1 Аналіз похибок заокруглення

1.1 Види похибок

Нехай необхідно розв'язати рівняння

$$Au = f. \tag{1}$$

Неточно задані вхідні дані призводять до рівняння

$$\tilde{A}\tilde{u} = \tilde{f}. \tag{2}$$

Назвемо $\delta_1 = u - \tilde{u}$ *незсувною похибкою.*

Застосування методу розв'язання (2) призводить до рівняння

$$\tilde{A}_h\tilde{u}_h = \tilde{f}_h, \tag{3}$$

де $h > 0$ – малий параметр. Назвемо $\delta_2 = \tilde{u} - \tilde{u}_h$ *похибкою методу.*

Реалізація методу на ЕОМ призводить до рівняння

$$\tilde{A}^\star_h\tilde{u}^\star_h = \tilde{f}^\star_h. \tag{4}$$

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

Назвемо $\delta = u - \tilde{u}^\star_h = \delta_1 + \delta_2 + \delta_3$ *повною похибкою.*

Визначення 1. Кажуть, що задача (1) *коректна*, якщо

1. $\forall f \in F\ \exists! u \in U$;
2. задача (1) *стійка*, тобто

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall f : \left\| A - \tilde{A} \right\| < \delta, \left\| f - \tilde{f} \right\| < \delta \Rightarrow \|u - \tilde{u}\| < \varepsilon.$$

Якщо задача (1) некоректна, то або розв'язок її не існує, або він не єдиний, або він нестійкий, тобто

$$\exists \varepsilon > 0\ \forall \delta > 0\ \exists f : \left\| A - \tilde{A} \right\| < \delta, \left\| f - \tilde{f} \right\| < \delta, \|u - \tilde{u}\| > \varepsilon.$$

Абсолютна похибка $\Delta x \le |x - x^\star|$.

Відносна похибка $\delta x \le \dfrac{\Delta x}{|x|}$ або $\dfrac{\Delta x}{|x^\star|}$.

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

Вірна цифра – це значуща, якщо абсолютна похибка за рахунок відкидання всіх молодших розрядів не перевищує одиниці розряду цієї цифри. Тобто, якщо $x^\star = \overline{\alpha_n \ldots \alpha_0}.\overline{\alpha_{-1} \ldots \alpha_{-p}}$, то a_{-p} – вірна, якщо $\Delta x \le 10^{-p}$ (інколи беруть $\Delta x \le q \cdots 10^{-p}$, $1/2 \le w < 1$, наприклад $w = 0.55$).

1.2 Підрахунок похибок в ЕОМ

Обчислимо відносну похибку заокруглення числа x на ЕОМ з плавоючою комою. У системі числення з основою β число x представляється у вигляді

$$x = \pm\left(\alpha_1\beta^{-1} + \alpha_2\beta^{-2} + \ldots + \alpha_t\beta^{-t} + \ldots\right)\beta^p, \tag{5}$$

де $0 \le \alpha_k < \beta$, $\alpha_1 \ne 0$, $k = 1, 2, \ldots$

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

$$x^\star = \pm\left(\alpha_1\beta^{-1} + \alpha_2\beta^{-2} + \ldots + \alpha_t\beta^{-t}\right)\beta^p$$

і, відповідно, похибка заокруглення $x - x^\star = \pm\beta^p\left(\alpha_{t+1}\beta^{-t-1} + \ldots\right)$. Її можна оцінити так:

$$|x - x^\star| \le \beta^{p-t-1}(\beta-1)(1+\beta^{-1}+\ldots) \le \beta^{p-t-1}(\beta-1)\frac{1}{1-\beta^{-1}} = \beta^{p-t}.$$

Враховуючи, що $\alpha_1 \ne 0$, маємо $|x| \ge \beta^p\beta^{-1} = \beta^{p-1}$. Звідси остаточно

$$\delta x \le \frac{\beta^{p-t}}{\beta^{p-1}} = \beta^{-t+1}.$$

При точніших способах заокруглення можна отримати оцінку $\delta x \le \beta^{-t+1}/2 = \varepsilon$. Число ε називаєтсья *машинним іпсилон.* Наприклад, для $\beta = 2$, $t = 24$, $\varepsilon = 2^{-24} \approx 10^{-7}$.

1.3 Обчислення похибок обчислення значення функції

Нехай задана функція $y = f(x_1, \ldots, x_n) \in C^{(1)}(\Omega)$. Необхідно обчислити її значення при наближеному значенні агрументів $\vec{x}^\star = (x_1^\star, \ldots, x_n^\star)$, де $|x_i - x_i^\star| \le \Delta x_i$ та оцінити похибку обчислення значення функції $y^\star = f\left(x_1^\star, \ldots, x_n^\star\right)$.

Маємо

$$|y - y^\star| = |f(\vec{x}) - f(\vec{x}^\star)| = \left|\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\vec{\xi})(x_i - x_i^\star)\right| \le \sum_{i=1}^n B_i \cdots \Delta x_i,$$

де $B_i = \max\limits_{\vec{x}\in U}\left|\dfrac{\partial f}{\partial x_i}(\vec{x})\right|$, $U = \{\vec{x} = (x_1, \ldots, x_n) : |x_i - x_i^\star| \le \Delta x_i, i = 1, \ldots, n\} \subset \Omega$.

Отже, з точністю до величин першого порядку малості по $\Delta x = \max\limits_i \Delta x_i$, $\Delta y = |y - y^\star| \prec \sum\limits_{i=1}^n b_i \cdots \Delta x_i$, де $b_i = \left|\dfrac{\partial f}{\partial x_i}\left(\vec{x}^\star\right)\right|$ і "$\prec$" означає *приблизно менше.*

Розглянемо похибки арифметичних операцій.

1. Сума: $y = x_1 + x_2$, $x_1, x_2 > 0$, $\Delta y \le \Delta x_1 + \Delta x_2$, $\delta y = \dfrac{\Delta y}{y} = \dfrac{\Delta y}{x_1 + x_2} \le \dfrac{\Delta x_1 + \Delta x_2}{x_1 + x_2} \le \max(\delta x_1, \delta x_2)$.

2. Різниця: $y = x_1 - x_2$, $x_1 > x_2 > 0$, $\Delta y \le \Delta x_1 + \Delta x_2$, $\delta y = \dfrac{\Delta y}{y} = \dfrac{\Delta y}{x_1 - x_2} \le \dfrac{\Delta x_1 + \Delta x_2}{x_1 - x_2} = \dfrac{x_1\delta x_1 + x_2\delta x_2}{x_1 - x_2}$.
   Як бачимо, при близьких аргументах зростає відносна похибка.

3. Добуток: $y = x_1 \cdot x_2$, $x_1, x_2 > 0$, $\Delta y = x_1\Delta x_2 + x_2\Delta x_1 + \Delta x_1\Delta x_2 \prec x_1\Delta x_2 + x_2\Delta x_1$, $\delta y = \dfrac{\Delta y}{y} = \dfrac{\Delta y}{x_1 x_2} \prec \dfrac{x_1\Delta x_2 + x_2\Delta x_1}{x_1 x_2} = \dfrac{\Delta x_1}{x_1} + \dfrac{\Delta x_2}{x_2} = \delta x_1 + \delta x_2$.

4. Частка: $y = \dfrac{x_1}{x_2}$, $x_1, x_2 > 0$, $\Delta y = \dfrac{x_2\Delta x_1 - x_1\Delta x_2}{x_2(x_2 + \Delta x_2)} < \dfrac{x_2\Delta x_1 + x_1\Delta x_2}{x_2(x_2 + \Delta x_2)} \prec \dfrac{x_2\Delta x_1 + x_1\Delta x_2}{x_2^2}$, $\delta y = \dfrac{\Delta y}{y} = \dfrac{x_2\Delta y}{x_1} \prec \dfrac{x_2\Delta x_1 + x_1\Delta x_2}{x_1 x_2} = \dfrac{\Delta x_1}{x_1} + \dfrac{\Delta x_2}{x_2} = \delta x_1 + \delta x_2$.
   Як бачимо, при малих x_2 зростає абсолютна похибка.

Пряма задача аналізу похибок: обчислення Δy, δy за заданими $\Delta x_i, i = 1, \ldots, n$.

Обернена задача: знаходження $\Delta x_i, i = 1, \ldots, n$ за заданими Δy, δy. Якщо $n > 1$ маємо одну умову $\sum_{i=1}^{n} b_i \Delta x_i < \varepsilon$ на багато невідомих Δx_i. Зазвичай вибирають їх із однієї з умов

$$\forall i : b_i \Delta x_i < \varepsilon / n \quad \text{або} \quad \forall i : \Delta x_i < \varepsilon / B, \quad \text{де } B = \sum_{i=1}^{n} b_i.$$