Zad. 2. Wyznaczanie dokładności maszynowej, długości mantysy oraz zakresu liczb zmiennopozycyjnych

e-mail: andrzej.kedziorski@fizyka.umk.pl
pokój: 485B
http://www.fizyka.umk.pl/~tecumseh/EDU/MNII/

Zadanie 2

Napisz program wyznaczający dokładność maszynową (jednostkę zaokrąglenia), a także liczbę bitów mantysy dla liczby zmiennopozycyjnej w pojedynczej i podwójnej precyzji zapisanej w standardzie IEEE 754. Ponadto, program ma wyznaczać możliwy zakres wykładników dla liczb w pojedynczej i podwójnej precyzji. Jaki wpływ na wyniki ma zastosowanie łagodnego niedomiaru?

Liczba zmiennopozycyjna $x = \pm m\beta^e$

- β podstawa systemu liczenia ($\beta = 2$)
- m - mantysa o długości t (liczba bitów w mantysie)

$$m = d_0 + \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \ldots + \frac{d_{t-1}}{\beta^{t-1}},$$

  gdzie $0 \leqslant d_i \leqslant \beta - 1$, $i = 0, \ldots, t-1$
- e cecha, $L \leqslant e \leqslant U$
- Normalizacja $1 \leqslant m \leqslant \beta$

Jednostka zaokrąglenia ϵ_{mach}

- Zaokrąglanie przez obcinanie $\epsilon_{\text{mach}} = \beta^{1-t}$
- Zaokrąglanie do najbliższej $\epsilon_{\text{mach}} = \frac{1}{2}\beta^{1-t}$
- Najmniejsza liczba ϵ taka, że $fl(1+\epsilon) > 1$
- Wyznaczyć ϵ_{mach} na podstawie ostatniej „definicji”; przy okazji możemy wyznaczyć liczbę bitów w mantysie t
- Mając t możemy porównać wyznaczoną ϵ_{mach} z powyższymi definicjami (które definicje są zgodne ze sobą?)
- Obliczenia wykonać w pojedynczej i podwójnej precyzji (czy na pewno wykonano obliczenia w pojedynczej precyzji?)
- Można też np. sprawdzić wartość $|3(4/3-1)-1|$

Jednostka zaokrąglenia ϵ_{mach} - output

- Na wyjściu program wypisuje w kolumnach numer iteracji, ϵ oraz $(1 + \epsilon)$, gdzie ϵ reprezentuje kolejne przybliżenia do ϵ_{mach}
- Na końcu program wypisuje wyniki: t oraz wartości ϵ_{mach} obliczone na różne sposoby
- Porównać wyniki ze standardem IEEE 754

Zakres liczb zmiennopozycyjnych

1. Wyznaczyć najmniejszą dodatnią liczbę zmiennopozycyjną
2. Wyznaczyć najmniejszą dodatnią liczbę zmiennopozycyjną o znormalizowanej mantysie
3. Do powyższego zadania można wykorzystać jednostkę zaokrąglenia ϵ zdefiniowaną jako najmniejszą liczbę taką, że $fl(1+\epsilon) > 1$
4. Wyznaczyć największą możliwą liczbę zmiennopozycyjną (mniejszą od ∞)
5. Obliczenia wykonać w pojedynczej i podwójnej precyzji (czy na pewno wykonano obliczenia w pojedynczej precyzji?)
6. Na wyjściu program wypisuje (trzy) wartości ekstremalne wyznaczone odpowiednio dla pojedynczej i podwójnej precyzji
7. Wyniki odnieść do wartości L, U, poziomu niedomiaru (*underflow*) i nadmiaru (*overflow*) zawartych w standardzie IEEE 754