Тема 2. Вступ до чисельних методів. Загальні поняття

2.1 Сутність чисельних методів. Загальні поняття

Для розв'язання математичних задач в основному існує три групи методів:

1. **Аналітичні методи**, в яких розв'язок задачі подається у вигляді аналітичних виразів. Їх **перевагами** є: запис розв'язку у загальному вигляді; висока точність і малий об'єм комп'ютерної пам'яті для зберігання розв'язку. Основний **недолік** – неуніверсальність, бо тільки невелика частина математичних задач може бути розв'язана аналітично.

2. **Графічні методи**, в яких розв'язок задачі знаходиться візуально. Їх **перевагою** є наочність. **Недоліками** графічних методів є: велика трудомісткість; низька точність (залежить від точності побудови графіків); неуніверсальність (графіки можна побудувати тільки для невеликої розмірності та ін.).

3. **Чисельні методи**, що дозволяють звести розв'язування задачі до виконання скінченного числа арифметичних і логічних дій з числами. При цьому розв'язок визначається як набір чисел, які надалі можуть бути інтерпретовані різним способом (наприклад, подані у вигляді таблиць, графіків, анімації тощо). Їх **перевагами** є: абсолютна універсальність, бо теоретично можуть бути застосовані для розв'язання будь-яких задач; добре пристосовані для реалізації на комп'ютері. **Недоліком** є велика трудомісткість у ході ручного рахунку, що, зазвичай, не є проблемою, оскільки вони призначені для використання на комп'ютері.

Таким чином, чисельні методи є основним апаратом розв'язання математичних задач, а їх значущість тільки збільшуватиметься у міру вдосконалення комп'ютерної техніки.

Чисельні методи бувають двох типів: **прямі** та **ітераційні**. В прямих методах розв'язок задачі досягається за скінченну кількість **кроків** методу після виконання останнього кроку, в ітераційних методах виконується ряд **ітерацій** методу до отримання наближеного розв'язку із заданою точністю.

Поняття ітераційного методу.

В основному чисельні методи є **ітераційними**. **Ітерація** – це повторення сукупності операцій або процедур для покращення наявного (поточного) наближеного розв'язку задачі. Нехай x^* – розв'язок задачі, тоді ітераційний метод будує так звану **ітераційну послідовність** $\left\{x^{(k)}\right\}$ $\left(k = 0, 1, 2, \ldots\right)$ наближень розв'язку, при цьому $x^{(k)}$ повинно наближатися до x^* зі збільшенням k.

Алгоритм ітераційного методу в найзагальнішому вигляді має таку схему:

1. Задається **початкове наближення** розв'язку $x^{(0)}$ (на основі апріорних знань про задачу).
2. На k-й ітерації методу $\left(k = 0, 1, 2, \ldots\right)$ буде поточне наближення розв'язку $x^{(k)}$. Далі обчислюється наступне наближення $x^{(k+1)} = \Phi(x^{(k)})$, де Φ і є сукупністю операцій або процедур для покращення наближеного розв'язку задачі, яка є суттю конкретного чисельного методу.
3. Перевіряється **критерій останову**, тобто перевіряється: чи є отримане наближення $x^{(k+1)}$ розв'язку x^* достатньо близьким. Якщо цього немає, то відбувається перехід до наступної ітерації, тобто до пункту 2.

Варто зазначити, що вид критерію останову (тобто припинення обчислень за ітераційним методом) залежить від виду розв'язуваної математичної задачі.

2.2. Характеристики чисельних методів

Для оцінки чисельних методів, тобто порівняння між собою методів для розв'язання однієї задачі, вводять такі їх основні характеристики:

трудомісткість;
порядок методу;
збіжність;
швидкість збіжності;
стійкість до погрішностей обчислень;
стійкість до погрішностей у відправних даних.

Під **трудомісткістю** методу розуміють кількість і якість обчислень, необхідних для досягнення достатньо близького наближення розв'язку задачі.

Під **порядком** методу розуміють вимоги до знань про функції, що входять у математичне формулювання задачі (наприклад, використання в методі похідних цих функцій):

- метод **нульового порядку**, якщо він використовує тільки значення цих функцій;
- метод **першого порядку**, якщо він використовує значення функцій і їх перших похідних;
- метод **другого порядку**, якщо він використовує значення і функцій та їх перших і других похідних і т. д.

Чисельний метод називається **таким, що збігається**, якщо наближення x^k прямує до розв'язку x^* зі збільшенням k. Очевидно, що методи, які не збігаються, не цікаві з прикладної точки зору. Тому одним з найважливіших етапів при введенні нового чисельного метода є теоретичне доведення його збіжності, тобто формулювання умов, за яких метод гарантовано збігається.

В основному розрізняють такі **швидкості збіжності** методів.

1. **Лінійна збіжність**. Говорять, що послідовність $\left\{x^{(k)}\right\}(k=0,1,2,\ldots)$ лінійно збігається до розв'язку x^* (або зішвидкістю геометричної прогресії), якщо існують числа $q\in(0,1)$ і $k_0>0$ такі, що

$$\left\|x^{(k+1)}-x^*\right\|\le q\left\|x^{(k)}-x^*\right\| \text{ для всіх } k\ge k_0.$$

Тут норма $\|x-y\|$ означає відстань між x і y.

2. **Надлінійна збіжність**. Говорять, що послідовність $\left\{x^{(k)}\right\}$ $(k=0,1,2,\ldots)$ надлінійно збігається до розв'язку x^*, якщо існує послідовність $\{q_k\}$ $(k=0,1,2,\ldots)$, $q_k\in(0,1)$ для всіх k, така, що

$$\left\|x^{(k+1)}-x^*\right\|\le q_k\left\|x^{(k)}-x^*\right\| \text{ і } q_k\to 0 \text{ при } k\to\infty.$$

3. **Квадратична збіжність**. Говорять, що послідовність $\left\{x^{(k)}\right\}$ $(k=0,1,2,\ldots)$ квадратично збігається до розв'язку x^*, якщо існують числа $C>0$ і $k_0>0$ такі, що

$$\left\|x^{(k+1)}-x^*\right\|\le C\left\|x^{(k)}-x^*\right\|^2 \text{ для всіх } k\ge k_0.$$

Під **стійкістю до погрішностей обчислень** розуміють те, що застосування чисельного методу приводить до розв'язку задачі на комп’ютері, незважаючи на помилки округлень і обчислень. Для цього в чисельних методах, якщо потрібно, передбачаються додаткові операції, що не змінюють суть методу, але забезпечують його стійкість до помилок обчислень.

Під **стійкістю до погрішностей у відправних даних** розуміють те, що при невеликих погрішностях у відправних даних застосування чисельного методу дозволяє отримати наближений розв'язок задачі з не дуже великою погрішністю. Стійкість до погрішностей у відправних даних досягається, як правило, шляхом модифікації чисельного методу, тобто внесенням змін до суті методу.

2.3. Похибка розв'язку

Для оцінювання точності чисельного методу слід ввести поняття абсолютної і відносної погрішності розв'язку.

Якщо x^* – точний розв'язок задачі, а $\tilde{x}$ – наближений розв'язок, то **абсолютною погрішністю** наближення $\tilde{x}$ називається деяка величина $\Delta(\tilde{x})$, про яку відомо, що

$$\left\|\tilde{x} - x^*\right\| \le \Delta(\tilde{x}).$$

Тут норма $\left\|\tilde{x} - x^*\right\|$ означає відстань між $\tilde{x}$ і x^*. Таким чином, якщо відомі наближений розв'язок задачі $\tilde{x}$ та абсолютна погрішність $\Delta(\tilde{x})$, то невідомий точний розв'язок задачі x^* знаходиться від $\tilde{x}$ на відстані не більше, ніж $\Delta(\tilde{x})$.

Також варто зазначити, що абсолютна погрішность розв'язку – це не різниця між точним і наближеним розв'язком задачі, а оцінка цієї різниці.

Відносною погрішністю наближення $\tilde{x}$ називають деяку величину $\delta(\tilde{x})$, про яку відомо, що

$$\frac{\left\|\tilde{x} - x^*\right\|}{\left\|\tilde{x}\right\|} \le \delta(\tilde{x}).$$

Варто звернути увагу на те, що наближений розв'язок задачі $\tilde{x}$, отриманий за допомогою чисельного методу, зазвичай залежить від вибраних параметрів цього методу. Без обмеження загальності, можна вважати, що параметр у методу один (умовно можна позначити його через h) і при цьому він задовольняє умову $0 < h < 1$. Тоді абсолютна погрішність $\Delta(\tilde{x})$ також залежить від h, тобто $\Delta(\tilde{x}) = \Delta(h)$.

Якщо існують деякі числа $M > 0$ і $k > 0$ (не обов'язково цілі) такі, що виконується нерівність $\Delta(h) \le M\,h^k$, то говорять, що абсолютна погрішність $\Delta(h)$ має порядок $O(h^k)$. Цим хочуть підкреслити **якісну** (на відміну від кількісної) залежність погрішності наближеного розв'язку задачі $\tilde{x}$ від параметра чисельного методу h. Наприклад, це показує, що зменшення значення параметра h на один порядок призводить до зменшення величини абсолютної погрішності $\Delta(\tilde{x})$ на k порядків.

2.4. Похибка округлення у ході розрахунків на комп'ютері з плаваючою крапкою

Одним із основних джерел обчислювальних погрішностей є наближене подання дійсних чисел у комп'ютері, обумовлене скінченністю розрядної сітки. При розв'язанні обчислювальних задач зазвичай використовують подання чисел у **формі з плаваючою крапкою** *(комою)*. Число a у формі з плаваючою крапкою подається у вигляді

$$a = sign(a)\,M\,r^p,$$

де r – основа системи числення;

p – порядок числа a;

M – мантиса числа a;

$sign(a)$ – знак числа a.

При цьому повинна виконуватися умова нормування $r^{-1} \le M \le 1$, тобто перша цифра в мантисі повинна бути відмінною від нуля.

Усі персональні комп'ютери і більшість робочих станцій підтримують **IEEE-стандарт двійкової арифметики** (тобто $r = 2$). Стандарт передбачає два типи чисел із плаваючою крапкою: *числа звичайної точності* (подаються 4 байтами – 32 бітами)

1	8	23
Знак	Порядок	Мантиса

і *числа подвійної точності* (подаються 8 байтами – 64 бітами)

1	11	52
Знак	Порядок	Мантиса

Через скінченність розрядної сітки (кількість знаків у мантисі) в комп'ютері можна подати не всі числа. Число a, яке не можна подати в комп'ютері, піддається округленню, тобто замінюється близьким числом $\tilde{a}$, яке подається в комп'ютері точно.

Слід знайти оцінку *відносної погрішності подання числа з плаваючою крапкою* [**Ошибка! Источник ссылки не найден.**]. Припускається, що застосовується просте округлення – відкидання всіх розрядів числа, що виходять за межі розрядної сітки. Нехай потрібно записати число, що є нескінченним двійковим дробом (система числення – двійкова)

$$a = \pm \underbrace{2^p}_{\text{порядок}} \underbrace{\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_t}{2^t} + \frac{a_{t+1}}{2^{t+1}} + \cdots \right)}_{\text{мантиса}}, \qquad (0.1)$$

де $a_j = \begin{cases} 0 \\ 1 \end{cases}, \ (j = 1, 2, \ldots)$ – цифри мантиси;

t – розрядність мантиси (23 або 52 залежно від типу числа).

Тоді, після відкидання зайвих розрядів, буде отримане округлене число

$$\tilde{a} = \pm 2^p \left(\frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_t}{2^t} \right).$$

Погрішність округлення в цьому випадку дорівнює

$$\tilde{a} - a = \pm 2^p \left(\frac{a_{t+1}}{2^{t+1}} + \frac{a_{t+2}}{2^{t+2}} + \cdots \right).$$

Найбільша погрішність буде у випадку, коли $a_{t+1} = 1$, $a_{t+2} = 1$, ... Тоді **абсолютна погрішність округлення**

$$|\tilde{a} - a| \le 2^p \frac{1}{2^{t+1}} \underbrace{\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \right)}_{=2} = 2^{p-t}.$$

З умови нормування виходить, що $M \geq \frac{1}{2}$, а значить завжди $a_1 = 1$. Тоді з (0.1) отримуємо, що $|a| \geq 2^p \times 2^{-1} = 2^{p-1}$ і тому $\frac{|\widetilde{a} - a|}{|a|} \leq 2^{-t+1}$, тобто відносна погрішність дорівнює 2^{-t+1}.

У реальності застосовують точніші методи округлення (ніж просте відкидання розрядів), тому **відносна погрішність** подання чисел в пам'яті комп'ютера дорівнює

$$\frac{|\widetilde{a} - a|}{|a|} \leq 2^{-t}, \tag{0.2}$$

тобто точність подання чисел визначається розрядністю мантиси t.

Слід зазначити, що з (0.2) витікає, що наближено подане в комп'ютері число $\widetilde{a}$ можна записати в вигляді $\widetilde{a} = a(1 \pm \varepsilon)$, де $|\varepsilon| \leq 2^{-t}$. Тому відносну погрішність подання чисел 2^{-t} ще називають "машинним епсілоном".

Більш детально похибки у ході розрахунків на комп'ютері з плаваючою крапкою розглянуті в роботах [**Ошибка! Источник ссылки не найден.**; **Ошибка! Источник ссылки не найден.**; **Ошибка! Источник ссылки не найден.**].

2.5. Математичні пакети

За останні 50 – 60 років, пов’язаних із використанням комп’ютерної техніки, в усьому світі накопичено велику кількість програмного забезпечення у вигляді бібліотек комп’ютерних підпрограм (написаних, у першу чергу, мовами Fortran та C), призначених для розв'язання типових математичних задач. Крім того, розроблено ряд універсальних математичних пакетів, за допомогою яких можна достатньо швидко як розв'язати багато відомих математичних задач, так і провести попередні розрахунки з подальшою реалізацією вже в проектованій комп’ютерній системі.

Найбільш відомими є такі математичні пакети: R, Mathematica, Maple, MatLab, MathCad [**Ошибка! Источник ссылки не найден.**; **Ошибка! Источник ссылки не найден.**].

Пакет **Mathematica** найпопулярніший у наукових колах, особливо у теоретиків. Він надає широкі можливості під час проведення символічних (аналітичних) перетворень, дозволяє швидко й ефективно розв’язувати багато задач обчислювальної математики. В пакеті Mathematica більшість задач розв’язується в діалоговому режимі, без традиційного програмування з використовуванням стандартних операторів. Крім того, в цьому пакеті реалізована мова програмування високого рівня, яка дозволяє створювати достатньо складні програми. Також цей пакет дозволяє конвертувати документи у формат LaTex, який є стандартним форматом переважної більшості наукових видавництв світу.

Пакет **Maple** – найдавніший серед пакетів символьної математики – також дуже популярний у наукових колах. Окрім здійснення аналітичних перетворень, з його допомогою можна розв’язувати різноманітні математичні задачі чисельно. Графічний редактор пакета дозволяє одержувати зображення тривимірних фігур із перерізами та конвертувати документи у формат LaTex.

Пакет **MatLab**, як і згадані пакети, є своєрідною мовою програмування високого рівня, що орієнтована переважно на інженерні розрахунки теорії управління, електро- і радіотехніки, а також моделювання технічних систем. Пакет MatLab став фактично міжнародним стандартом сучасного навчального програмного забезпечення і використовується в багатьох провідних університетах світу [**Ошибка! Источник ссылки не найден.**].

Пакет **MathCad** відрізняється від інших такою особливістю, як використання звичайних математичних позначень. Документ пакета MathCad на екрані монітору виглядає як звичайний математичний розрахунок. Крім того, MathCad є середовищем візуального програмування, тобто не вимагає знання спеціального набору команд. Хоча пакет орієнтований, у першу чергу, на проведення чисельних розрахунків, він має вбудований символьний процесор Maple, що дозволяє виконувати аналітичні перетворювання. Також у пакеті MathCad передбачена можливість створювати зв’язки його документів з документами пакета MatLab.

Система **R** відрізняється від наведених пакетів в основному тим, що є вільним програмним продуктом, який розповсюджується на умові ліцензії GNU. Вона включає в себе об’єктно-орієнтовану мову програмування R і середовище в сукупності з великим набором бібліотек (більш ніж 1600), які доступні для усіх основних платформ – Linux, Windows і MacOS. Постійно виходять нові версії та розширення, які доступні на офіційному сайті http://cran.r-project.org [**Ошибка! Источник ссылки не найден.**], розробники пишуть для системи R пакети, адаптовані до конкретних галузей. Важливим показником популярності цього програмного продукту є і те, що постійно публікуються нові книги з описом застосування R для різних видів аналізу даних [**Ошибка! Источник ссылки не найден.**; **Ошибка! Источник ссылки не найден.**]. Мова R вбудовується в усілякі системи. Наприклад, корпорація Oracle інтегрувала мову програмування R у СУБД Oracle Database 11g.

З цих причин саме пакет R використовується у цьому навчальному посібнику для програмування, виконання розрахунків та наочного подання результатів розв'язання. Математичний пакет R все частіше застосовується у ході проведення лабораторних занять із різних навчальних дисциплін, тому тим, хто бажає більш досконально опанувати цей пакет, пропонується звернутися до додаткових джерел [**Ошибка! Источник ссылки не найден.** – **Ошибка! Источник ссылки не найден.**].

Треба зазначити, що останнім часом позначилася тенденція до зближення та інтеграції різних математичних пакетів. Так, останні версії пакетів Mathematica і Maple мають потужні засоби для візуального програмування; в пакети MatLab та MathCad включена бібліотека аналітичних перетворень Maple; MathCad дозволяє працювати разом із MatLab і т. д. Тому, за великим рахунком, не має значення, який математичний пакет застосовується у навчальному процесі та засвоюється студентами. Головне – принципово навчитися використовувати будь-який математичний пакет для розв'язання практичних задач.

2.6. Висновки

1. Чисельні методи є одним із основних апаратів розв'язання математичних задач.
2. Універсальних чисельних методів розв'язання математичної задачі, як правило, не існує. Тому треба розглядати кілька методів.
3. Вибір чисельного методу для розв'язання математичної задачі треба здійснювати з урахуванням його характеристик.
4. Є ряд математичних пакетів, за допомогою яких можна достатньо швидко розв'язати багато відомих математичних задач.

2.7. Контрольні запитання та завдання

1. Назвіть основні групи методів розв'язання математичних задач та їх характеристики.
2. Назвіть основні характеристики чисельних методів. Розкрийте їх суть.

3. Які є основні швидкості збіжності ітераційних методів?
4. Назвіть та охарактеризуйте основні етапи розв'язання практичних задач на комп'ютері.
5. Що називають "ітерацією"? Наведіть загальну схему ітераційного методу.
6. Які методи розв'язання математичних задач називають ітераційними?
7. Які методи розв'язання математичних задач називають чисельними?
8. Дайте визначення поняттям абсолютної та відносної погрішностей.
9. Дайте визначення лінійної і квадратичної швидкості збіжності ітераційного методу.
10. Що є причинами появи погрішності при арифметичних обчисленнях на комп'ютері?
11. Як подані числа в комп'ютері? Вкажіть розмір абсолютної та відносної погрішностей округлення чисел у комп'ютері.
12. Які універсальні математичні пакети ви знаєте? Дайте їх порівняльну характеристику. Для чого вони використовуються?