

Семинар 2

Варламов Антоний Михайлович

15 сентября 2021 г.

1 Погрешности округления. Погрешности арифметики с плавающей точкой

Существуют различные типы данных:

1. Integer (целые числа) – используется 4 байта/8 байт
2. Real (действительные числа) – используется 4 байта (single precision) или 8 байт (double precision) (Альтернативные названия – float и double). Число хранится в виде:

$$a = \pm 2^{\pm e} (1 + f) \quad (1)$$

single precision: OFL = 10^{38}

double precision: OFL = 10^{324}

Представление чисел в компьютере – дискретное. Плотность распределения чисел непостоянна.

Характерные особенности компьютерной арифметики:

1. $a := 0.1$ но $a \neq 0.1$

2. $10^{20} + 1 = 10^{20}$
 $(10^{20} + 1) - 10^{20} = 0$
 $(10^{20} - 10^{20}) + 1 = 0$

Данные следствия говорят о том, что ноль в компьютерном представлении не единственен.

3. Все числа в компьютере неточны. Степень неточности:

$$\delta a = \frac{\Delta a}{a} \leq \varepsilon_{\text{маш}} \quad (2)$$

4. Все операции неточны. Для обозначения "компьютерных" операций используются символы операции в кружке.

$$a \oplus b = (a + b) (1 \pm \delta) \quad (3)$$

Для анализа обычно используются два метода:

- а Все операции точны, но числа имеют погрешность.
- б Все операции неточны, но числа точны

Рассмотрим пример применения разных методов:

$$a_1 + a_2 + a_3 \quad (4)$$

$$(a_1 + a_2) (1 + \delta_1) \oplus a_3 = \quad (5)$$

$$((a_1 + a_2) (1 + \delta_1) + a_3) (1 + \delta_2) = \quad (6)$$

$$a_1 + a_2 + a_3 + \delta_1 (a_1 + a_2) + \delta_1 \delta_2 (a_1 + a_2) + \delta_2 (a_1 + a_2 + a_3) \approx \quad (7)$$

$$a_1 + a_2 + a_3 + a_1 (\delta_1 + \delta_2) + a_2 (\delta_1 + \delta_2) + a_3 \delta_2 \quad (8)$$

Вспомним прошлые результаты:

$$\left| f'_i - \frac{f_{i+1} - f_i}{h} \right| \leq \frac{h}{2} M_2 \quad (9)$$

$$\left| f'_i - \frac{f_{i+1} - f_{i-1}}{2h} \right| \leq \frac{h^2}{6} M_3 \quad (10)$$

Проведем некоторый анализ:

$$\tilde{f}_i = f_i (1 + \delta) \quad (11)$$

$$\frac{\tilde{f}_{i+1} - \tilde{f}_i}{h} - \frac{f_{i+1} - f_i}{h} = \frac{\delta_1 f_{i+1} - \delta_2 f_i}{h} \quad (12)$$

$$\left| \frac{\delta_1 f_{i+1} - \delta_2 f_i}{h} \right| = \frac{|\delta_1| f_{i+1} + |\delta_2| f_{i+1}}{h} \leq \frac{2\varepsilon_{\text{маш}} M}{h} \quad (13)$$

$$err_1 = \frac{h}{2} M_2 + \frac{2\varepsilon_{\text{маш}} M}{h} \quad (14)$$

$$\left| \frac{\tilde{f}_{i+1} - \tilde{f}_{i-1}}{2h} - \frac{f_{i+1} - f_{i-1}}{2h} \right| = \frac{\varepsilon_{\text{маш}} M}{h} \quad (15)$$

$$err_2 = \frac{h^2}{6} M_3 + \frac{\varepsilon_{\text{маш}} M}{h} \quad (16)$$

Нарисуем картинку:

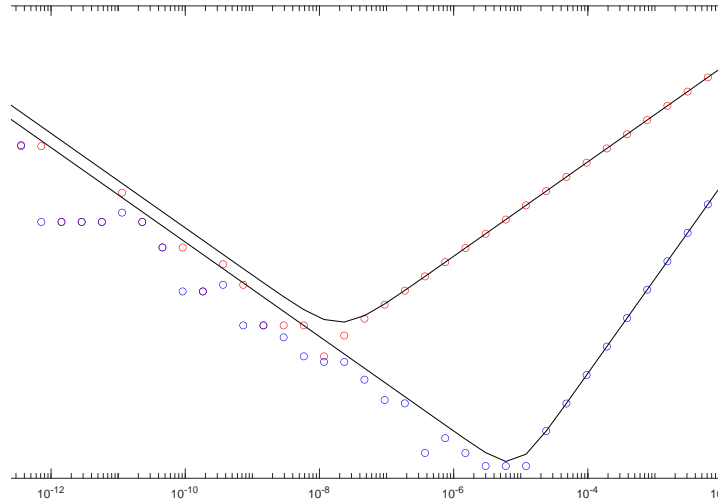


Рис. 1: Визуализация погрешностей вычисления на компьютере

2 Основы линейной алгебры

Норма