

## Zadanie 2

March 9, 2021

**Autor: Mateusz Chojnowski**

### Zadania:

1. Wyznaczyć najmniejszą dodatnią liczbę zmiennopozycyjną
2. Wyznaczyć najmniejszą dodatnią liczbę zmiennopozycyjną o znormalizowanej mantysie
3. Do powyższego zadania można wykorzystać jednostkę zaokrąglenia zdefiniowaną jako najmniejszą liczbę taką, że  $fl(1 + e) > 1$
4. Wyznaczyć największą możliwą liczbę zmiennopozycyjną (mniejszą od  $\infty$ )
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

# 1 Program

Na potrzebę zadania został przygotowany program w języku C. Kod programu znajduje się w pliku `main.c` załączonym do pracy.

## 2 Realizacja

2.1 Wyznaczyć najmniejszą dodatnią liczbę zmiennopozycyjną

[illegible]

Dla liczby typu `double` uzyskano liczbę o wartości około  $2e-324$ . Alternatywną drogą do uzyskania minimalnej liczby `double` wykonania potęgowania  $2^{-(1024+51)}$  ( $1024 = 2^{10}$  - wykładnik, 51 - bity mantysy wykorzystane do przechowania wykładnika).

Należy zauważyć że wchodząc wykładnikiem w mantysę (łagodny niedomiar) stopniowo tracimy precyzję zawartą w najmniej znaczących bitach mantysy, spowodowane jest to przesunięciem liczby w mantysie, aby zrobić miejsce na bity niosące informacje o wykładniku.

[illegible]

3 Wyznaczyć jednostkę zaokrąglenia  $e_{mach}$

Uzyskane wartości, 0.00000011920928955078125 dla typu float oraz 0.00000000000000022204460492503131 dla typu double, pokrywają się z wartościami zawartymi w standardzie (w kodzie znajdują się odpowiednie asercje).

Do wyznaczenia maksymalnych wartości `float` i `double` zostały użyte funkcje `calculate_highest_float` oraz `calculate_highest_double`. Funkcje te najpierw wypełniają mantysę typu 1, a następnie multiplikują uzyskaną liczbę o 2 do czasu osiągnięcia wartości maksymalnej mniejszej od  $\infty$ .

2