

MOwNiT – arytmetyka komputerowa

1. Wyznaczyć doświadczalnie parametry reprezentacji liczb zmiennoprzecinkowych (*float*, *double*, *long double*) i porównać uzyskane wartości dla różnych architektur, systemów operacyjnych i kompilatorów. Sprawdzić, czy reprezentacje są zgodne ze standardem IEEE.

Parametry do wyznaczenia:

- o liczba bajtów używana do przechowywania zmiennej danego typu,
 - o liczba bitów na mantysę,
 - o liczba bitów na cechę (wliczając znak),
 - o "maszynowe epsilon" - najmniejsza liczba, dla której $1.0 + e > 1.0$,
 - o występowanie i sposób reprezentacji wartości specjalnych (± 0 , $\pm \text{Inf}$, NaN).
2. Wykonać obliczenia (dla zmiennych typu *float*, *double*, *long double*) wg podanych poniżej wzorów dla 101 równoodległych wartości x z przedziału $[0.99, 1.01]$:

- $f(x) = x^8 - 8x^7 + 28x^6 - 56x^5 + 70x^4 - 56x^3 + 28x^2 - 8x + 1$
- $f(x) = ((((((x - 8)x + 28)x - 56)x + 70)x - 56)x + 28)x - 8)x + 1$
- $f(x) = (x - 1)^8$
- $f(x) = e^{(8 \ln(\text{abs}(x-1)))}$, $x \neq 1$

Porównać wyniki. Objasnić różnice w wynikach.

3. Wyznaczyć wartości funkcji $f(x) = \sqrt{x^2 + 1} - 1$, $g(x) = x^2 / (\sqrt{x^2 + 1} + 1)$, dla argumentu $x = 8^{-1}, 8^{-2}, 8^{-3}, \dots$. Sprawdzić, czy wyznaczone wartości dla obu funkcji (matematycznie tożsamy) są takie same i spróbować uzasadnić ewentualne różnice. Jak obliczać z kolei wartości dla dużych argumentów (np. x bliskiego największej liczbie typu *double*)? Obliczenia wykonać dla zmiennych typu *float*, *double*, *long double*.

4. Rozważmy funkcję $f(x) = (e^x - 1) / x$. Można wykazać (jak?), że $\lim_{x \rightarrow 0} f(x) = 1$

- a) Sprawdzić tę własność obliczając wartość funkcji dla $x = 10^{-k}$ dla $k = 1, \dots, 15$. Czy rezultaty są zgodne z oczekiwaniami?
- b) Powtórzyć obliczenia tym razem korzystając ze wzoru:

$$f(x) = (e^x - 1) / \ln(e^x).$$

Porównać rezultaty z tymi uzyskanymi w punkcie a). Spróbować wyjaśnić ewentualne różnice.

Obliczenia wykonać dla zmiennych typu *float*, *double*, *long double*. Zwrócić uwagę na typ argumentów i wyników dla funkcji bibliotecznych wykorzystywanych w obliczeniach.

5. W dokładnej arytmetyce ciąg $x_{k+1} = 111 - (1130 - 3000/x_{k-1})/x_k$, $x_0 = 11/2$, $x_1 = 61/11$ jest rosnący i zbieżny do 6. Obliczyć x_{34} (dla zmiennych typu *float*, *double*, *long double*) i spróbować wyjaśnić uzyskane wyniki. Dokładna wartość (zaokrąglona do 4 cyfr znaczących) jest równa $x_{34} = 5.998$.
6. Wyznaczyć kolejne elementy ciągu $x_{k+1} = x_k + 3x_k(1 - x_k)$, $x_0 = 0.1$, i porównać otrzymane wartości dla różnej precyzji zmiennych (*float*, *double*, *long double*). Powtórzyć doświadczenie dla przekształconej postaci wzoru: $x_{k+1} = 4x_k - 3x_kx_k$. Spróbować wyjaśnić otrzymane wyniki.
7. Dana jest zależność rekurencyjna $3x_{k-1} - 10x_k + 3x_{k+1} = 0$. Wartości początkowe $x_0 = 1$, $x_1 = 1/3$. Wyznaczyć wartości x_k i x_{k+1} dla $k = 45$. Następnie korzystając z wyznaczonych wartości x_k i x_{k+1} obliczyć x_1 i x_0 , wykonując rekurencję w tył. Porównać wyznaczone wartości x_1 i x_0 z dokładnymi wartościami początkowymi 1 i 1/3. Wykonać obliczenia dla różnej precyzji zmiennych (*float*, *double*, *long double*). Skomentować różnice. Co będzie, jeśli wszędzie liczbę 3 zastąpimy przez liczbę 2 lub 20, lub 30?
8. Niech ciąg x_k będzie zdefiniowany:

$$x_1 = 4, \quad x_{k+1} = 2^{2^{(k+1)+1}} \cdot \frac{\sqrt{1 + x_k^2 / 2^{2^{(k+1)}}} - 1}{x_k}$$

Zaproponować inną postać tego związku i obliczyć x_{30} dwoma sposobami. Obliczenia wykonać dla zmiennych typu *float*, *double*, *long double*. Skomentować i spróbować objaśnić otrzymane wyniki.

9. Dla $x_0 > 1$ ciąg

$$x_{k+1} = 2^{k+1}(\sqrt{1 + 2^{-k}x_k} - 1)$$

jest zbieżny do $\log(x_0+1)$. Przekształcić wzór tak, by uniknąć utraty dokładności. Porównać dwa sposoby liczenia wyrazów ciągu. Obliczenia wykonać dla zmiennych typu *float*, *double*, *long double*.

10. Przybliżoną wartość pochodnej funkcji $f(x)$ w punkcie x można obliczyć ze wzoru:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

Wykorzystać ten wzór do obliczenia pochodnej funkcji $f(x) = \sin(x) + \cos(3x)$ w punkcie $x = 1$ dla $h = 2^{-n}$ ($n = 0, 1, 2, \dots, 40$). Wykonać obliczenia dla różnej precyzji zmiennych. Zwrócić uwagę na typ argumentów i wyników dla funkcji bibliotecznych wykorzystywanych w obliczeniach. Jak wytłumaczyć, że od pewnego momentu zmniejszenie wartości h nie poprawia przybliżenia wartości pochodnej? Jak zachowują się wartości $1+h$? Obliczone przybliżenia pochodnej porównać z dokładną wartością pochodnej.

11. Niech

$$x_n = \int_0^1 t^n (t+5)^{-1} dt$$

Całka powyższa może być obliczona za pomocą wzoru rekurencyjnego:

$$x_n = 1/n - 5x_{n-1} \quad (*)$$

przy czym $x_0 = \ln 6 - \ln 5 = \ln 1.2$.

Stosując ten wzór rekurencyjny obliczyć x_0, x_1, \dots, x_{20} . Zwrócić uwagę, czy i dla jakich n obliczony wynik jest ujemny. Czy tak powinno być?

Obliczyć analitycznie x_{20} wykorzystując do funkcji podcałkowej szereg Taylora, a następnie na podstawie tego wzoru wyliczyć w komputerze wartość x_{20} . Po wyznaczeniu x_{20} zastosować rekurencję wstecz wykorzystując odpowiednio przekształcony wór (*), tj. obliczyć kolejno $x_{19}, x_{18}, \dots, x_0$. Czy otrzymane x_0 jest poprawne? Wykonać obliczenia dla różnej precyzji zmiennych (*float, double, long double*). Skomentować i spróbować wyjaśnić otrzymane wyniki.

12. Dany jest wielomian Wilkinsona $W(x) = (x-1)(x-2)\dots(x-20)$. Przekształcić wielomian do postaci $W(x) = x^{20} + a_{19}x^{19} + \dots + a_1 x + a_0$. Podać wyznaczone wartości współczynników a_i . Dla $x = 1$ oraz $x = 20$ dokładna wartość wielomianu wynosi 0. Obliczyć wartość wielomianu dla $x = 1$ oraz $x = 20$. Do wyznaczenia wartości wielomianu wykorzystać schemat Hornera oraz dowolny schemat sumacyjny. Obliczenia wykonać dla zmiennych typu *float, double, long double*.

13. Niech

$$r_n = \frac{1}{2n+1} + \frac{2}{\pi} \sum_{k=1}^n \frac{1}{k} \operatorname{tg} \frac{\pi k}{2n+1}$$

Obliczyć r_1, \dots, r_{100} . Sprawdzić, czy zachodzi:

$$0 \leq \frac{4}{\pi^2} \ln(2n+1) + 1 - r_n \leq 0.0106.$$

Obliczenia wykonać dla zmiennych typu *float, double, long double*. Zwrócić uwagę na typ argumentów oraz wyników dla funkcji bibliotecznych wykorzystywanych w obliczeniach. Skomentować wyniki obliczeń ciągu r_n .