^* - x_0| \; \forall n \ge 1 \; \textbf{і} \; |x^* - x_n| \le \frac{K^n |x_1 - x_0|}{1 - K} \; \forall n \ge 1.$$

Для того, щоб існував розв‘язок рівняння $x = g(x)$, необхідно, щоб графік кривої $g(x)$ і пряма x перетинались в точці (x^*, x^*). Як було зазначено раніше, ітерації можуть бути збіжними (безпосередньо або збіжність може носити коливальний характер) або розбіжними. Приклади збіжних ітерацій наведені на рис. 8 та 9.

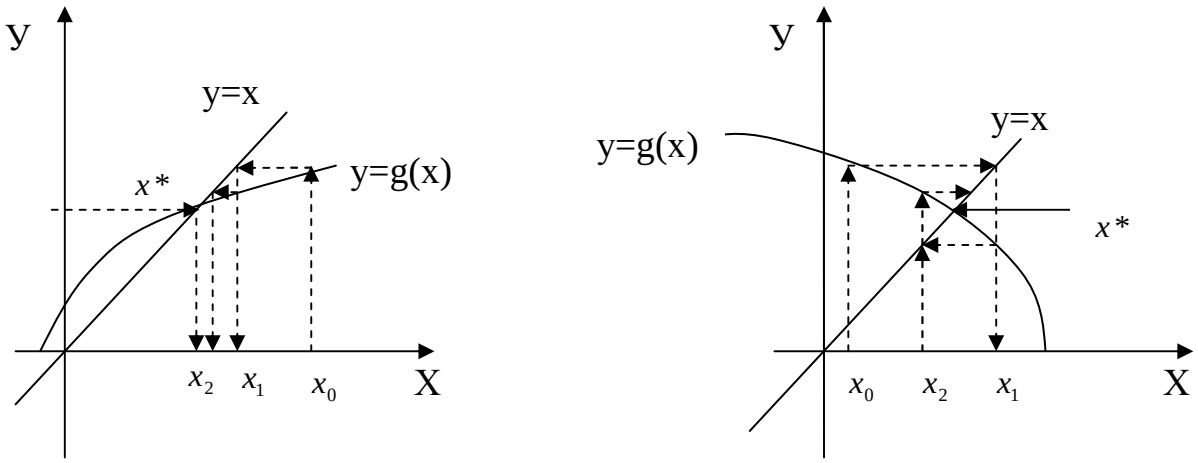


Рис. 8. Графічна інтерпретація збіжних ітерацій

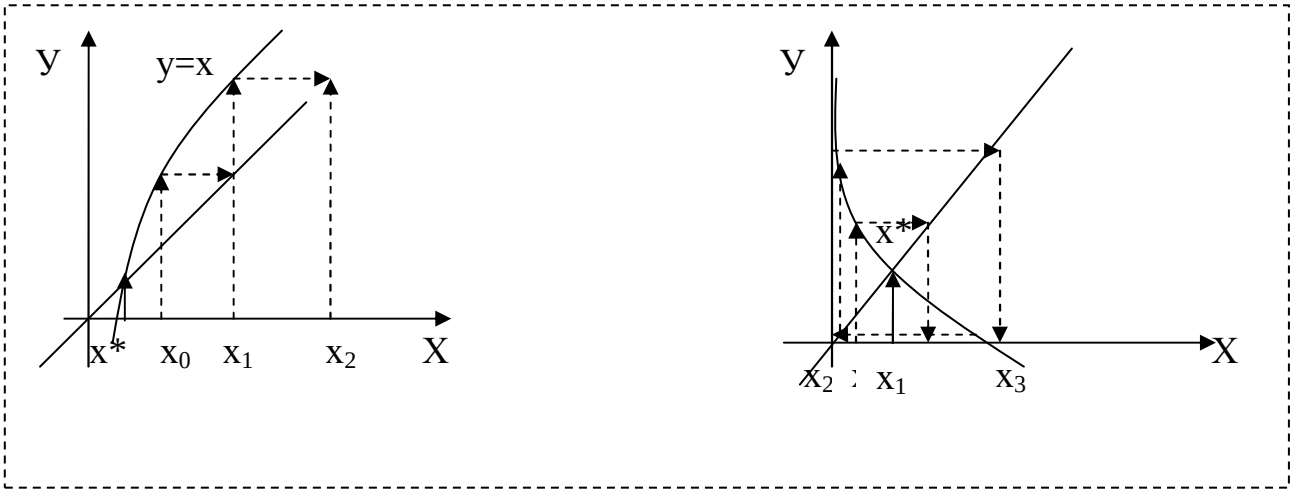


Рис. 9. Графічна інтерпретація розбіжної послідовності простих ітерацій

Приклад. Розглянемо ітерації для функції $g(x)=1+x-x^2/4$. Розв‘язавши рівняння $x=g(x)$, можна знайти нерухомі точки $x_1{}^*=2$ і $x_2{}^*=-2$, похідна від функції $g'(x)=1-x/2$. Розглянемо обидва випадки.

Випадок 1

$x^*=-2$

$x_0=-2{,}05$

$x_1=-2{,}100625$

$x_2=-2{,}20378135$

$x_3=-2{,}41794441$

…….

$\lim\limits_{n\to\infty} x_n=-\infty$

$|g'(x)|>3/2,\ x\in[-3,-1]$

Випадок 2

$x^*=2$

$x_0=1{,}5$

$x_1=1{,}96$

$x_2=1{,}9996$

$x_3=1{,}99999996$

…….

$\lim\limits_{n\to\infty} x_n=2$

$|g'(x)|<1/2,\ x\in[1,3]$

Зауваження. Теорема про збіжність ітерацій нічого не говорить щодо випадку, коли $|g'(x)| = 1$.

Приклад. $x_n = 2(x-1)^{1/2}$ для $x > 1$ існує лише одна нерухома точка $x^* = 2$. $g'(x) = 1$, отже теорему про збіжність послідовності ітерацій застосовувати не можна. Розглянемо два випадки, коли початкове наближення лежить праворуч і ліворуч від нерухомої точки $x^* = 2$:

$x_0 = 1{,}5$	$x_0 = 2{,}5$
$x_1 = 1{,}41421356$	$x_1 = 2{,}44948974$
$x_2 = 1{,}28718851$	$x_2 = 2{,}40789513$
$x_3 = 1{,}07179943$	$x_3 = 2{,}37309514$
$x_4 = 0{,}53590832$	$x_4 = 2{,}34358284$
$x_5 = 2(-0{,}46409168)^{1/2}$	……
x_5 – невизначено	$\lim\limits_{n\to\infty} x_n = 2$

Друга послідовність буде дуже повільно збігатись до нерухомої точки.

Коли задано рівняння $f(x) = 0$, то часто виразити з нього x і записати його у вигляді $x = g(x)$ можна декількома способами. З можливих варіантів необхідно вибирати такий, який буде задовольняти умові збіжності ітерацій.

Приклад. $f(x) = 4 * \sin(x) + 2 * x + 1$.

Це рівняння можна переписати двома способами:

1) $x = -(\sin(x) + 1)/2$ 2) $x = \arcsin(-(2 * x + 1)/4)$ (рис. 10).

У першому випадку $g'(x) = -2\cos(x)$, у другому – $g'(x) = -2/\sqrt{15 - 4x^2 - 4x}$.

З графіків функцій видно, що корінь рівняння лежить на відрізку $[-1{,}0]$ і в точці $x = -0{,}5$ похідна першої функції =-1,755, а другої – -0,5. Отже для першої функції не виконується умова теореми про збіжність ітерацій, і тому для відшукання кореня краще взяти другу функцію.

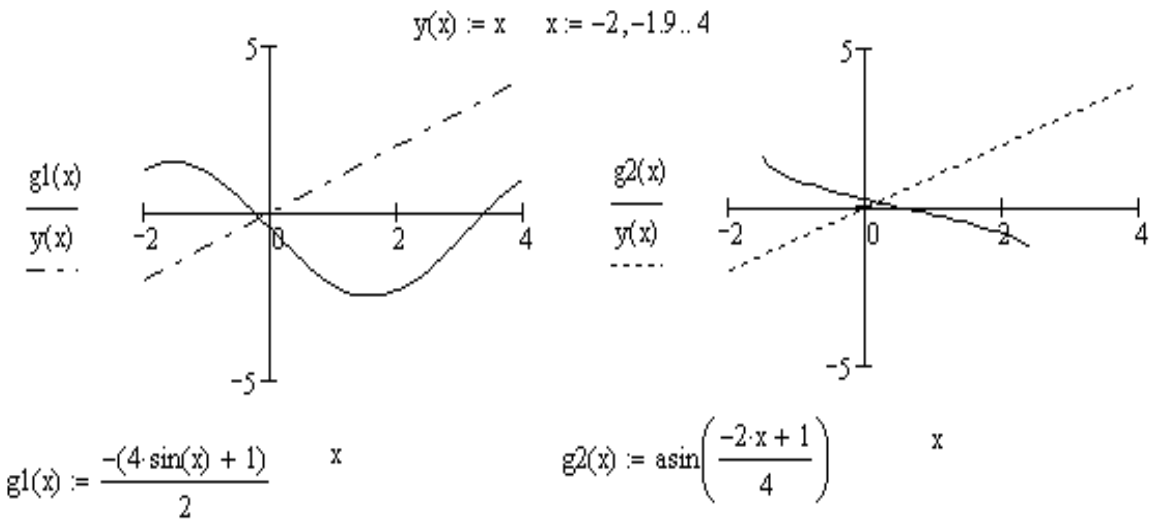


Рис. 10. Графіки функцій $g(x)$

Як критерій зупинки ітераційного процесу можна застосувати нерівність $|x_{n+1} - x_n| < \varepsilon$. Для програмної реалізації методу вхідними даними будуть:

1. Функція $g(x)$.
2. Початкове наближення нерухомої точки $x*$.
3. Точність обчислення нерухомої точки ε.

Вихідними даними будуть:

1. Значення нерухомої точки $x*$.
2. Кількість виконаних операцій.
3. Значення функції в нерухомій точці для контролю результату.

Лекція 13. Методи локалізації коренів ($f(x)=0$)

Нехай $f(x)$ неперервна функція на деякому відрізку $[a,b]$. Будь-яке число $x*$, для якого $f(x*)=0$, називається коренем рівняння $f(x)=0$, або його ще називають нулем функції $f(x)$.

Метод бісекцій (Больцано або дихотомії)

Будемо розглядати рівняння $f(x)=0$, де $f(x)\in C[a,b]$. Відрізок $[a,b]$ необхідно вибирати таким чином, щоб на ньому існував лише один корінь рівняння.

Для методу умовою існування кореня буде: $sign(f(a)) \neq sign(f(b))$ (функція на кінцях має різні знаки, що геометрично означає перетин графіком функції $f(x)$ осі абсцис). Основна ідея методу полягає у зменшенні відрізка шляхом відсікання від нього половини таким чином, щоб локалізувати корінь заданого рівняння.

Отже, основним ітераційним кроком методу є визначення середньої точки відрізку $x = (a+b)/2$ і аналіз наступних ситуацій, які можуть виникнути:

a) $sign(f(a)) \neq sign(f(x)) \Rightarrow$ що корінь $x^* \in [a, x]$;

b) $sign(f(a)) = sign(f(x)) \Rightarrow$ що корінь $x^* \in [x, b]$;

c) $f(x) = 0$, що означає $x^* = x \Rightarrow$ корінь знайдено.

Перші дві ситуаціє є основою ділення відрізка навпіл і вибору необхідної половини, третя – критерієм зупинки ітераційного процесу.

Основою застосування методу є наступна теорема:

Теорема. Нехай $f(x) \in C[a,b]$ і існує таке число $x^* \in [a,b]$, що $f(x^*) = 0$. Якщо $sign(f(a)) \neq sign(f(b))$ і $\{x_n\}_{n=0}^{\infty}$ – послідовність середніх точок, отриманих в результаті поділу відрізка навпіл, то $|x^* - x_n| > (b-a)/2^{n+1}$, а отже, послідовність $\{x_n\}_{n=0}^{\infty}$ збігається до x^* $\lim\limits_{n\to\infty} x_n = x^*$.

Геометрична інтерпретація методу наведена на рис. 11.

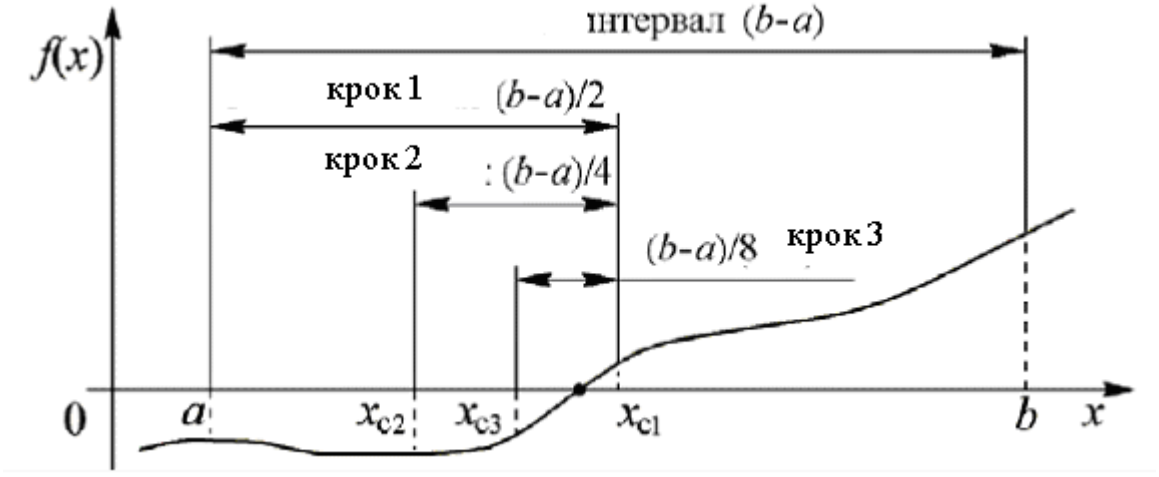


Рис. 11. Геометрична інтерпретація методу поділу навпіл

Перевагою методу бісекцій є те, що він не висуває значних вимог до функції, що задає рівняння (метод нульового порядку) і існує формула, яка дає оцінку точності кореня. Якщо для відшукання розв‘язку виконано N ітерацій, то гарантія того, що похибка наближення менша, ніж наперед задане значення δ, визначається формулою: $N = \dfrac{\ln(b-a) - \ln(\delta)}{\ln 2}$.

Хоч в цілому метод має повільну збіжність, але на відміну від табулювання або методу простих ітерацій кількість кроків ітерацій зменшується у рази.

Для комп‘ютерної реалізації метода вхідними даними є:

1. Вигляд функції, що задає рівняння.
2. Відрізок $[a,b]$.
3. Точність відшукання кореня ε.

Вихідні дані:

1. Значення кореня.
2. Кількість ітерацій для оцінки похибки.
3. Значення функції в точці кореня для контролю.

Метод хорд (хибного положення – regula falsi)

Оскільки метод бісекцій має повільну збіжність, то часто застосовують інший метод нульового порядку – метод хорд. Наближення кореня є кращим, якщо використовувати не середину відрізка, а точку перетину січної, що стягує кінці відрізка з віссю абсцис (рис. 12).

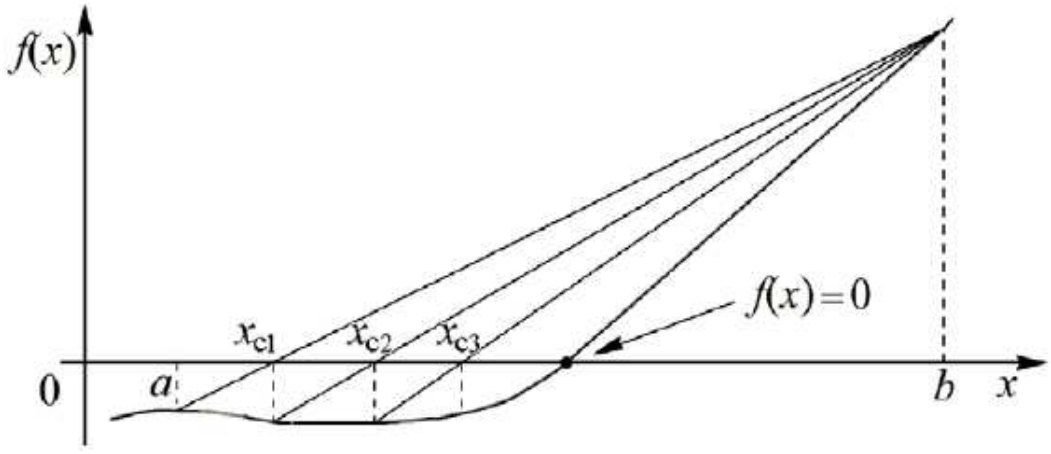


Рис. 12. Геометрична інтерпретація методу хорд

Тангенс кута нахилу хорди $\mathrm{tg}\varphi = \dfrac{f(b)-f(a)}{b-a}$ для точки перетину осі $(x,0)$ і $(b, f(b))$ – $\mathrm{tg}\varphi = \dfrac{0-f(b)}{x-b}$. Прирівнявши рівняння, отримаємо ітераційну формулу: $x = b - \dfrac{f(b)(b-a)}{f(b)-f(a)}$. Як і для методу поділу навпіл, тут існує три варіанти:

d) $sign(f(a)) \neq sign(f(x)) \Rightarrow$ що корінь $x^* \in [a,x]$;

e) $sign(f(a)) = sign(f(x)) \Rightarrow$ що корінь $x^* \in [x,b]$;

f) $f(x) = 0$, що означає $x^* = x \Rightarrow$ корінь знайдено.

Вони будуть лежати в основі ітераційного процесу і служити критерієм його продовження.

Якщо функція вгнута або опукла в околі точки кореня, то один кінець відрізка лишається нерухомим. Критерій зупинення ітерацій $|f(x)| < \varepsilon$ вже непридатний для завершення ітерацій методу хорд, і кращим критерієм буде $|b-a| < \varepsilon$ – критерій виродження відрізка у точку.

Метод золотого перерізу

Ще одним методом нульового порядку локалізації кореня є метод золотого перерізу, який також дає кращу збіжність, ніж метод бісекцій. В загальному випадку цей метод застосовується для відшукання локальних екстремумів, але якщо рівняння $f(x) = 0$ модифікувати, взявши ліву частину по модулю $|f(x)|$, або піднести до квадрата $f^2(x)$, то точка кореня для модифікованої функції стане точкою мінімуму, і метод можна застосувати до знаходження кореня як мінімуму функції.

Як і в попередніх методах, $f(x) \in C[a,b]$ і задається відрізок $[a,b]$, на якому існує корінь. Необхідно зауважити, що на відміну від попередніх методів, на відрізку може існувати декілька коренів, але метод може відшукати лише один.

Основна ідея методу теж полягає у звуженні відрізка, але відсікається вже не половина, а частка у відношенні золотого перерізу. Коефіцієнт золотого перерізу береться рівним приблизно $\Delta = 0{,}38$, і алгоритм полягає у виконанні наступних ітерацій (рис. 13):

a) знаходяться точки $x_1 = a + \Delta|b - a|$ і $x_2 = b - \Delta|b - a|$;

b) в цих точках порівнюється значення функції і відсікається той підінтервал, де функція має більше значення: $|f(x_1)| > |f(x_2)|$ то точка $a = x_1$, якщо – $|f(x_1)| < |f(x_2)|$, то $b = x_2$.

Критерієм припинення ітерацій доцільно взяти умову $|b - a| < \varepsilon$.

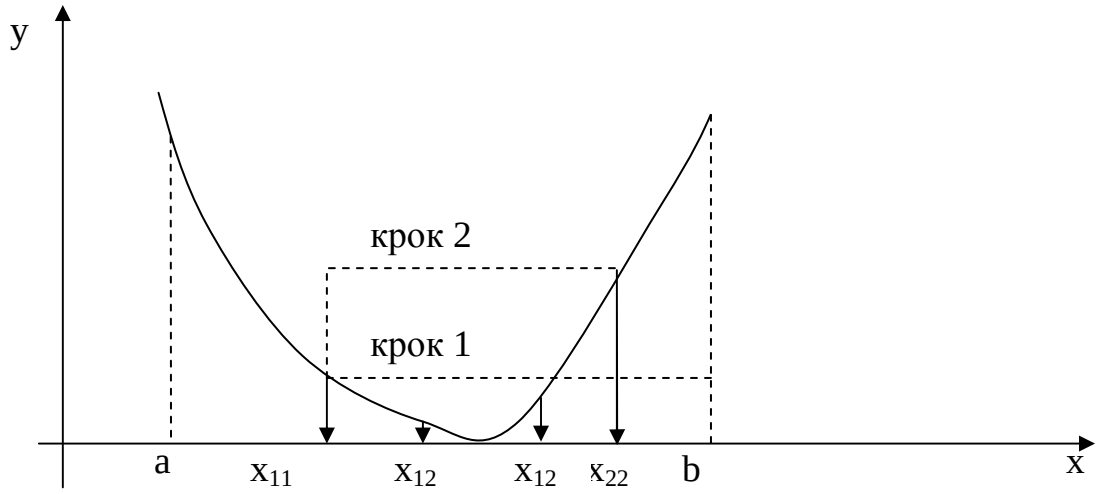


Рис. 13. Геометрична інтерпретація методу золотого перерізу

Аналіз методів локалізації кореня

Початкове наближення

Методи локалізації кореня можна продовжувати, поки він не буде знайдений, тому вони називаються глобально збіжними. Але для застосування методів необхідно вибирати інтервали, на яких існує лише один корінь, і якщо рівняння мають декілька коренів на малих інтервалах, то задача підбору інтервалів стає досить складною. Тому, якщо задача розв‘язання нелінійного рівняння є частиною проекту, доцільно спочатку побудувати графік функції, на основі якого прийняти необхідні рішення стосовно початкових наближень. При комп‘ютерній побудові графіків необхідно бути обережним, бо вони

відтворюються з певними спотворюваннями, що важливо враховувати у тих випадках, коли корені співпадають або розташовані досить близько або функція має локальний екстремум, близький до нуля.

Перевірка збіжності

Як було вже розглянуто у наведених методах, за критерій зупинки ітерацій можна взяти наступні критерії (рис. 14, 15):

1. $|f(x)| < \varepsilon$;
2. $|b-a| < \varepsilon$, де значення кінців інтервалу на кожному кроці ітерацій прямують один до одного.

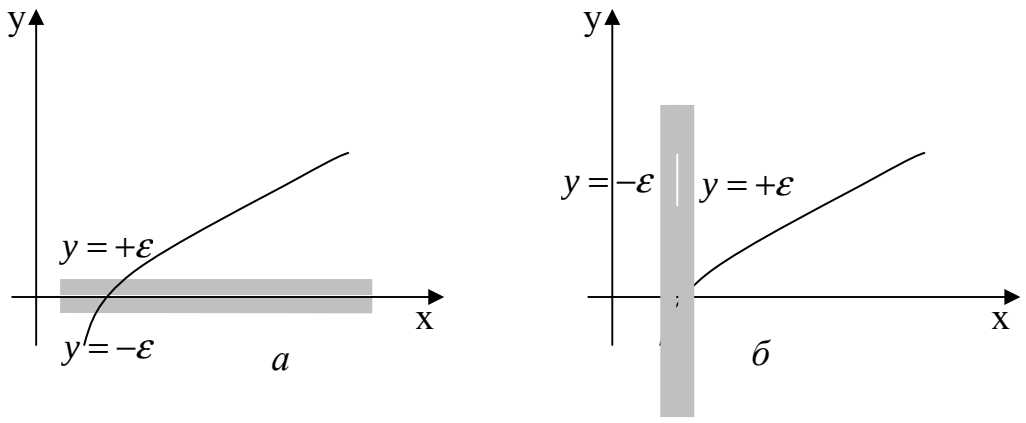


Рис. 14. Геометрична інтерпретація критеріїв збіжності:
а) $|f(x)| < \varepsilon$; *б*) $|b-a| < \varepsilon$

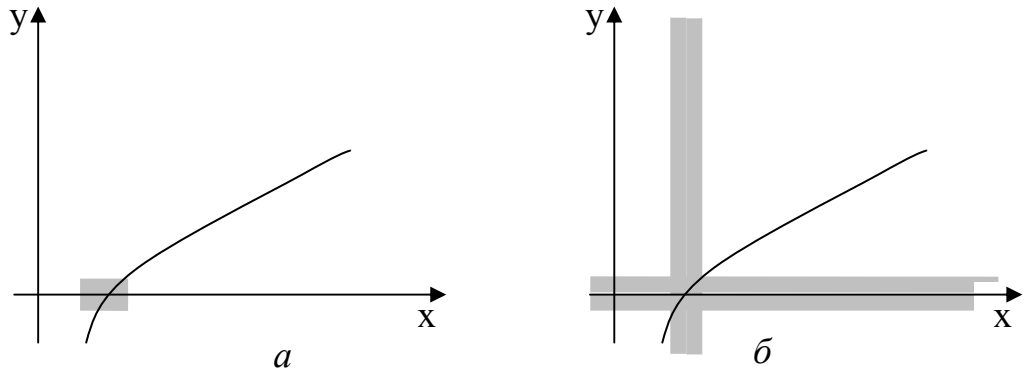


Рис. 15. Комбіновані критерії збіжності:
а) $|f(x)| < \varepsilon$ і $|b-a| < \varepsilon$; *б*) $|f(x)| < \varepsilon$ або $|b-a| < \varepsilon$ – необмежена область

Як видно з рисунків, якщо за ε взяти дуже мале значення, то ітерації можуть продовжуватися безкінечно. Це значення доцільно вибирати в 100 разів більше за 10^{-m}, де m – число десяткових знаків для вибраного типу даних.

Розв‘язок рівняння $f(x)=0$ містить помилку, обумовлену округленням проміжних обчислень або нестійкістю обчислень. Якщо графік функції в околі точки кореня досить крутий, то критерій $|f(x)|<\varepsilon$ може призвести до великої кількості операцій (інтервал буде майже рівний 0, а значення функції – більше припустимої точності). Якщо графік функції, навпаки, пологий – то критерій $|f(x)|<\varepsilon$ може виконуватись вже в точці, досить віддаленій від його істинного розташування. Такі функції досить важкі для застосування чисельних методів і по можливості вимагають нормування.

Метод Ньютона-Рафсона (дотичних)

Недоліком розглянутих раніше методів є їхня досить повільна збіжність. Значно прискорити збіжність ітерацій для відшукання коренів можна, застосувавши метод першого порядку Ньютона-Рафсона (або просто Ньютона). Застосування методу вимагає, щоб функція $f(x)\in C_2[a,b]$ була двічі неперервно диференційованою на відрізку $[a,b]$ і бажано, щоб на відрізку відшукання кореня не існувало критичних точок $f'(x)\neq 0$.

Ітераційна формула методу Ньютона випливає з геометричного змісту похідної. Тангенс кута нахилу дотичної до графіка функції $f(x)$, що проходить через точки x_0, x_1, можна записати двома способами:

$$\operatorname{tg}\varphi=\frac{0-f(x_0)}{x_1-x_0} \text{ і } \operatorname{tg}\varphi=f'(x_0).$$

Прирівнявши обидва значення, отримаємо ітераційну формулу: $x_{n+1}=x_n-f(x_n)/f'(x_n)$, $n=1,2,\ldots$, яка за визначених вище умов породжує послідовність, що збігається до кореня рівняння (рис. 16, *а*)). Але як видно з ітераційної формули, метод Ньютона на кожному кроці ітерацій вимагає більше обчислень (не тільки значення

функції, а й її похідної). Незважаючи на це, метод Ньютона має квадратичну збіжність, що означає, що на кожній ітерації похибка зменшується за квадратичним законом (іншими словами, кількість вірних значущих цифр подвоюється). Так, якщо на певному кроці досягнута похибка рівна 0,5, то за п‘ять, шість кроків вона зменшиться на 2^{-64}. Для методу бісекцій для досягнення такого результату необхідно було б на порядок більше кроків.

Визначення порядку збіжності. Нехай $\{x_n\}_{n=0}^{\infty}$ збігається до x^* і $E_n = x^* - x_n$ для $n \geq 0$. Якщо існують такі константи $A \neq 0$ і $R > 0$, що $\lim\limits_{n\to\infty} \dfrac{|x^* - x_{n+1}|}{|x^* - x_n|^R} = \lim\limits_{n\to\infty} \dfrac{|E_{n+1}|}{|E_n|^R} = A$, то говорять, що послідовність $\{x_n\}_{n=0}^{\infty}$ збігається з порядком збіжності R, а число A називається постійною асимптотичною помилкою. Якщо $R = 1$, то збіжність називається лінійною (методи локалізації кореня), якщо $R = 2$ – то квадратичною (метод Ньютона).

Збіжність методу можна навіть прискорити при відшуканні кратних коренів, скориставшись формулою: $x_{n+1} = x_n - Mf(x_n)/\, f'(x_n)$, M – кратність кореня.

Вихідними даними для методу Ньютона є:

1. Вигляд функції, що задає рівняння.
2. Початкове наближення x_0 (немає потреби визначати відрізок існування одного кореня, що також є перевагою методу).
3. Точність знаходження кореня ε.

При застосуванні методу Ньютона необхідно брати до уваги певні особливості його застосування, а саме: метод є локально збіжним і збіжність залежить від того, наскільки вдало вибране початкове наближення. Якщо початкове наближення є „поганим”, то метод може розбігатись і корінь не буде знайдено. Також при існуванні в околі точки кореня критичних точок ($f'(x) = 0$) може призвести до фатальної помилки ділення на 0. Також небажана ситуація може виникнути, коли похідна має досить мале значення, тоді результатом ділення на неї буде велике число.

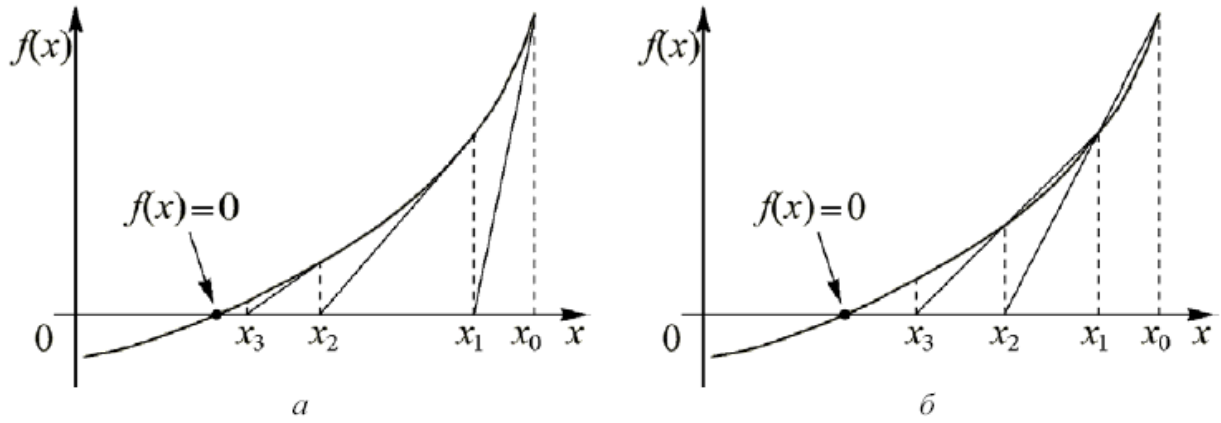


Рис. 16. Метод Ньютона і метод січних

Модифікації методу Ньютона (метод січних, метод Рибакова)

Щоб уникнути обчислення на кожному кроці похідної та недоліків, пов'язаних з її значенням, на практиці застосовують модифікацію методу Ньютона яка має назву методу „січних" (рис. 16, *б*)). В методі січних похідна заміщується її наближеним значенням: $f'(x)=\lim\limits_{\Delta x\to 0}\dfrac{f(x+\Delta x)-f(x)}{\Delta x}\Rightarrow F'(x)=\dfrac{f(x_{k+1})-f(x_k)}{x_{k+1}-x_k}$.

Тоді основна ітераційна формула буде мати наступний вигляд:

$$x_{n+1}=x_n-\frac{f(x_n)(x_n-x_{n-1})}{f(x_n)-f(x_{n-1})}.$$

В методі Рибакова похідна заміщується деяким числом $M\geq f'(\xi)$, де ξ є значенням x, при якому $f'(x)$ максимальне. При такому заміщенні збіжність не порушується, але трохи сповільнюється. Основна ітераційна формула має вигляд: $x_{n+1}=x_n+|f(x_n)|/M$. Кількість ітерацій для методу Рибакова $N=(b-a)M/\varepsilon$. Достоїнством застосуванням методу Рибакова є те, що функція може мати похідну з розривами першого роду.

Метод Ейткена-Стеффенсона

Для прискорення збіжності послідовності ітерацій знаходження кореня нелінійного рівняння застосовують процес Ейткена.

Теорема (Прискорення Ейткена). Нехай послідовність $\{x_n\}_{n=0}^{\infty}$ лінійно збігається до границі x^* і $x^*-x_n\neq 0$ для всіх $n\geq 0$. Якщо

існує таке число $A \in R$, що $|A<1|$ і $\lim\limits_{n\to\infty}\dfrac{(x^*-x_{n+1})}{x^*-x_n}=A$, то послідовність $\{y_n\}_{n=0}^{\infty}$, що визначається формулою $y_n = x_n - \dfrac{(x_{n+1}-x_n)^2}{x_{n+2}-2x_{n+1}+x_n}$, збігається більш швидко в тому сенсі, що $\lim\limits_{n\to\infty}\dfrac{(x^*-y_n)}{x^*-x_n}=0$.

Об‘єднання процесу Ейткена з ітерацією нерухомої точки дає рекурентну формулу для методу Ейткена-Стеффенсена:

$x_{n+1}=\dfrac{x_0x_2-x_1^2}{x_0-2x_1+x_2}$, де x_0 – початкове наближення, і на першому кроці $x_n=x_0$, $x_1=f(x_0)$ і $x_2=f(x_1)$.

Треба зазначити, що для складних функцій цей метод дає значно кращу збіжність, ніж інші, раніше розглянуті методи, але для простих функцій розрахунки майже не полегшуються, що пов‘язане з додатковими діями в ітераційній формулі.

Метод Уолла

Для досить складних і гладких функцій можна застосовувати метод Уолла, який дає кубічну збіжність, але накладає на функції додаткові умови диференційованості, бо до його рекурентної формули входить друга похідна (метод другого порядку). Основна рекурентна формула має вигляд:

$$x_{n+1}=\frac{f(x_n)}{f'(x_n)-f(x_n)f''(x_n)\big/2f'(x_n)}.$$

Контрольні запитання

1. Яка точка називається нерухомою?
2. Чи завжди можна розв’язати рівняння методом простих ітерацій?
3. Як перевірити збіжність методу простих ітерацій?
4. Яка умова застосування методу:

- дихотомії;
- золотого перерізу;
- Ньютона?

5. В яких випадках метод Ньютона може не знаходити кореня, і як ці ситуації можна подолати?
6. В чому різниця застосування методу Ньютона та методу січних?
7. Який порядок збіжності має метод хибного положення?
8. Який метод має найвищий порядок збіжності і які умови його застосування?
9. Які критерії застосовують для зупинки ітерацій при відшукуванні коренів рівнянь?

Лекція 14. Системи лінійних та нелінійних рівнянь

Лінійні рівняння. Основні поняття

Методи розв‘язання систем лінійних рівнянь поділяють на точні та наближені. До точних методів відносяться метод Крамера, розв‘язання системи матричним рівнянням (за допомогою оберненою матриці) і метод Гаусса. Хоч метод Крамера здається досить простим, але в чисельних методах він не застосовується. Це пов‘язано з великою кількістю операцій, які необхідно виконати для відшукання коренів системи рівнянь, а отже, і значною накопиченою похибкою. Говорячи про системи лінійних рівнянь та методи їх розв‘язку на практиці, необхідно зазначити, що ці задачі пов‘язані з системами великих розмірностей, що обумовлює певні проблеми їх розв‘язання. В чисельних методах розглядається така характеристика матриць, як обумовленість, – міра чутливості розв‘язку системи до збурень у правій частині. Також надзвичайно важливим аспектом комп‘ютерної реалізації методів є різниця порядків коефіцієнтів матриці, що може призвести до помилок зникнення порядку.

Визначення. Матриця $A = \lfloor a_{ij} \rfloor$ розмірності $n \times n$ називається верхньою трикутною матрицею, якщо її елементи $a_{ij} = 0, i > j$.

Матриця $A = \lfloor a_{ij} \rfloor$ розмірності $n \times n$ називається нижньою трикутною матрицею, якщо $a_{ij} = 0, i < j$.

Розв‘язання трикутних матриць є досить простою задачею. Для розв‘язання верхньої трикутної матриці застосовують зворотну підстановку. З останнього рівняння відшукується значення $x_n = b_n / a_{nn}$, яке потім підставляється в $n-1$ рівняння, і з нього знаходиться $x_{n-1} = (b_{n-1} - a_{n-1n} x_n) / a_{n-1n-1}$ і так далі $x_k = \dfrac{b_k - \sum_{j=k+1}^{n} a_{kj} x_j}{a_{kk}}$, $k = n-1, n-2, ...1$. Аналогічно можна розв‘язати і систему рівнянь, яка представлена нижньою трикутною матрицею, але тільки процедура обчислення починається не з кінця, а з першого рівняння.

Говорять, що системи рівнянь є еквівалентними, якщо вони мають однакову множину розв‘язків.

Теорема. Елементарні перетворення:

- перестановки – рівняння можна міняти місцями;
- масштабування – множення рівняння на число, не рівне 0;
- заміщення – рівняння можна замінити сумою самого рівняння і будь-якого іншого – приводять до еквівалентної системи рівнянь.

Метод Гаусса

Розглянемо систему лінійних алгебраїчних рівнянь розмірності $n \times n$:

$a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1$

$a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2$

.....

$a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n = b_n$

У матричному вигляді: $A\bar{x} = b$. Матриця A є невиродженою і на головній діагоналі відсутні елементи, рівні 0.

Визначення. Коефіцієнт матриці $a_{kk} \neq 0$, який використовується для виключення елементів a_{ik}, називається головним, а k-й рядок матриці – головним рядком.

Метод Гаусса складається з двох частин:

1. Прямий хід – приведення початкової матриці до еквівалентної верхньої трикутної.
2. Зворотний хід – розв‘язання верхньої трикутної матриці.

Прямий хід. Нехай a_{11} – головний елемент. Розділивши перше рівняння на головний елемент, ми отримаємо: $x_1 + a_{12}^{(1)} x_2 + ... + a_{1n}^{(1)} x_n = b_1^{(1)}$, де $a_{1j}^{(1)} = a_{1j} / a_{11}$. Домножуючи послідовно отримане рівняння на a_{i1} кожного наступного рядка $(i = 2,3,...n)$ і віднімаючи від цього рівняння, отримаємо в першому стовпчику нулі:

$$\begin{array}{l} x_1 + a_{12}^{(1)} x_2 + ... + a_{1n}^{(1)} x_n = b_1^{(1)} \\ 0 + a_{22}^{(1)} x_2 + ... + a_{2n}^{(1)} x_n = b_2^{(1)} \\ ..... \\ 0 + a_{n2}^{(1)} x_2 + ... + a_{nn}^{(1)} x_n = b_n^{(1)} \end{array}.$$

Далі, послідовно обираючи в ролі ведучого головного елемента a_{ii}, де $i = \overline{2, n-1}$, і повторюючи попередні дії для наступних рядків, в результаті отримаємо верхню трикутну матрицю, еквівалентну заданій:

$$\begin{array}{l} x_1 + a_{12}^{(n-1)} x_2 + ... + a_{1n}^{(n-1)} x_n = b_1^{(n-1)} \\ 0 + x_2 + ... + a_{2n}^{(n-1)} x_n = b_2^{(n-1)} \\ ..... \\ 0 + 0 + ... + a_{nn}^{(n-1)} x_n = b_n^{(n-1)} \end{array},$$

де на кожному кроці нові елементи матриці будуть визначатись за рекурентною формулою: $a_{ij} = a_{ij} - a_{kj} * a_{ik} / a_{kk}$.

Для того щоб в комп‘ютерній реалізації не виконувати зайвих операцій, достатньо зробити перетворення лише коефіцієнтів верхньої трикутної матриці, а коефіцієнти, що лежать під головною діагоналлю, можна лишити без змін. Тоді програмно реалізувати

метод Гаусса можна за допомогою трьох, узгоджених за значенням лічильників, вкладених покрокових циклів:

- for k := 1 to n - 1 do – номер головного рядка;
- for i := k + 1 to n do – перебір рядків, що лежать нижче головного;
- for j := k + 1 to n do – перебір стовпчиків (якщо розглядається розширена матриця – то n+1).

$a_{ij} = a_{ij} - a_{kj} * a_{jk} / a_{kk}$ – рекурентна формула перетворення коефіцієнтів матриці.

Загальна кількість операцій для виключення x_1 складає $N_2 = n + 2n(n-1)$, аналогічно можна підрахувати кількість операцій на інших кроках (вони будуть виражатись рекурентною формулою $N_i = 2(n-i+1)^2 - (n-i+1)$). Тоді загальна кількість операцій прямого ходу буде становити:

$$N = \sum_i N_i = \sum_{i=1}^{n} \left(2(n-i+1)^2 + (n-i+1)\right) = \frac{2}{3}n^3 + \frac{n^2}{2} - \frac{n}{6}.$$

Якщо розглядати тільки праві частини матриці, то кількість операцій перетворення матриці буде становити: $M = n(n+1) - n = n^2$.

Зауваження. Якщо при прямому ході деякий діагональний елемент матриці прийме нульове значення, то даний спрощений варіант методу Гаусса непридатний для розв‘язання системи рівнянь.

Зворотний хід. Зворотній хід, в якому починаючи з кінця визначаються невідомі x_i, задається формулами:

$$x_n = a_{nn+1}^{(n)},\ x_i = a_{in+1}^{(n)} - \sum_{j=i+1}^{n} a_{ij}^{(n)} x_j, \quad i = n-1, n-2, \ldots 1$$

Кількість операцій множення та віднімання на зворотному ході дорівнює: $N^* = n(n-1) = n^2 - n$.

Отже, загальна кількість операцій, які треба виконати для розв‘язання системи лінійних рівнянь методом Гаусса, буде становити:

$$N_{\text{заг}} = N + N^* = \frac{2}{3}n^3 + \frac{3}{2}n^2 - \frac{7}{6}n \approx \frac{2}{3}n^3.$$

Виконуючи перетворення матриці окремо для правої і лівої частини, метод можна ефективно застосувати для розв‘язання цілого сімейства систем лінійних рівнянь, які відрізняються лише правими частинами. Якщо задано k таких систем, то загальна кількість операцій буде становити: $N = \frac{2}{3}n^3 - \frac{n^2}{2} - \frac{n}{6} + kn(2n-1) \approx \frac{2}{3}n^3 + 2kn^2$.

Обчислення визначника

Прямий хід методу Гаусса над нерозширеною матрицею можна застосовувати для обчислення визначника. Визначник матриці при діленні її рядка на головний елемент також ділиться на цей елемент, а при відніманні з будь-якого іншого рядка, помноженого на деяке число, визначник не змінюється. В результаті прямого ходу буде отримана трикутна матриця:

$$\begin{matrix} x_1 + a_{12}^{(n-1)}x_2 + \ldots + a_{1n}^{(n-1)}x_n = b_1^{(n-1)} \\ 0 + x_2 + \ldots + a_{2n}^{(n-1)}x_n = b_2^{(n-1)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 0 + 0 + \ldots + a_{nn}^{(n-1)}x_n = b_n^{(n-1)} \end{matrix}$$

, де на головній діагоналі стоять одиниці, отже, якщо в процесі перетворення $a_{ii}^{(k)} \neq 0$, то визначник можна обчислити:

$$\det a = a_{11}^{(0)} a_{22}^{(1)} a_{33}^{(2)} \ldots a_{nn}^{(n-1)}.$$

Обчислення оберненої матриці.

Якщо задана невироджена матриця $A = (a_{ij})$, то j-й стовпчик оберненої матриці A^{-1} співпадає зі стовпчиком $(x_{1j}, x_{2j}, \ldots, x_{nj})^T$ ($(x_{1j}, x_{2j}, \ldots, x_{nj})$ – розв‘язок системи) $Ax_j = \delta_j$ (δ – символ Кронекера $\delta_j = (\delta_{1j}, \delta_{2j}, \ldots, \delta_{nj})$, де $\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$).

Отже, для відшукання оберненої матриці необхідно розв‘язати n таких систем, які відрізняються лише правими частинами. При цьому загальна кількість операцій $N \approx 8/3\, n^3$.

Вибір головного елемента. Для зменшення похибки обчислень та уникнення появи нульового елемента на головній діагоналі на практиці застосовується стратегія з вибором головного елемента. Ця стратегія полягає у тому, що на кожному кроці шляхом перестановки двох рядків або стовпчиків найбільший за абсолютним значенням коефіцієнт $\max|a_{ij}|$ ставиться на головну діагональ і стає головним елементом. Така модифікація методу зменшує розповсюдження похибки округлення на кожному кроці обчислень.

Норма та обумовленість матриць

Введемо поняття норми вектора в лінійному просторі R^n двома способами:

$$\|x\| = \begin{cases} \|x\|_1 = \max|x_i| \\ \|x\|_2 = (x,x)^{1/2} = \sqrt{\sum_{i=1}^{n} x_i^2} \end{cases}$$

В лінійному просторі квадратних числових матриць n-го порядку задамо норму матриці формулою $\|A\| = \sup\limits_{x \neq 0} \dfrac{\|Ax\|}{\|x\|}$. Введена таким чином норма матриці називається узгодженою з нормою вектора.

$$\|A\|_1 = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|, \quad \|A\|_2 \le \left(\sum_{i,j=1}^{n} a_{ij}^2 \right)^{1/2}.$$

Приклад.

$$A = \begin{pmatrix} 4 & 3 & -1 \\ -2 & -4 & 5 \\ 1 & 2 & 6 \end{pmatrix} \quad \|A\|_1 = 4+4+6 = 14, \; \|A\|_2 = \sqrt{112}.$$

Якщо матриця A симетрична, то $\|A\|_2 = \max\limits_{1 \le i \le n} |\lambda_i(A)|$, $\left\|A^{-1}\right\|_2 = \dfrac{1}{\min\limits_{1 \le i \le n} |\lambda_i(A)|}$, де λ_i – власні числа матриці A.

Якщо A і B – квадратні матриці, то для них виконуються наступні співвідношення:

$$\|A+B\| \le \sup_{x\neq 0}\frac{\|Ax\|}{\|x\|} + \sup_{x\neq 0}\frac{\|Bx\|}{\|x\|} = \|A\| + \|B\|,\ \|Ax\| \le \|A\|\cdot\|x\|,\ \left\|A^k\right\| \le \|A\|^k.$$

Розглянемо систему лінійних алгебраїчних рівнянь $Ax = b$, де $\det A \neq 0$ і $b \neq 0$, отже система має єдиний розв‘язок. На практиці при розв‘язанні системи будь-яким методом (в тому числі і методом Гаусса) обчислення виконуються з округленням і отриману похибку можна інтерпретувати як похибку правої частини: $A(+r)x = b + \eta$, r – похибка розв‘язку, а η – похибка правої частини. Тоді $Ar = \eta$ і $r = A^{-1}\eta$ Відношення відносної похибки розв‘язку $\|r/\|x\|\|$ до відносної похибки правої частини $\|\eta\|/\|b\|$

$$\frac{\|r\|/\|x\|}{\|\eta\|/\|b\|} = \frac{\|r\|\cdot\|b\|}{\|\eta\|\cdot\|x\|} = \frac{\|Ax\|}{\|x\|}\frac{\left\|A^{-1}\eta\right\|}{\|\eta\|} \le \frac{\|A\|\cdot\|x\|}{\|x\|}\frac{\left\|A^{-1}\right\|\cdot\|\eta\|}{\|\eta\|} = \|A\|\cdot\left\|A^{-1}\right\| = \nu(A).$$

Величина $\nu(A) = \|A\|\cdot\left\|A^{-1}\right\|$ називається мірою обумовленості матриці A. Міра обумовленості дорівнює максимально можливому коефіцієнту підсилення відносної похибки від правої частини до розв‘язку системи. Якщо матриця симетрична, то $\nu(A) = \dfrac{\max\limits_{1\le i\le n}|\lambda_i(A)|}{\min\limits_{1\le i\le n}|\lambda_i(A)|}$.

Якщо значення $\nu(A)$ велике, то матриця A називається погано обумовленою, в протилежному випадку ($\nu(A)$ не велике) – добре обумовленою.

Приклад погано обумовленої системи. Розглянемо систему:

$$A = \begin{pmatrix} 1 & -1 & -1 & \ldots & -1 & -1 \\ 0 & 1 & -1 & \ldots & -1 & -1 \\ 0 & 0 & 1 & \ldots & -1 & -1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & -1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix},\ \det A \neq 0 \text{ і має єдиний розв‘язок } x = (0,0,\ldots,1).$$

При розв‘язанні системи методом Гаусса прямий хід відсутній. Нехай у зворотному ході була припущена єдина похибка – замість $x_n = 1$ підставили $\tilde{x}_n = 1 + \varepsilon$, де $\varepsilon \neq 0$ – досить мале число. Тоді в результаті розв‘язку системи отримаємо $\tilde{x} = x + r$, де $r = (r_1, r_2, ..., r_n)$ – похибка, що задовольняє системі рівнянь

$$\begin{aligned} r_1 - r_2 - r_3 - ... - r_n &= 0 \\ r_2 - r_3 - ... - r_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots & , \\ r_{n-1} - r_n &= 0 \\ r_n &= 0 \end{aligned}$$

звідки отримаємо:

$$r_n = \varepsilon$$

$$r_{n-1} = r_n = \varepsilon$$

$$r_{n-2} = r_n + r_{n-1} = \varepsilon + \varepsilon = 2\varepsilon$$

$$r_{n-3} = r_n + r_{n-1} + r_{n-2} = \varepsilon + \varepsilon + 2\varepsilon = 2^2\varepsilon$$

……

$$r_{n-k} = r_n + r_{n-1} + \ldots + r_{n-(k-1)} = 2^{k-1}\varepsilon$$

……

$$r_1 = r_{n-(n-1)} = 2^{(n-1)-1}\varepsilon = 2^{n-2}\varepsilon$$

Тоді $\|r\|_1 2^{n-2}|\varepsilon|$, $\|x\|_1 = 1$, $\|\eta\|_1 = |\varepsilon|$, $\|b\|_1 = 1$, і тоді міра обумовленості такої системи буде $\nu(A) = \|A\|_1 \cdot \|A^{-1}\|_1 \geq \dfrac{\|r\|_1 / \|x\|_1}{\|\eta\|_1 / \|b\|_1} = 2^{n-2}$.

Так, наприклад, якщо $n = 102$, то $\nu(A) \geq 2^{100} > 10^{30}$, а $\|r\|_1 = 2^{100}|\varepsilon| > 10^{30}|\varepsilon|$, і якщо навіть допущена похибка в зворотному ході була дуже малою $\varepsilon = 10^{-15}$, то похибка отриманого результату буде досить значною $\|r\|_1 > 10^{15}$.Така значна похибка є наслідком поганої обумовленості матриці.

Необхідно зауважити, що матриця великої розмірності завжди погано обумовлена.

Метод LU-розкладу

Невироджену матрицю A завжди можна представити у вигляді добутку нижньої трикутної матриці L і верхньої трикутної матриці U: $A = LU$. У матричному вигляді:

$$\begin{pmatrix} a_{11}\ a_{12}\ a_{13}\dots a_{1n} \\ a_{21}\ a_{22}\ a_{23}\dots a_{2n} \\ a_{31}\ a_{32}\ a_{33}\dots a_{3n} \\ \dots\dots\dots\dots\dots \\ a_{n1}\ a_{n2}\ a_{n3}\dots a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0\dots 0 \\ l_{21} & 1 & 0\dots 0 \\ l_{31} & l_{32} & 1\dots 0 \\ \dots\dots\dots\dots & & \\ l_{n1} & l_{n2} & l_{n3}\ \dots 1 \end{pmatrix} \begin{pmatrix} u_{11}\ u_{12}\ u_{13}\cdots u_{1n} \\ 0 \quad u_{22}\ u_{23}\cdots u_{2n} \\ 0 \quad 0 \quad u_{23}\cdots u_{2n} \\ \dots\dots\dots\dots\dots\dots \\ 0 \quad 0 \quad 0\cdots u_{nn} \end{pmatrix}$$

Тоді розв‘язання системи лінійних рівнянь можна представити $Ax = LUx = b$. Розв‘язок такої системи можна розбити на два кроки:

1. $Ly = b$ – застосовуючи прямий хід;
2. $Ux = y$ – застосовуючи зворотний хід.

Для розкладення матриці A на нижню та верхню трикутні матриці застосовують метод виключення Гаусса.

Приклад.

$$A = \begin{pmatrix} 4 & 3 & -1 \\ -2 & -4 & 5 \\ 1 & 2 & 6 \end{pmatrix} \quad b = \begin{pmatrix} 8 \\ 7 \\ 22 \end{pmatrix}$$

Розкладення матриці:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 3 & -1 \\ -2 & -4 & 5 \\ 1 & 2 & 6 \end{pmatrix}$$ Домножимо послідовно перший рядок на 0,5 і 0,25 та віднімемо з 2-го та 3-го рядка, отримаємо:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -0,5 & 1 & 0 \\ 0,25 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 3 & -1 \\ 0 & -2,5 & 4,5 \\ 0 & 1,25 & 6,25 \end{pmatrix}$$

Тепер за провідний рядок візьмемо другий рядок матриці. Домножимо його на -0,5 і віднімемо від 3-го рядка, отримаємо:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -0,5 & 1 & 0 \\ 0,25 & -0,5 & 1 \end{pmatrix} \begin{pmatrix} 4 & 3 & -1 \\ 0 & -2,5 & 4,5 \\ 0 & 0 & 8,5 \end{pmatrix}$$

Розв‘яжемо систему

$$Ly = b:\ L = \begin{pmatrix} 1 & 0 & 0 \\ -0{,}5 & 1 & 0 \\ 0{,}25 & -0{,}5 & 1 \end{pmatrix},\ b = \begin{pmatrix} 8 \\ 7 \\ 22 \end{pmatrix},\ y = \begin{pmatrix} 7 \\ 19 \\ 34 \end{pmatrix}.$$

Розв‘яжемо систему

$$Ux = y:\ U = \begin{pmatrix} 4 & 3 & -1 \\ 0 & -2{,}5 & 4{,}5 \\ 0 & 0 & 8{,}5 \end{pmatrix},\ y = \begin{pmatrix} 8 \\ 11 \\ 25.5 \end{pmatrix},\ x = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

Метод LU-розкладу дуже ефективний, коли необхідно розв‘язати декілька систем лінійних рівнянь, які відрізняються лише правими частинами.

Суттєвим недоліком методів, які включають перетворення матриці на діагональні, є можливість появи нулів на головній діагоналі, але якщо матриця невироджена, цього можна уникнути шляхом перестановки рядків та стовпчиків, що матрично можна записати $PA = LU$, де P – матриця перестановок.

Приклад.

$$A = \begin{pmatrix} 1 & 2 & 6 \\ 4 & 8 & -1 \\ -2 & 3 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

Помножимо матриці L і U, отримаємо:

1-й стовпчик	Перетворення
$1 = u_{11}$	$2 = 1u_{12}$
$4 = l_{21}u_{11}$	$8 = l_{21}u_{12} = 4 \cdot 2 + u_{22} \Rightarrow u_{22} = 0$
$-2 = l_{31}u_{11} = l_{31}$	$3 = l_{31}u_{12} + l_{32}u_{22} = -2 \cdot 2 + l_{32}0 = -4 \Rightarrow$ *протиріччя*

Отже, матрицю A неможливо розкласти, але якщо її домножити на матрицю перестановок P, то розклад стане можливим:

$$PA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 6 \\ 4 & 8 & -1 \\ -2 & 3 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 6 \\ -4 & 3 & 5 \\ 4 & 8 & -1 \end{pmatrix}$$

Методи розв‘язання систем лінійних рівнянь для матриць спеціального вигляду

Метод Холецького

Для симетричних матриць застосовують спрощену модифікацію методу – метод Холецького. Якщо матриця A симетрична, то її LU-розклад буде мати вигляд $A = LL^T$, де L^T – транспонована матриця нижньої трикутної матриці L. Для обчислення елементів матриці L застосовуються наступні формули:

$$l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{j-1} l_{ik}^2}, \quad j+1 \le i \le k \text{ і } l_{ij} = (a_{ii} - \sum_{k=1}^{j-1} l_{ik} l_{jk}) / l_{ii}, \quad i = 1,2,...,n$$

Метод прогонки

Для стрічкових матриць, які виникають при розв‘язанні крайової задачі (наприклад, розрахунок жорсткості будівельних конструкцій), розроблений спеціальний метод – прогонка. Нехай задана 3-діагональна матриця

$$M = \begin{pmatrix} 1 & x_1 & 0 & \cdots & 0 & 0 \\ a_1 & -c_1 & b_1 & \cdots & 0 & 0 \\ 0 & a_2 & -b_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -x_n & 1 \end{pmatrix}$$

$$A_j x_{j-i} - C_j x_j + B_j x_{j+1} = F_j, \quad j = 1,2,...,n \quad (1)$$

$$x_0 = \alpha_0 x_1 + \beta_0, \; x_N = \alpha_N x_{N-1} + \beta_N \quad (2)$$

$$|C_j| \ge |A_j|, \; |B_j| \ge |A_j| > 0, \; |\alpha_0| < 1, \; |\alpha_N| \le 1. \quad (3)$$

Рівняння (1)-(2) називаються різностною крайовою задачею другого порядку або трьохточковою різностною схемою. Умова (3) гарантує існування єдиного кореня різностної крайової задачі.

Підставимо $x_0 = \alpha_0 x_1 + \beta_0$ в рівняння (1): $A_1(\alpha_0 x_1 + \beta_0) - C_1 x_1 + B_1 x_2 = F_1$ і знайдемо x_1: $x_1 = \alpha_1 x_2 + \beta_1$, де $\alpha_1 = B_1/(C_1 - A_1\alpha_0)$ і $\beta_1 = (A_1\beta_0 - F_1)/(C_1 - A_1\alpha_0)$. Підставимо x_1 в

наступне рівняння і відшукаємо x_2 і так далі, продовживши ці дії, отримаємо рекурентні формули:

$$x_k = \alpha_k x_{k+1} + \beta_k,\ \alpha_k = B_k/(C_k - A_k\alpha_{k-1}),\ \beta_k = (A_k\beta_{k-1} - F_k)/(C_k - A_k\alpha_{k-1})$$

$$x_N = \alpha_N(\alpha_{N-1}x_N + \beta_{N-1}) + \beta N \ \Rightarrow\ x_N = \frac{\beta_N - \alpha_N\beta_{N-1}}{1-\alpha_N\alpha_{N-1}}.$$

Як і метод Гаусса, метод прогонки складається з двох частин алгоритму:

1. Прямого ходу, який полягає у обчисленні всіх коефіцієнтів α_k, β_k.
2. Зворотного ходу, на якому знаходяться всі невідомі x_i за рекурентною формулою: $x_i = \alpha_i x_{i+1} + \beta_i$.

В методі прогонки для знаходження коренів необхідно виконати лише

$N = 6(N-1) + 5 + 2N = 8(N+1) - 9$ операцій (з формули видно, що кількість операцій відносно розмірності задачі зростає лише лінійно).

Лекція 15. Наближені методи розв‘язання лінійних систем

Метод простих ітерацій (Якобі)

Нехай задана система лінійних рівнянь $Ax = b$ і $a_{ii} \neq 0$. Для застосування методу простих ітерацій представимо цю систему у вигляді:

$$x = Cx + d,\ \text{де}\ d = (d_i) = b_i/a_{ii},\ \text{а}\ c_{ij} = \begin{cases} -a_{ij}/a_{ii},\ i \neq j \\ 0, \quad i = j \end{cases}. \quad (1)$$

Вхідними даними є вектор початкових наближень $\bar{x}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}\right)$, на основі якого будується ітераційна послідовність векторів наближень за формулою:

$$x^{(k)} = Cx^{(k-1)} + d,\ k = 1,2,...\ \text{або}\ x_i^{(k+1)} = x_i^{(k)} - \frac{1}{a_{ii}}\left(\sum_{j=1}^{n} a_{ij}x_j^{(k)} - b_i\right) \quad (2)$$

і точність ε, з якою відшукуються корені (процес пошуку продовжується, поки $|x_{k+1}x_k| < \varepsilon$).

Приклад. Нехай задана система рівнянь. ЇЇ розв‘язок $(2;4;3)$:

-3x	+y	+5z	=15		x=	(-15+y+5z)/3		x_k=	$(-15+y_{k-1}+5z_{k-1})/3$	2
4x	-8y	+z	=-21	⇒	y=	(21+4x+z)/8	⇒	y_k=	$(21+4x_{k-1}+z_{k-1})/8$	4
4x	-y	+z	=7		z=	7-4x+y		z_k=	$7-4x_{k-1}+y_{k-1}$	3

Задамо початкове наближення $(x_0, y_0, z_0) = (1; 2; 2)$. Якщо виконати ітерації, то стане очевидно, що метод розбігається.

k	x_1	x_2	x_3
0	1	2	2
1	-1,5	3,375	5
2	6,6875	2,5	16,
3	34,6875	8,0115625	37,5

Теорема. Якщо $\|C\| < 1$, то система рівнянь $x = Cx + d$ має єдиний розв‘язок $\bar{x}^*$ і ітерації, задані формулою (2) (послідовність наближень), збігаються до розв‘язку зі швидкістю геометричної прогресії.

Умові теореми задовольняють матриці з явно домінуючою головною діагоналлю, і для таких систем метод є досить ефективним. Так, в нашому прикладі $\|C\|_1 = 17$ або $\|C\|_2 = \sqrt{129}$ – обидві норми значно більші за одиницю і не задовольняють умові теореми.

Якщо матриця не відповідає вимозі теореми, то елементарними перетвореннями її можна привести до матриці з домінуючою діагоналлю. Перестановкою рядків (1-го і 3-го) матрицю нашого прикладу можна переписати наступним чином:

4x	-y	+z	=7		x=	(7+y-z)/4		x_k=	$(7+y_{k-1}-z_{k-1})/4$	2
4x	-8y	+z	=-21	⇒	y=	(21+4x+z)/8	⇒	y_k=	$(21+4x_{k-1}+z_{k-1})/8$	4
-2x	+y	+5z	=15		z=	(15+2x-y)/5		z_k=	$(15+2x_{k-1}-y_{k-1})/5$	3

Отримаємо матрицю з домінуючою головною діагоналлю. Візьмемо ті ж самі початкові наближення $(x_0, y_0, z_0) = (1; 2; 2)$ і отримаємо збіжну послідовність ітерацій:

k	x_1	x_2	x_3
0	1	2	2
1	1,75	3,375	3,0
2	1,84375	3,875	3,025
3	1,9625	3,925	2,965

На похибку простих ітерацій можна дати наступну оцінку:

$\left\|x^* - x^{(k-1)}\right\| \le \left\|x^{(k)} - x^{(k-1)}\right\| + \|C\| \cdot \left\|x^* - x^{(k-1)}\right\| \le \frac{1}{1-\|C\|}\left\|x^{(k)} - x^{(k-1)}\right\|$ і

$\left\|x^* - x^{(k)}\right\| \le \|C\| \cdot \left\|x^* - x^{(k-1)}\right\|$, остаточно отримаємо

$\left\|x^* - x^{(k)}\right\| \le \frac{\|C\|}{1-\|C\|}\left\|x^{(k)} - x^{(k-1)}\right\|$.

Метод Зейделя

Збіжність послідовності ітерацій можна прискорити, якщо на кожному кроці ітерацій у рекурентну формулу визначення $x_i^{(k)}$ підставляти вже знайдені значення $x_j^{(k)}$, $j < i$. Така підстановка називається ітераціями Зейделя. Ітерації Зейделя можна записати наступною рекурентною формулою:

$$x_i^{(k)} = \sum_{j=1}^{i-1} c_{ij} x_j^{(k)} + \sum_{j=i+1}^{n} c_{ij} x_j^{(k-1)} + d\,.$$

Умови збіжності методу простих ітерацій і методу Зейделя не співпадають, але перетинаються.

Теорема. Для існування єдиного розв‘язку системи (1) і збіжності методу Зейделя достатньо, щоб виконувалась хоча б одна з умов:

1) $\sum_{j\neq i}|c_{ij}|<|c_{ii}|, \quad i=\overline{1,n}$ (матриця строго діагонально домінуюча);
2) матриця C є симетричною додатньо визначеною (всі її власні числа додатні).

Приклад. Візьмемо матрицю з прикладу простих ітерацій:

4x	-y	+z	=7		x=	(7+y-z)/4		x_k=	$(7+y_{k-1}-z_{k-1})/4$	2
4x	-8y	+z	=-21	$\Rightarrow$	y=	(21+4x+z)/8	$\Rightarrow$	y_k=	$(21+4x_k+z_{k-1})/8$	4
-2x	+y	+5z	=15		z=	(15+2x-y)/5		z_k=	$(15+2x_k-y_k)/5$	3

Скориставшись ітераціями Зейделя, можна побачити, що збіжність буде більш швидкою:

k	x_1	x_2	x_3
0	1	2	2
1	1,75	3,75	2,95
2	1,95	3,968	2,98625
3	1,995	3,996	2,99901

Нормування матриці

Нехай задана система лінійних рівнянь:

$$\begin{cases} 7,6x_1+0,5x_2+2,4x_3=1,9 \\ 2,2x_1+9,1x_2+4,4x_3=9,7 \\ -1,3x_1+0,2x_2+5,8x_3=-1,4 \end{cases}$$

Необхідно знайти розв‘язок цієї системи методом простих ітерацій з точністю $\varepsilon=10^{-3}$.

Матриця цієї системи є діагонально домінуючою, отже, вона задовольняє умові збіжності ітерацій. Якщо почати виконувати ітерації, то буде очевидним, що вони збігаються дуже повільно. Прискорити їх збіжність можна, якщо коефіцієнти матриці будуть

задовольняти умові $\sum_{j=1}^{n}\left|a_{ij}\right|<1,\ i=\overline{1,n}$. Для досягнення виконання цієї умови можна зробити наступні перетворення:

$$\begin{cases} 10x_1 - 2{,}4x_1 + 0{,}5x_2 + 2{,}4x_3 = 1{,}9 \\ 2{,}2x_1 + 10x_2 - 0.99x_2 + 4{,}4x_3 = 9{,}7 \\ -1{,}3x_1 + 0{,}2x_2 + 10x_3 - 4{,}2x_3 = -1{,}4 \end{cases}$$

Коефіцієнти на головній діагоналі ми представляємо у вигляді $a_{ii}x_i = mx_i - lx_i$, підбираючи значення числа m так, щоб сума діагональних елементів була меншою за одиницю. Тоді ітерації можна записати наступним чином:

$$\begin{array}{ll} 10x_1 = 1{,}9 + 2{,}4x_1 - 0{,}5x_2 - 2{,}4x_3 & x_1 = 0{,}19 + 0{,}24x_1 - 0{,}05x_2 - 0{,}24x_3 \\ 10x_2 = 9{,}7 - 2{,}2x_1 + 0{,}9x_2 - 4{,}4x_3 \Rightarrow & x_2 = 0{,}97 - 0{,}22x_1 + 0{,}09x_2 - 0{,}44x_3 \\ 10x_3 = -1.4 + 1{,}3x_1 - 0{,}2x_2 + 4{,}2x_3 & x_3 = -0{,}14 + 0{,}13x_1 - 0{,}02x_2 + 0{,}42x_3 \end{array}$$

Тоді будемо мати:

$$\sum_{j=1}^{3}\left|a_{1j}^{*}\right| = 0{,}53 < 1, \quad \sum_{j=1}^{3}\left|a_{2j}^{*}\right| = 0{,}75 < 1, \quad \sum_{j=1}^{3}\left|a_{3j}^{*}\right| = 0{,}57 < 1 \ \sum_{i=1}^{3}\left|b_i^{*}\right| 1{,}3.$$

Можна дати оцінку на кількість кроків для отримання розв‘язку:

$$\frac{q^{i+1}}{1-q}\sum_{i=1}^{n}\left|b_i^{*}\right| < \varepsilon, \quad \text{де} \quad q = \max_i \sum_j \left|a_{ij}^{*}\right| \quad \Rightarrow \quad \frac{0{,}75^{i+1}}{1-0{,}75}1{,}3 < 10^{-3} \Rightarrow$$

$$\Rightarrow 0{,}75^{i+1} \le 0{,}19 \cdot 10^{-3} \Rightarrow i \le 29.$$

Як видно, після виконаного нормування матриці ітерації будуть збігатися значно швидше.

Власні числа

Власними числами $\lambda_i, i=\overline{1,n}$ квадратної матриці A n-го порядку називаються дійсні або комплексні числа, що задовольняють умові $Ax = \lambda_i x$. Власні числа застосовуються для оцінки існування розв‘язку системи алгебраїчних рівнянь і рівнянь, які описуються поліномами. Їх застосування має також важливий практичний зміст, наприклад, в механіці вони характеризують внутрішню напругу об‘єкта, що виникає під впливом зовнішніх сил, в радіоелектроніці

власні числа визначають часові характеристики та режими роботи пристроїв.

Власні числа знаходяться з характеристичного рівняння:

$$\Delta A = \begin{pmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{pmatrix} = 0.$$

Результатом обчислення визначника буде поліном, корені якого і будуть визначати власні числа. Для відшукання коренів характеристичного рівняння застосовують метод Данилевського. Для цього будують послідовність матриць $A^{(1)}, A^{(2)},..., A^{(n)}$ за формулами:

$$a_{ij}^{(k+1)} = b_{ij}^{(k)},\ j \neq k+1,\ a_{ik+1}^{(k+1)} = \sum_{j=1}^{n} b_{ij}^{(k)} a_{jk}^{(k)},$$

де $b_{k+1j}^{(k)} = a_{k+1j}^{(k)} / a_{k+1k}^{(k)}$ і $b_{ij}^{(k)} = a_{ij}^{(k)} - a_{ik}^{(k)} b_{k+1j}^{(k)}, (i \neq k+1), \quad i, j = \overline{1,n}$.

Причому перетворення побудовані таким чином, що у отриманій матриці $A^{(n)}$ значення останнього стовпчика і є власними числами.

Максимальне власне число матриці і відповідний йому власний вектор можна знайти за допомогою степеневого методу, алгоритм якого складається з наступних кроків:

1. За формулою $\bar{y}^{(c)} = A\bar{y}^{(c-1)}$ обчислюється вектор $\bar{y}^{(c)}$, де вхідними даними є початкове довільне значення цього вектора, A – задана матриця.
2. Знаходяться наближення до максимального власного значення: $\lambda_{mqx} = \frac{1}{n}\sum_{i=1}^{n} y_i^{(c)} / y_i^{(c-1)}$.
3. Виконується нормування вектора $\bar{y}^{(c)}$: $\bar{y}^{(c)} = \frac{1}{\max y_i^{(c)}} \bar{y}^{(c)}$.
4. Перевіряється виконання умови $\left|\lambda_{\max}^{(c)} - \lambda_{\max}^{(c-1)}\right| < \varepsilon$, де ε – задана точність обчислення. Якщо умова виконується, то вважається, що максимальне власне число і відповідний йому власний вектор знайдено.

Якщо всі власні числа заданої матриці є дійсними і максимальне власне число не є кратним, то їх можна знайти за допомогою методу скалярних добутків, який складається з наступних кроків:

1. Задається початкове наближення вектора $\bar{y}^{(0)}$ і обчислюється $\bar{y}^{(k)} = A^{(i)}\bar{y}^{(k-1)}$.
2. Обчислюється наближення до максимального власного значення: $\lambda_{\max}^{(k)} = \left(\bar{y}^{(k)}, \bar{y}^{(k)}\right) / \left(\bar{y}^{(k-1)}, \bar{y}^{(k)}\right)$.
3. Нормуємо вектор $\bar{y}^{(k)}$: $\bar{y}^{(k)} = \bar{y}^{(k)} / \sqrt{\left(y^{(k)}, y^{(k)}\right)}$.
4. Перевіряємо виконання умови $\left|\lambda_{\max}^{(k)} - \lambda_{\max}^{(k-1)}\right| < \varepsilon$. Якщо вона виконується, то власне число знайдено, якщо ні – то повторюємо виконання алгоритму, починаючи з пункту 1.

Для знаходження чергового власного числа повторюємо виконання пунктів алгоритму 1-4, перетворивши початкову матрицю за формулою: $A^{(i)} = A^{(i-1)} - \lambda_i (U_i, U_i')$, де (U_i, U_i') – добуток вектора-стовпчика на вектор-рядок

Розв‘язок систем нелінійних рівнянь

Загальні поняття

Розглянемо систему нелінійний рівнянь, яку у загальному вигляді можна записати:

$$\begin{aligned} &f_1(x_1, x_2, ..., x_n) = 0 \\ &f_2(x_1, x_2, ..., x_n) = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &f_n(x_1, x_2, ..., x_n) = 0 \end{aligned}$$ або у векторній формі $\bar{f}(\bar{x}) = 0$ (1)

Отже, будемо розглядати систему рівнянь у n-вимірному просторі. Введемо для n-вимірного простору поняття відстані $\rho(x, y) = \|x - y\|$ через норму вектора, або згідно до раніше введених понять норми:

$$\rho(x,y)=\begin{cases}\rho_1(x,y)=\max\limits_{1\le i\le n}|x_i-y_i|\\ \rho_2(x,y)=\left(\sum\limits_{i=1}^{n}(x_i-y_i)^2\right)^{1/2}\end{cases}$$

Множина точок x, для яких справджується нерівність $\overline{S}\left(y^0,r\right)=\left\{x:\rho\left(x,y^0\right)\le r\right\}$, називається замкненою кулею радіуса r з центром в точці y^0.

Матриця вигляду:

$$F(x)=\begin{pmatrix}\frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \dots & \frac{\partial f_1(x)}{\partial x_n}\\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \dots & \frac{\partial f_2(x)}{\partial x_n}\\ \dots & \dots & \dots & \dots\\ \frac{\partial f_n(x)}{\partial x_1} & \frac{\partial f_n(x)}{\partial x_2} & \dots & \frac{\partial f_n(x)}{\partial x_n}\end{pmatrix}$$

називається якобіаном або матрицею Якобі для системи функцій $f_1(x),\ f_2(x),...,\ f_n(x)$.

Узагальнений метод простих ітерацій і Зейделя

Представимо систему рівнянь (1) у формі:

$$x=\varphi(x), \tag{5}$$

де $\varphi(x)=(\varphi_1(x),\varphi_2(x),...,\varphi_{n1}(x))$ – задана вектор-функція від змінних $x=(x_1,x_2,...,x_n)$ так, що $\varphi_i(x)=\varphi_i(x_1,x_2,...,x_n)$. Задавшись вектором початкових наближень $\bar{x}$ і виконуючи ітерації за формулою (5), можна знайти розв‘язок заданої системи нелінійних рівнянь.

Теорема. Нехай на замкненій кулі $\overline{S}\left(y^0,r\right)=\overline{S}$ задана вектор-функція $\varphi(x)$, причому для будь-яких $x,y\in\overline{S}$ виконується нерівність $\rho(\varphi(x),\varphi(y))\le\alpha\rho(x,y)$ і $\rho\left(\varphi\left(y^0\right),y^0\right)\le(1-\alpha)r$, де $\alpha\,(0\le\alpha\le1)$ – деяка числова константа. Тоді на $\overline{S}$ існує єдиний розв‘язок системи рівнянь, причому $x^*=\lim\limits_{k\to\infty}x^{(k)}$, де при довільному $x^{(0)}\in\overline{S}$ $x^{(k)}=\varphi\left(x^{(k-1)}\right),\quad k=1,2,...$ і при цьому виконуються нерівності:

$$\rho\left(x^*, x^{(k)}\right) \le \alpha\rho\left(x^*, x^{(0)}\right) \le 2\alpha^k r,\ \rho\left(x^*, x^{(k)}\right) \le \frac{\alpha}{1-\alpha}\rho\left(x^{(k)}, x^{(k-1)}\right).$$

Також збіжність простих ітерацій до нерухомої точки можна визначити, аналогічно як і для нелінійного рівняння, за похідною.

$\sum_{j=1}^{n} \partial\varphi_i(x)/\partial x_j < 1, i = \overline{1,n}$ (сума часткових похідних для кожного рядка системи повинна бути строго менше 1).

Якщо на кожному k-му кроці ітерацій за формулою (2) у наступне рівняння $x_j^{(k)} = \varphi_j(x)$ підставляти значення $x_i^{(k)}$, де $j > i$, то отримаємо метод Зейделя для нелінійних систем.

Метод Ньютона

Якщо матриця $F(x)$ – якобіан заданої системи нелінійних рівнянь, то метод Ньютона для систем нелінійних рівнянь можна представити ітераційною формулою:

$$x^{(k)} = x^{(k-1)} - F^{-1}\left(x^{(k-1)}\right) f\left(x^{(k-1)}\right), \quad k = \overline{1,n}.$$

Як видно з вигляду рекурентної формули, метод Ньютона для систем пов‘язаний з великою кількістю обчислень (на кожному кроці необхідно обчислювати $F^{-1}(x)$), тому на практиці часто користуються спрощеним рекурентним співвідношенням: $x^{(k)} = x^{(k-1)} - F^{-1}\left(x^{(0)}\right) f\left(x^{(k-1)}\right), \quad k = \overline{1,n}$, де матриця є оберненою до якобіана і відшукується лише один раз для початкового наближення вектора $x^{(0)}$.

При добре вибраному початковому наближенні метод Ньютона дає квадратичну збіжність, тоді як метод простих ітерацій – лише лінійну.

Контрольні запитання

1. Як визначається норма матриці?
2. Яка матриця є погано обумовленою?
3. Які недоліки має метод Гаусса для розв’язання практичних задач?

4. Які точні методи використовують на практиці для розв’язання систем лінійних рівнянь?
5. Які наближені методи використовують на практиці для розв’язання систем лінійних рівнянь?
6. Який метод є оптимальним для систем із симетричною матрицею?
7. Який метод застосовують для систем із стрічковою матрицею?
8. Які методи використовують для розв’язання систем нелінійних рівнянь?
9. Яке спрощення можна прийняти для зменшення кількості операцій в методі Ньютона для нелінійних систем?
10. Яка матриця називається якобіаном?

Список літератури

1. *Бахвалов Н.С.* Численные методы / Н. С. Бахвалов, Н. П. Жидков, Г. Н. Кобельков. – М.: БИНОМ. Лаб. знаний, 2003. – 632 с.
2. *Бахвалов Н.С.* Численные методы в задачах и упражнениях / Н. С. Бахвалов А. В. Лапин, Е. В. Чижонков. – М.: Высш. шк., 2000. – 192 с.
3. *Вержбицкий В.М.* Численные методы. Линейная алгебра и нелинейные уравнения / В.М. Вержбицкий. – М.: Высш.шк., 2000. – 268 с.
4. *Вержбицкий В.М.* Численные методы. Математический анализ и обыкновенные дифференциальные уравнения / В.М. Вержбицкий. – М.: Высш. шк., 2001. – 383 с.
5. *Волков Е.А*. Численные методы / Е.А. Волков. – СПб.: Лань, 2004. – 248 с.

Зміст

Навчальне видання

ГОРДА Олена Володимирівна

ЧИСЕЛЬНІ МЕТОДИ

РОЗВ'ЯЗАННЯ НЕЛІНІЙНИХ РІВНЯНЬ ТА СИСТЕМ РІВНЯНЬ

Конспект лекцій

для студентів, які навчаються за напрямом підготовки 6.050101 "Комп'ютерні науки"

Редагування і коректура *Г.Є. Голіциної*
Комп'ютерна верстка *О.В. Кириченка*

Підписано до друку Формат 60x84 1/16.
Ум. друк. арк. 4,18. Обл.-вид. арк. 4,5.
Тираж 60 прим. Вид. № 34/I-09. Зам. №
КНУБА, Повітрофлотський проспект, 31, Київ, Україна, 03680

E-mail: red-isdat@knuba.edu.ua

Віддруковано в редакційно-видавничому відділі
Київського національного університету будівництва і архітектури

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
ДК № 808 від 13.02.2002 р.