Министерство образования и науки Российской Федерации

ФГБОУ ВО «НИУ МЭИ»

Кафедра Вычислительных машин, систем и сетей

**Лабораторная работа № 3**

по дисциплине: «Машинная арифметика в рациональных числах»

**Выполнил**: студент группы А-08м-23

Мильчаков Я. С.

Москва 2023

**Целью** **данной лабораторной работы** является изучение работы со специальными значения формата с плавающей точкой и реализацию высокоточных вычислений с плавающей точкой.

Задание №1

Проверить, что для вычисления значения полного сопротивления по формуле



значение выражения при , *R2 = 0* являются специальные числа в формате с плавающей точкой одинарной и двойной точности.

Каков результат при *R1 = 1, а R2 = -0* ипри *R1 = 0, а R2 = -0*

Сравнить значения формулы для полного сопротивления с эквивалентной формулой:

Результаты тестирования представлены на рисунке 1.

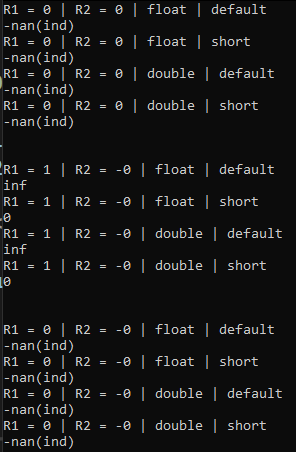


Рисунок 1 – Результаты тестирования

Из результатов тестирования можно сделать выводы, что при значениях 1 и -0 мы имеем расхождение в результате вычисления для значений одинарной и двойной точности. Также в результате тестирования мы получаем специальные значения -nan(ind) – отрицательная неопределенность и inf – бесконечность.

Задание №2

Продемонстрировать нарушение закона ассоциативности в формате с плавающей точкой одинарной точности:



При *a = 2-k , b = 1, c= -1* подобрать *k* чтобы получилось, что правая часть выражения равна *2-k*, а левая часть равна нулю или отличается от правой части.

Коэффициент k подбирался перебором от 0 до 30. Результаты подбора представлены на рисунке 2.

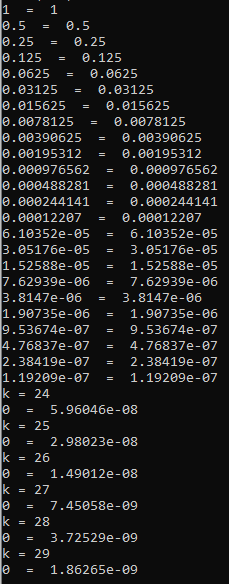


Рисунок 2 – Подбор коэффициента k

Из результатов тестирования можно сделать вывод, что закон ассоциативности нарушается при значении k равном 24 (количество битов для одинарной точности).

Выводы

В ходе лабораторной работы был изучены специальные значения формата с плавающей точкой и реализация высокоточных вычислений с плавающей точкой. Вычисления в ходе лабораторной работы были выполнения при помощи средств программного языка C++.

Из результатов первого задания можно сделать выводы, что при значениях 1 и -0 мы имеем расхождение в результате вычисления для значений одинарной и двойной точности. Также в результате тестирования мы получаем специальные значения -nan(ind) – отрицательная неопределенность и inf – бесконечность.

Из результатов второго задания можно сделать вывод, что закон ассоциативности нарушается при значении k равном 24 (количество битов для одинарной точности).

ПРИЛОЖЕНИЕ А. ЛИСТИНГ КОДА

|  |
| --- |
| #include <iostream>  #include <cmath>  using namespace std;  template<typename T>  void calculateDefault(T a, T b) {  cout << 1 / ( (1 / a) / ( 1 / b) ) << endl;  }  template<typename T>  void calculateShort(T a, T b) {  cout << ( a \* b ) / ( a + b ) << endl;  }  void task1() {  cout << "R1 = 0 | R2 = 0 | float | default" << endl;  calculateDefault((float)0, (float)0);  cout << "R1 = 0 | R2 = 0 | float | short" << endl;  calculateShort((float)0, (float)0);  cout << "R1 = 0 | R2 = 0 | double | default" << endl;  calculateDefault((double)0, (double)0);  cout << "R1 = 0 | R2 = 0 | double | short" << endl;  calculateShort((double)0, (double)0);  cout << endl << endl;  cout << "R1 = 1 | R2 = -0 | float | default" << endl;  calculateDefault((float)1, (float)-0);  cout << "R1 = 1 | R2 = -0 | float | short" << endl;  calculateShort((float)1, (float)-0);  cout << "R1 = 1 | R2 = -0 | double | default" << endl;  calculateDefault((double)1, (double)-0);  cout << "R1 = 1 | R2 = -0 | double | short" << endl;  calculateShort((double)1, (double)-0);  cout << endl << endl;  cout << "R1 = 0 | R2 = -0 | float | default" << endl;  calculateDefault((float)0, (float)-0);  cout << "R1 = 0 | R2 = -0 | float | short" << endl;  calculateShort((float)0, (float)-0);  cout << "R1 = 0 | R2 = -0 | double | default" << endl;  calculateDefault((double)0, (double)-0);  cout << "R1 = 0 | R2 = -0 | double | short" << endl;  calculateShort((double)0, (double)-0);  }  float sum(float a, float b) {  return a + b;  }  void task2() {  float b = 1;  float c = -1;  float a;  for (int i = 0; i < 30; i++) {  a = pow(2, -i);  float sum1 = sum(sum(a, b), c);  float sum2 = sum(a, sum(b, c));  if (sum1 != sum2) {  cout << "k = " << i << endl;  }  cout << sum1 << " = " << sum2 << endl;  }  }  int main()  {  task1();  task2();  } |