CENG 371 SCIENTIFIC COMPUTING

HW1

Name: Andaç Berkay Seval

ID: 2235521

Requirements: I did my homework with Python. Therefore, there are 2 requirements:

- NumPy
- Matplotlib

Q1)

a)

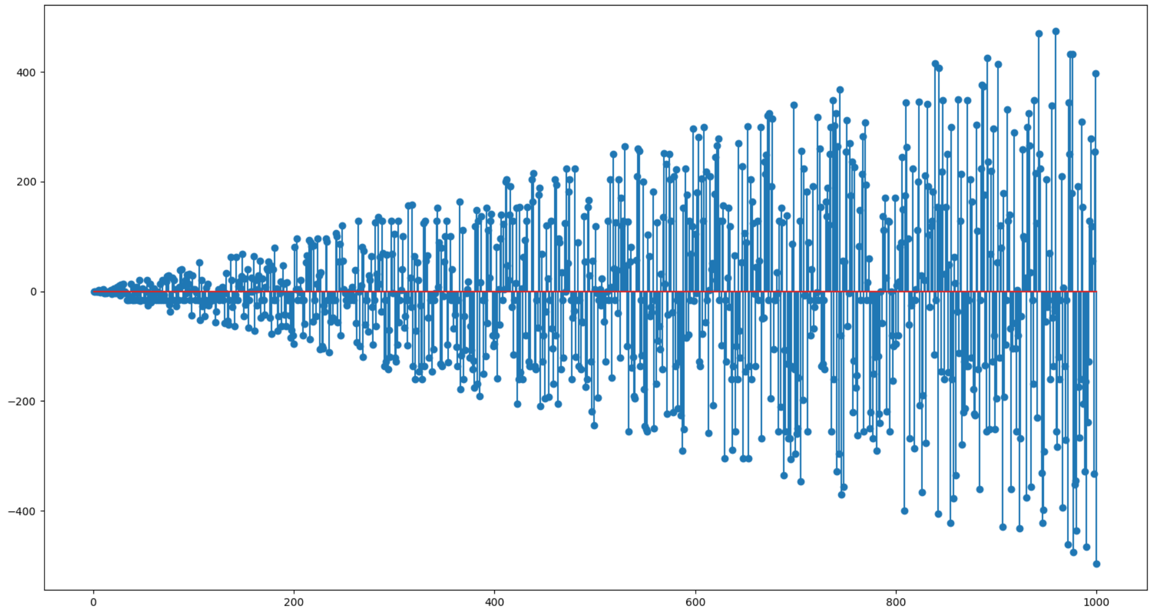

Figure 1: Graph of n - g(n) . (x-axis is n and y-axis is g(n))

b) These are the n values where g(n) = 0 : 1, 2, 4, 8, 16, 32, 64, 128, 256, 512.

c) For majority of ns g(n) is not equal to 0 because of the rounding error which is the difference between the result produced by an exact arithmetic algorithm and the result produced by finite precision, rounded algorithm. It is the result of the cancellation error where two approximate and nearly equal numbers are subtracted. In the equation, for majority of ns, (n+1) / n – 1 does not give the exact arithmetic solution, but an approximation. Therefore, rounding error occurs in these situations.

d) Since the cancellation error increases. When n gets large, the number (n+1) / n gets smaller, and the difference between (n+1) / n and 1 also gets smaller. Since (n+1) / n is an approximate number for majority of ns and, (n+1) / n and 1 are nearly equal numbers for large ns, rounding error increases while n is getting larger.

Q2)

a)

$$\text{sum} = \sum_{n=1}^{10^6} 1 + (10^6 + 1 - n).10^{-8} = \sum_{n=1}^{10^6} 1 + 10^{-2} + 10^{-8} - n10^{-8} = 10^6 + 10^4 + 10^{-2} - 10^{-8}.\sum_{n=1}^{10^6} n = 10^6 + 10^4 + 10^{-2} - 10^{-8}(10^6)(10^6+1)/2 = 1005000.005$$

b) It sums up a sequence of finite precision floating point numbers and reduces the accumulated rounding error compared to naïve summation. Pairwise summation is almost as good as compensated summation in terms of rounding errors while having lower computational cost.

c)

i. **naïve summation:**

single precision: 1005000.004999876

double precision: 1005000.0049999995

ii. **compensated summation:**

single precision: 1005000.004999876

double precision: 1005000.005

iii. **pairwise summation:**

single precision where the partitions have at most 1000 elements: 1005000.004999876

double precision where the partitions have at most 1000 elements: 1005000.0050000001

single precision where the partitions have at most 10000 elements: 1005000.004999876

double precision where the partitions have at most 10000 elements: 1005000.005

d)

i. **naïve summation with single precision:**

The absolute error of naive summation with single precision is : 1.2398231774568558e-07

The time of execution of single precision naive summation is : 290.78030586242676 ms

ii. **naïve summation with double precision:**

The absolute error of naive summation with double precision is : 4.656612873077393e-10

The time of execution of double precision naive summation is : 243.0415153503418 ms

iii. **compensated summation with single precision:**

The absolute error of compensated summation with single precision is : 1.2398231774568558e-07

The time of execution of single precision compensated summation is : 628.8518905639648 ms

iv. **compensated summation with double precision:**

The absolute error of compensated summation with double precision is : 0.0

The time of execution of double precision compensated summation is : 590.2070999145508 ms

v. **pairwise summation with single precision where the partitions have at most 1000 elements:**

The absolute error of pairwise summation with single precision is : 1.2398231774568558e-07

The time of execution of single precision pairwise summation is : 244.6136474609375 ms

vi. **pairwise summation with double precision where the partitions have at most 1000 elements:**

The absolute error of pairwise summation with double precision is : 1.1641532182693481e-10

The time of execution of double precision pairwise summation is : 190.37318229675293 ms

vii. **pairwise summation with single precision where the partitions have at most 10000 elements:**

The absolute error of pairwise summation with single precision is : 1.2398231774568558e-07

The time of execution of single precision pairwise summation is : 236.7103099822998 ms

viii. **pairwise summation with double precision where the partitions have at most 10000 elements:**

The absolute error of pairwise summation with double precision is : 0.0

The time of execution of double precision pairwise summation is : 188.53020668029785 ms

Therefore, in single precision, all the methods give the same error. However, in double precision, compensated summation does not give an error, pairwise summation gives a smaller error than naïve summation where partitions have at most 1000 elements. Moreover, pairwise summation does not give an error where partitions have at most 10000 elements. Also, all the methods give less error in double precision than in single precision.

Furthermore, compensated summation has the largest execution time, then naïve summation, and pairwise summation has the least execution time. Also, all the methods have larger execution time in single precision than in double precision.

e) From my findings, I can say that summations should done in double precision as much as possible since the accuracy is higher for any method than in single precision. Also, in a situation where the accuracy is the main purpose and the execution time is not that important as accuracy, compensated summation should be chosen since it gives the best accuracy with reducing the accumulated rounding errors in best amount of. However, in a situation where execution time should also be considered, pairwise summation is the way to go. It is almost as much accurate with compensated summation (there is a logarithmic difference) and the computational cost is much lower.