

# Numerical Analysis

established about 4'000 years ago

- 1. Аналіз похибок заокруглення
  - 1.1. Види похибок
  - 1.2. Підрахунок похибок в ЕОМ
  - 1.3. Підрахунок похибок обчислення значення функції

## 1. Аналіз похибок заокруглення

Література:

- Самарский, Гулин, стор. 16–25: [djvu](#), [pdf](#);
- Березин, Жидков, том I, стор. 16–35: [djvu](#), [pdf](#).

### 1.1. Види похибок

Нехай необхідно розв'язати рівняння

$$Au = f. \quad (1)$$

За рахунок неточно заданих вхідних даних насправді ми маємо рівняння

$$\tilde{A}\tilde{u} = \tilde{f}. \quad (2)$$

**Означення:** Назвемо  $\delta_1 = u - \tilde{u}$  неусувною похибкою.

Застосування методу розв'язання (2) приводить до рівняння

$$\tilde{A}_h \tilde{u}_h = \tilde{f}_h, \quad (3)$$

де  $h > 0$  — малий параметр.

**Означення:** Назвемо  $\delta_2 = \tilde{u} - \tilde{u}_h$  похибкою методу.

Реалізація методу на ЕОМ приводить до рівняння

$$\tilde{A}_h^* \tilde{u}_h^* = \tilde{f}_h^*. \quad (4)$$

**Означення:** Назвемо  $\delta_3 = \tilde{u}_h - \tilde{u}_h^*$  похибкою заокруглення.

**Означення:** Тоді повна похибка  $\delta = u - \tilde{u}_h^* = \delta_1 + \delta_2 + \delta_3$ .

**Означення:** кажуть, що задача (1) коректна, якщо

- $\forall f \in F: \exists! u \in U$ ;
- Задача (1) *стійка*, тобто  $\forall \varepsilon > 0: \exists \delta > 0$ :

$$|A - \tilde{A}| < \delta, |f - \tilde{f}| < \delta \implies |u - \tilde{u}| < \varepsilon. \quad (5)$$

Якщо задача (1) некоректна, то або розв'язок її не існує, або він неєдиний, або він нестійкий, тобто  $\exists \varepsilon > 0: \forall \delta > 0$ :

$$|A - \tilde{A}| < \delta, |f - \tilde{f}| < \delta \implies |u - \tilde{u}| > \varepsilon. \quad (6)$$

**Означення:** Абсолютна похибка  $\Delta x \leq |x - x^*|$ .

**Означення:** Відносна похибка  $\delta x \leq \Delta x / |x|$ , або  $\Delta x / |x^*|$ .

**Означення:** *Значущими цифрами* називаються всі цифри, починаючи з першої ненульової зліва.

**Означення:** *Вірна цифра* — це значуща, якщо абсолютна похибка за рахунок відкидання всіх молодших розрядів не перевищує одиниці розряду цієї цифри.

Тобто, якщо  $x^* = \alpha_n \dots \alpha_0 . \alpha_{-1} \dots \alpha_{-p} \dots$ , то  $\alpha_{-p}$  вірна, якщо  $\Delta x \leq 10^{-p}$  (інколи  $\Delta x \leq w \cdot 10^{-p}$ , де  $1/2 \leq w < 1$  наприклад,  $w = 0.55$ ).

## 1.2. Підрахунок похибок в ЕОМ

Підрахуємо відносну похибку заокруглення числа  $x$  на ЕОМ з плаваючою комою. В  $\beta$ -ічній системі числення число представляється у вигляді

$$x = \pm(\alpha_1 \beta^{-1} + \alpha_2 \beta^{-2} + \dots + \alpha_t \beta^{-t} + \dots) \cdot \beta^p, \quad (7)$$

де  $0 \leq \alpha_k < \beta$ ,  $\alpha_1 \neq 0$ ,  $k = 1, 2, \dots$

Якщо в ЕОМ  $t$  розрядів, то при відкиданні молодших розрядів ми оперуємо з наближеним значенням

$$x^* = \pm(\alpha_1 \beta^{-1} + \alpha_2 \beta^{-2} + \dots + \alpha_t \beta^{-t}) \cdot \beta^p \quad (8)$$

і відповідно похибка заокруглення

$$x - x^* = \pm \beta^p \cdot (\alpha_{t+1} \beta^{-t-1} + \dots). \quad (9)$$

Тоді її можна оцінити так

$$\begin{aligned} |x - x^*| &\leq \beta^{p-t-1} \cdot (\beta - 1) \cdot (1 + \beta^{-1} + \dots) \leq \\ &\leq \beta^{p-t-1} \cdot (\beta - 1) \cdot \frac{1}{1 - \beta^{-1}} = \beta^{p-t}. \end{aligned} \quad (10)$$

Якщо в представленні (7) взяти  $\alpha_1 = 1$ , то  $|x| \geq \beta^p \cdot \beta^{-1}$ . Звідси остаточно

$$\delta x \leq \frac{\beta^{p-t}}{\beta^{p-1}} = \beta^{-t+1}. \quad (11)$$

При більш точних способах заокруглення можна отримати оцінку  $\delta x \leq \frac{1}{2} \cdot \beta^{-t+1} = \varepsilon$ . Число  $\varepsilon$  називається «машинним іпсилон». Наприклад, для  $\beta = 2$ ,  $t = 24$ ,  $\varepsilon = 2^{-24} \approx 10^{-7}$ .

### 1.3. Підрахунок похибок обчислення значення функції

Нехай задана функція  $y = f(x_1, \dots, x_n) \in C^1(\Omega)$ . Необхідно обчислити її значення при наближеному значенні аргументів  $\vec{x}^* = (x_1^*, \dots, x_n^*)$ , де  $|x_i - x_i^*| \leq \Delta x_i$  та оцінити похибку обчислення значення функції  $y^* = f(x_1^*, \dots, x_n^*)$ . Маємо

$$\begin{aligned} |y - y^*| &= |f(\vec{x}) - f(\vec{x}^*)| = \\ &= \left| \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\vec{\xi}) \cdot (x_i - x_i^*) \right| \leq \\ &\leq \sum_{i=1}^n B_i \cdot \Delta x_i, \end{aligned} \quad (12)$$

$$\text{де } B_u = \max_{\vec{x} \in U} \left| \frac{\partial f}{\partial x_i}(\vec{x}) \right|.$$

Тут

$$U = \{ \vec{x} = (x_1, \dots, x_n) : |x_i - x_i^*| \leq \Delta x_i \} \subset \Omega, \quad (13)$$

для  $i = \overline{1, n}$ . Отже з точністю до величин першого порядку малості по

$$\Delta x = \max_i \Delta x_i, \quad (14)$$

$$\Delta y = |y - y^*| \prec \sum_{i=1}^n b_i \cdot \Delta x_i, \quad (15)$$

$$\text{де } b_i = \left| \frac{\partial f}{\partial x_i}(\vec{x}^*) \right| \text{ та «}\prec\text{» означає приблизно менше.}$$

Розглянемо похибки арифметичних операцій.

- Сума:  $y = x_1 + x_2, x_1, x_2 > 0$ :

$$\Delta y \leq \Delta x_1 + \Delta x_2, \quad (16)$$

$$\delta y \leq \frac{\Delta x_1 + \Delta x_2}{x_1 + x_2} \leq \max(\delta x_1, \delta x_2). \quad (17)$$

- Різниця:  $y = x_1 - x_2, x_1 > x_2 > 0$ :

$$\Delta y \leq \Delta x_1 + \Delta x_2, \quad (18)$$

$$\delta y \leq \frac{x_2 \delta x_1 + x_1 \delta x_2}{x_1 - x_2}. \quad (19)$$

При близьких  $x_1, x_2$  зростає відносна похибка (за рахунок втрати вірних цифр).

- Добуток:  $y = x_1 \cdot x_2, x_1, x_2 > 0$ :

$$\Delta y \leq x_2 \Delta x_1 + x_1 \Delta x_2, \quad (20)$$

$$\delta y \leq \delta x_1 + \delta x_2. \quad (21)$$

- Частка  $y = x_1/x_2, x_1, x_2 > 0$ :

$$\Delta y \leq \frac{x_2 \Delta x_1 + x_1 \Delta x_2}{x_2^2}, \quad (22)$$

$$\delta y \leq \delta x_1 + \delta x_2. \quad (23)$$

При малих  $x_2$  зростає абсолютна похибка (за рахунок зростання результату ділення).

**Означення:** *Пряма задача* аналізу похибок: обчислення  $\Delta y, \delta y$  по заданих  $\Delta x_i, i = \overline{1, n}$ .

**Означення:** *Обернена задача:* знаходження  $\Delta x_i, i = \overline{1, n}$  по заданих  $\Delta y, \delta y$ . Якщо  $n > 1$ , маємо одну умову

$$\sum_{i=1}^n b_i \cdot \Delta x_i < \varepsilon \quad (24)$$

для багатьох невідомих  $\Delta x_i$ .

Вибирають їх із однієї з умов:

$$\forall i : b_i \cdot \Delta x_i < \frac{\varepsilon}{n} \quad (25)$$

або

$$\Delta x_i < \frac{\varepsilon}{\sum_{i=1}^n b_i}. \quad (26)$$

[Назад до лекцій](#)

[Назад на головну](#)

---

**numerical-analysis** is maintained by **csc-knu**.

© 2019 Київський національний університет імені Тараса Шевченка, Андрій Риженко,  
Скибицький Нікіта