Obliczenia Naukowe Sprawozdanie Lista 1

Mateusz Gancarz

8 listopada 2022

1 Zadanie 1.

1.1 Opis zadania

W zadaniu musieliśmy wyznaczyć w arytmetyce Float16, Float32 oraz Float64:

- epsilon maszynowy, czyli *macheps*, będący odległością od 1.0 do następnej liczby w arytmetyce zmiennopozycyjnej
- *eta*, czyli najmniejszą dodatnią liczbę
- *max*, czyli największą liczbę

1.2 Opis rozwiązania i wyniki

Aby otrzymać dane wartości, możemy iteracyjnie dzielić lub mnożyć początkowe wartości, dopóki nie będą spełniały określonego warunku. Dane wartości możemy również otrzymać za pomocą wbudowanych funkcji języku *Julia* i z ich pomocą sprawdzimy poprawność programu. Kod źródłowy znajduje się w pliku *ex1.jl*.

- *Float16*:
  - $macheps = 9.77 \cdot 10^{-4}$
  - $eps = 9.77 \cdot 10^{-4}$
  - $eta = 5.96 \cdot 10^{-8}$
  - $nextfloat = 5.96 \cdot 10^{-8}$
  - $maxnum = 6.55 \cdot 10^{4}$
  - $floatmax = 6.55 \cdot 10^{4}$
- *Float32*:
  - $macheps = 1.1920929 \cdot 10^{-7}$
  - $eps = 1.1920929 \cdot 10^{-7}$
  - (*float.h*) FLT_EPSILON $= 1.1920928955078125 \cdot 10^{-7}$

- *eta* $= 1.401 \cdot 10^{-45}$
- *nextfloat* $= 1.401 \cdot 10^{-45}$
- *maxnum* $= 3.4028235 \cdot 10^{38}$
- *floatmax* $= 3.4028235 \cdot 10^{38}$

- *Float64*:
  - *macheps* $= 2.220446049250313 \cdot 10^{-16}$
  - *eps* $= 2.220446049250313 \cdot 10^{-16}$
  - (*float.h*) DBL_EPSILON $= 2.2204460492503131 \cdot 10^{-16}$
  - *eta* $= 4.941 \cdot 10^{-324}$
  - *nextfloat* $= 4.941 \cdot 10^{-324}$
  - *maxnum* $= 1.7976931348623157 \cdot 10^{308}$
  - *floatmax* $= 1.7976931348623157 \cdot 10^{308}$

W poleceniu również otrzymaliśmy zadanie wytłumaczenia następujących pytań:

- Jaki związek ma liczba *macheps* z precyzją arytmetyki (oznaczaną na wykładzie przez ϵ)?
  Obie liczby mają tą samą wartość, lecz ϵ opisuje największy możliwy błąd względny przy zaokrąglaniu liczby rzeczywistej, a *macheps* opisuje największy możliwy błąd względny, jaki może pojawić się przy liczbie 1.0.
- Jaki związek ma liczba *eta* z liczbą MIN_{sub}? Obie liczby opisują tę samą wartość, czyli najmniejszą dodatnią liczbę w arytmetyce float.

2 Zadanie 2.

W tym zadaniu mieliśmy sprawdzić metodę obliczenia *macheps* za pomocą metody Kahana $3 \cdot (4/3 - 1)$. Kod źródłowy znajduje się w pliku *ex2.jl*.

- *Float16*:
  - *Kahan macheps* $= -9.77 \cdot 10^{-4}$
  - *macheps* $= 9.77 \cdot 10^{-4}$
- *Float32*:
  - *Kahan macheps* $= 1.1920929 \cdot 10^{-7}$
  - *macheps* $= 1.1920929 \cdot 10^{-7}$
- *Float64*:
  - *Kahan macheps* $= -2.220446049250313 \cdot 10^{-16}$
  - *macheps* $= 2.220446049250313 \cdot 10^{-16}$

3 Zadanie 3.

3.1 Opis zadania

W tym zadaniu musieliśmy sprawdzić czy w arytmetyce *Float64* liczby zmiennoprzecinkowe są równomiernie rozmieszczone w $[1, 2]$ z krokiem $\delta = 2^{-52}$. W dalszej części zadania musieliśmy również sprawdzić jak rozmieszczone są liczby zmiennoprzecinkowe w przedziałach $[0.5, 1.0]$ oraz $[2.0, 4.0]$. Kod źródłowy znajduje się w pliku *ex3.jl*.

3.2 Opis rozwiązania i wyniki

Rozwiązanie polega na sprawdzeniu czy eksponenty w reprezentacji liczb 1.0 oraz 2.0 są takie same (sprzeczność tej równości wyklucza równomierne rozłożenie w tym przedziale, o czym później się przekonamy), a następnie sprawdzeniu czy odległość między następnymi liczbami w tym przedziale jest równa $\delta = 2^{-52}$.

Dla sprawdzenia rozmieszczenia liczb zmiennoprzecinkowych w przedziałach $[0.5, 1.0]$ oraz $[2.0, 4.0]$ wykonujemy podobne postępowanie, zwracając na koniec wartość między kolejnymi liczbami w przedziale. Oto zwrócone wartości:

- przedział $[0.5, 1.0]$ rozmieszczenie z krokiem 1.1102230246251565^{-16}
- przedział $[2.0, 4.0]$ rozmieszczenie z krokiem 4.440892098500626^{-16}

4 Zadanie 4.

4.1 Opis zadania

W tym zadaniu musieliśmy znaleźć eksperymentalnie w arytmetyce *Float64* liczbę zmiennopozycyjną x w przedziale $1 < x < 2$, taką, że $x \cdot (1/x) \neq 1$ oraz następnie znaleźć najmniejszą taką liczbę. Kod źródłowy do zadania znajduje się w pliku *ex4.jl*.

4.2 Opis rozwiązania i wyniki

Aby znaleźć taką liczbę, mogliśmy iteracyjnie sprawdzić każdą liczbę zmiennoprzecinkową z przedziału $[1.0, 2.0]$ oraz zatrzymać pętlę i zwrócić wartość, jeśli dana liczba spełniłaby warunek z treści zadania. Tym samym sposobem mogliśmy znaleźć najmniejszą taką liczbę.

- $1 < x < 2; x \cdot (1/x) \neq 1; x = 1.000000057228997$
- $y \cdot (1/y) \neq 1; y = 1.0^{-323}$

5 Zadanie 5.

5.1 Opis zadania

W tym zadaniu musieliśmy obliczyć iloczyn skalarny podanych wektorów na cztery sposoby w arytmetyce *Float32* oraz *Float64*. Kod źródłowy znajduje się w pliku *ex5.jl.*

$x = [2.718281828, 3.141592654, 1.414213562, 0.5772156649, 0.3010299957]$
$y = [1486.2497, 878366.9879, 22.37492, 4773714.647, 0.000185049]$

5.2 Rozwiązanie i wyniki

- *Float32*
  - prawdziwa wartość - -1.00657107000000^{-11}
  - metoda "w przód -0.3472038161853561
  - metoda "w tył -0.3472038162872195
  - od największego do najmniejszego - -0.5
  - od najmniejszego do największego - -0.5
- *Float64*
  - prawdziwa wartość - -1.00657107000000^{-11}
  - metoda "w przód 1.0251881368296672^{-10}
  - metoda "w tył $-1.5643308870494366^{-10}$
  - od największego do najmniejszego - 0.0
  - od najmniejszego do największego - 0.0

6 Zadanie 6.

W tym zadaniu musieliśmy policzyć wartości funkcji $f(x) = \sqrt{x^2+1} - 1$ oraz $g(x) = \frac{x^2}{\sqrt{x^2+1}+1}$ dla wartości argumentu $x = 8^{-1}, 8^{-2}, 8^{-3}, \ldots$. Kod źródłowy znajduje się w pliku *ex6.jl.*

6.1 Rozwiązanie i wyniki

- 8^{-1}
  $f(x) = 0.0077822185373186414$
  $g(x) = 0.0077822185373187065$
- 8^{-2}
  $f(x) = 0.00012206286282867573$
  $g(x) = 0.00012206286282875901$

- 8^{-3}
  $f(x) = 1.9073468138230965^{-6}$
  $g(x) = 1.907346813826566^{-6}$
- 8^{-4}
  $f(x) = 2.9802321943606103^{-8}$
  $g(x) = 2.9802321943606116^{-8}$
- 8^{-5}
  $f(x) = 4.656612873077393^{-10}$
  $g(x) = 4.6566128719931904^{-10}$
- 8^{-6}
  $f(x) = 7.275957614183426^{-12}$
  $g(x) = 7.275957614156956^{-12}$
- 8^{-7}
  $f(x) = 1.1368683772161603^{-13}$
  $g(x) = 1.1368683772160957^{-13}$
- 8^{-8}
  $f(x) = 1.7763568394002505^{-15}$
  $g(x) = 1.7763568394002489^{-15}$
- 8^{-9}
  $f(x) = 0.0$
  $g(x) = 2.7755575615628914^{-17}$
- 8^{-10}
  $f(x) = 0.0$
  $g(x) = 4.336808689942018^{-19}$

6.2 Wnioski

Inna postać funkcji może wpłynąć na końcowe wyniki naszego programu. Widać to po wynikach funkcji f(x), która po pewnym czasie zwraca wartość 0.0, gdy pod inną postacią funkcja g(x) zwraca wartości różne od 0.0.

7 Zadanie 7.

7.1 Opis zadania

W tym zadaniu musieliśmy obliczyć pochodną ze wzoru $f'(x_0) \approx \tilde{f}'(x_0) = \frac{f(x_0+h)-f(x_0)}{h}$ dla funkcji $f(x) = \sin x + \cos 3x$ oraz porównać wartość z faktyczną pochodną.

7.2 Rozwiązanie i wyniki

- 2^{-25}
  $\tilde{f}'(x) = 0.116942398250103$
  $|f'(x) - \tilde{f}'(x)| = 1.1656156484463054^{-7}$
- 8^{-28}
  $\tilde{f}'(x) = 0.11694228649139404$
  $|f'(x) - \tilde{f}'(x)| = 4.802855890773117^{-9}$
- 8^{-37}
  $\tilde{f}'(x) = 0.1169281005859375$
  $|f'(x) - \tilde{f}'(x)| = 1.4181102600652196^{-5}$
- 8^{-45}
  $\tilde{f}'(x) = 0.11328125$
  $|f'(x) - \tilde{f}'(x)| = 0.003661031688538152$
- 8^{-50}
  $\tilde{f}'(x) = 0.0$
  $|f'(x) - \tilde{f}'(x)| = 0.11694228168853815$

7.3 Wnioski

Wartość błędu maleje aż do 2^{-28}, a później już rośnie. Dzieje się to ponieważ dochodzimy coraz bardziej do granic możliwości dokładnego określenia ułamka przez bardzo małą wartość liczb napotkanych w kalkulacjach.