Analiza numeryczna

1. Analiza błędów

Rafał Nowak

Błędy

Niech $\tilde{x}$ będzie przybliżoną wartością wielkości x.

- błąd bezwzględny
$$\Delta x := |\tilde{x} - x|;$$
- błąd względny
$$\delta x := |\tilde{x} - x| / |x| \qquad (x \neq 0).$$

Symbol $|\cdot|$ może oznaczać dowolną normę, tzn. $|x - y|$ jest odległością x od y.

$$x := (e_n \ldots e_1 e_0 . e_{-1} e_{-2} \ldots)_B = \pm \left(\sum_{i=0}^{n} e_i B^i + \sum_{j=1}^{\infty} e_{-j} B^{-j} \right).$$

- $B \geqslant 2$ - liczba całkowita - podstawa systemu; najczęściej $B = 2, 10$.
- $0 \leqslant e_i \leqslant B - 1$ - liczby całkowite - cyfry liczby x

Cyfry dokładne vs cyfry znaczące

Niech $B = 10$ ($B = 2$) oraz niech $\tilde{a}$ będzie przybliżoną wartością wielkości a.

- jeśli $|a - \tilde{a}| \leqslant \frac{1}{2} \cdot B^{-p}$ to $\tilde{a}$ ma p **dokładnych cyfr** dziesiętnych (dwójkowych) ułamkowych.
- ponadto, jeśli w reprezentacji liczby $\tilde{a}$ jest $e_n = e_{n-1} = \ldots = e_{q+1} = 0$, $e_q \neq 0$ to cyfry $e_q, e_{q-1}, \ldots, e_p$ nazywamy dziesiętnymi (dwójkowymi) cyframi znaczącymi liczby $\tilde{a}$.

Cyfry dokładne vs cyfry znaczące

Niech $B = 10$ $(B = 2)$ oraz niech $\tilde{a}$ będzie przybliżoną wartością wielkości a.

- jeśli $|a - \tilde{a}| \leqslant \frac{1}{2} \cdot B^{-p}$ to $\tilde{a}$ ma p **dokładnych cyfr** dziesiętnych (dwójkowych) ułamkowych.
- ponadto, jeśli w reprezentacji liczby $\tilde{a}$ jest $e_n = e_{n-1} = \ldots = e_{q+1} = 0$, $e_q \neq 0$ to cyfry $e_q, e_{q-1}, \ldots, e_p$ nazywamy dziesiętnymi (dwójkowymi) cyframi znaczącymi liczby $\tilde{a}$. **Przykład**: niech będzie $a = 0.00045675$; liczba $\tilde{a} = 0.000\underline{4567}9$ ma 7 dokładnych cyfr ułamkowych oraz cztery cyfry znaczące: 4, 5, 6, 7.
- Przykład: liczba 0.001234 ± 0.000004 ma pięć cyfr dokładnych, z czego trzy są znaczące.
- Przykład: liczba 0.001234 ± 0.000006 ma cztery cyfry dokładne i tylko dwie cyfry znaczące.

- znormalizowana zmiennopozycyjna postać

$$x = s\, m\, B^c,$$

- $s = \operatorname{sgn} x$ — znak liczby x
- $1 \leqslant m < B$ — mantysa
- c - liczba całkowita — cecha

Reprezentacja dwójkowa

- $B = 2,\ x = s\, m\, 2^c$
- $m = (1.e_{-1}e_{-2}\dots)_2 = 1 + \sum_{i=1}^{\infty} e_{-i}2^{-i} \in [1, 2)$
- $d + 1$ — długość słowa (32 = float, 64 = double w języku C)
- $t \in \mathbb{N}$ — liczba bitów na mantysę
- $m_t = (1.e^*_{-1}e^*_{-2}\dots e^*_{-t})_2,$ — zaokrąglenie mantysy

Definicja (Reguła zaokrąglenia)

Zaokrąglenie liczby x

$$\mathrm{rd}\,(x) := s\,\bar{m}\,2^c, \tag{1}$$

gdzie

$$\bar{m} = (1.e_{-1}e_{-2}\dots e_{-t})_2 + (0.\underbrace{00\cdots0}_{t-1 \text{ razy}}\, e_{-t-1})_2$$

Twierdzenie

Liczbę $\mathrm{rd}\,(x)$ *można zapisać w postaci*

$$\mathrm{rd}\,(x) = s\, m_t\, 2^{c_t}, \tag{2}$$

gdzie mantysa $m_t = 1.e^*_{-1}e^*_{-2}\dots e^*_{-t}$ *i cecha* $c_t \in \mathbb{Z}$ *są dane wzorami*

$$m_t := 1.0, \quad c_t := c+1$$

jeśli

$$e_{-k} = 1 \quad \textit{dla} \quad k = 1, 2, \dots, t+1,$$

lub wzorami

$$m_t := \bar{m}, \quad c_t := c$$

w przeciwnym wypadku.

Precyzja arytmetyki

Twierdzenie

Błąd bezwzględny zaokrąglenia spełnia nierówność

$$|\operatorname{rd}(x) - x| \leqslant 2^{-t-1} \cdot 2^c.$$

Twierdzenie

Błąd względny zaokrąglenia spełnia nierówność

$$\left|\frac{\operatorname{rd}(x) - x}{x}\right| \leqslant \frac{1}{2} 2^{-t}.$$

Definicja

Precyzją arytmetyki danego komputera nazywamy liczbę

$$\mathsf{u} := \frac{1}{2} 2^{-t}.$$

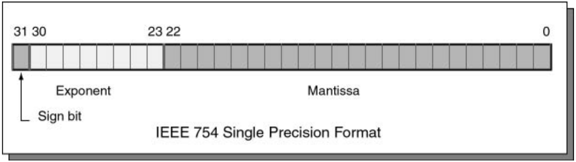


Tabela: Formaty liczb zmiennopozycyjnych (IEEE 754)

		single	double
$d+1$	długość słowa (w bitach)	32	64
t	długość mantysy (w bitach)	23	52
$d-t$	długość cechy (w bitach)	8	11
$c_{\max}$	największa cecha	127	1023
$c_{\min}$	najmniejsza cecha	-126	-1022
	największa liczba dod.	$3.4 \cdot 10^{38}$	$1.8 \cdot 10^{308}$
	najmniejsza liczba dod.	$1.2 \cdot 10^{-38}$	$2.2 \cdot 10^{-308}$
	najmn. dod. liczba subnorm.	$1.4 \cdot 10^{-45}$	$4.9 \cdot 10^{-324}$
u	precyzja arytmetyki	$5.96 \cdot 10^{-8}$	$1.11 \cdot 10^{-16}$

Zbiór reprezentacji arytmetyki zmiennopozycyjnej

$$X_{fl} := \mathrm{rd}\,(X) = \{\mathrm{rd}\,(x) : x \in X\}$$

Założenie (**Model standardowy arytmetyki**)

Niech będzie a, $b \in X_{fl}$, $\diamond \in \{+, -, \times, /\}$, $a \diamond b \in X'$,
$\mathrm{fl}(a \diamond b) := \mathrm{rd}\,(a \diamond b)$ — **obliczony** wynik spełnia

$$fl(a \diamond b) = (a \diamond b)(1 + \varepsilon_\diamond), \tag{3}$$

gdzie $\varepsilon_\diamond = \varepsilon_\diamond(a, b)$, $|\varepsilon_\diamond| \leqslant \mathsf{u}$.

Twierdzenie

Jeśli $|\alpha_j| \leqslant u$ *i* $\rho_j = \pm 1$ *dla* $j = 1, 2, \ldots, n$ *oraz* $nu < 1$, *to zachodzi równość*

$$\prod_{j=1}^{n} (1 + \alpha_j)^{\rho_j} = 1 + \theta_n, \tag{4}$$

gdzie θ_n *jest wielkością spełniającą nierówność*

$$|\theta_n| \leqslant \gamma_n,$$

gdzie z kolei

$$\gamma_n := \frac{nu}{1 - nu} \approx nu. \tag{5}$$

Twierdzenie

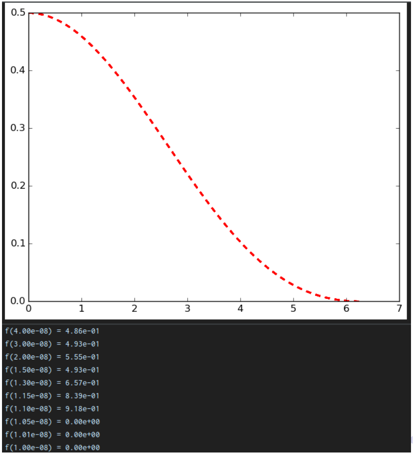
Jeśli $|\alpha_j| \leqslant u$ *dla* $j = 1, 2, \ldots, n$ *oraz* $nu < 0.01$*, to zachodzi równość*

$$\prod_{j=1}^{n}(1 + \alpha_j) = 1 + \eta_n, \tag{6}$$

gdzie $|\eta_n| \leqslant 1.01nu$*.*

Utrata cyfr znaczących

Utrata cyfr znaczących występuje wtedy, gdy odejmujemy dwie prawie równe liczby.
Przykład: $f(x) = (1 - \cos(x))/x^2$



Uwarunkowanie zadania

Definicja

Jeśli niewielkie względne zmiany danych zadania powodują duże względne zmiany jego rozwiązania, to zadanie takie nazywamy **źle uwarunkowanym**. Wielkości charakteryzujące wpływ zaburzeń danych na odkształcenia rozwiązania nazywamy **wskaźnikami uwarunkowania** zadania.

Przykład

Zadanie: obliczyć wartość funkcji f w punkcie $x \in \mathbb{R}$.

$$\frac{|f(x+h) - f(x)|}{|f(x)|} \approx \frac{|hf'(x)|}{|f(x)|} = \frac{|xf'(x)|}{|f(x)|} \frac{|h|}{|x|} = C_f(x) \cdot \frac{|h|}{|x|}.$$

Czynnik $C_f(x) = |xf'(x)|/|f(x)|$ można traktować jako *wskaźnik uwarunkowania* zadania.

Algorytmy numerycznie poprawne

Problem: jak dokładny może być dla wybranego zadania wynik obliczony w arytmetyce zmiennopozycyjnej?

Definicja

Algorytmem *numerycznie poprawnym* nazywamy taki algorytm, dla którego **obliczone rozwiązanie jest mało zaburzonym rozwiązaniem dokładnym dla mało zaburzonych danych.** Przez „małe zaburzenia" rozumiemy tu zaburzenia na poziomie błędu reprezentacji.