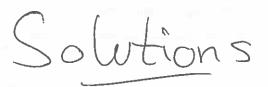
# Introduction to Scientific Computing, PSCB57, Midterm

Professor Hanno Rein University of Toronto, Scarborough

Monday, October 29th, 2018



- Duration: 50 minutes.
- This midterm is worth 25% of your final grade.
- No aid sheets, books or other notes are allowed.
- A basic calculator is allowed, although not needed. No programmable calculators or any other electronic devices are allowed.
- Write your answers directly on the question sheet. If you need scrap paper raise your hand.
- The University of Toronto's Code of Behaviour on Academic Matters applies to all University of Toronto Scarborough students. The Code prohibits all forms of academic dishonesty including, but not limited to, cheating, plagiarism, and the use of unauthorized aids. Students violating the Code may be subject to penalties up to and including suspension or expulsion from the University.

7 Points

Convert the following IEEE754 double floating point numbers from their binary representation to the standard decimal format:

0.375

<u>-9</u>

Also convert the following IEEE754 double floating point number to decimal notation. Here, it is ok to give an approximate answer.

<u>2 −3e305</u>

2 +8e-305

Circle all numbers that can be approximated by a finite IEEE754 double floating point number:





$$7^{8312} - 1$$

Circle all expressions that result in a division by zero when implemented using IEEE754 double floating point numbers:

$$\frac{11}{(11111+10^{-15})-11100-11}$$

$$\frac{101}{(11+10^{-11})-10-1}$$

$$\begin{array}{c|c}
 & 101 \\
\hline
 & (11 + \pi^{-10} \cdot 10^{-11}) - 10 - 1
\end{array}$$

In python 3, what are the result of the following expressions?

$$4.0 / 2.0 == 2.0$$

False

True

Question 2 5 Points

Instead of using an integer or floating point representation, some computer programs also use what is called a fixed point number representation. Assume we have 8 bits,  $b_i$ , i = 0, ... 7 to represent a number x in the following way:

$$x = (-1)^{b_0} \sum_{i=1}^{7} b_i 2^{i-4}.$$

What is the largest number than can be represented (an approximate answer is ok)?

Add the following 8 bit fixed point numbers in their binary representation, then convert the result into the normal decimal format: 01100110, 00000101.

Mark all statements which are true:

- Negative numbers can be represented using this number representation.
- Additions and subtractions are exact when using this number representation.
- Any number between 0 and 28 can be represented exactly using this scheme.
- An 8-bit fixed point number uses more memory than a double precision floating point number.
- Some fractions can be represented exactly using this number representation, for example  $\frac{1}{8}$ .

Question 3 4 Points

Write down the central finite difference formulas (second order) for the first, second and third derivatives!

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

$$f'''(x) \approx \frac{f(x+2h) - 2 f(x+h) + 2 f(x-h) - f(x-2h)}{2h^3}$$

Hint: we did not discuss the equation for the third derivative in detail during the lectures. Try to construct it in an analogous way to the first two derivatives, then test it with a function where you know the answer, for example with a third order polynomial.

5 Points

Consider the follwing integral:

$$\int_0^{2\pi} \sin(x) dx.$$

What is the exact value of this integral?

		<u>ر</u>												
Now,	discretize	the	integral	by	splitting	the	interval	$[0,2\pi]$	into	four	sub-intervals.	Then	use	the
trapezoidal rule to calculate the value of the integral numerically.														

Consider the integral once again, but this time over the interval  $[0, \pi]$ . Split this integral into four sub-intervals and use the trapezoidal rule to calculate the value of the integral numerically. Hint:  $\sin\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$ .

其(12+1)

Explain the concept behind Simpson's rule in contrast to the trapezoidal rule! Would you expect a better approximation to the integral for the above case or not?

- Simpson's rule uses piece-mise que	adratic interpolation
- Simpson's rule uses piece-mise que - Simpson's rule doubt give a be	Her result
300000000000000000000000000000000000000	

6 Points

Use the bisection method on the function  $f(x) = x^2 - 1$ . Start with the interval [0, 16]. After 4 iterations, what is the interval in which you expect the root?

[0,1]

How often would you have to iterate to reach machine precision?

~52 Eimes

Write down the update rule (how to calculate  $x_{n+1}$  from  $x_n$ ) for Newton's method and the specific function f(x) from above:

 $\frac{\chi_{h+1} = \chi_h - \frac{\chi_h^2 - 1}{2 \chi_h}}{\chi_{h+1}}$ 

Which method will converge to the true solution faster in the above case?

any linear function

Describe a real world situation where you might use a root finding method.

any type of optimization problem

3 Points

Below is a rather long assembler program. You'll find the syntax for the assembler commands used in the appendix (note that there are empty lines). What does this program do? Be concise and precise in your answer! Also add some comments to the program describing the various parts!

Standard map

```
SET 1. r0
SET 0. r1
SET 3. r2
SET
SET
SET
SET
SET
        30 r10
0 r7
1000 r8
         1 r9
6.283185307179586 r4
        3 r5
-3 r6
SET
SET
SIN r2 r3
MUL r0 r3
ADD r3 r1
ADD r2 r1
                                               actual standard Map
                    r3
r1
r2
IF r4 r1
JUMP r5
SUB r1 r4 r1
JUMP r6
IF r4 r2
JUMP r5
SUB r2 r4
JUMP r6
IF r1 r7
JUMP r5
ADD r1 r4 r1
JUMP r6
IF r2 r7
JUMP r5
ADD r2
JUMP r6
               r4 r2
PRINT r1
PRINT r2
SUB r8 r9 r8 IF r8 r7 JUMP r10
                                                                    1000 Eines
```

Question 7 2 Bonus Points

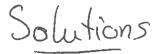
Convert the assembler program from the previous question into a python 3 program! Small syntax errors will not be penalized.

(see tutorie (assignment)

## Introduction to Scientific Computing, PSCB57, Midterm

Professor Hanno Rein University of Toronto, Scarborough

Monday, October 29th, 2018

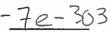


- Duration: 50 minutes.
- This midterm is worth 25% of your final grade.
- No aid sheets, books or other notes are allowed.
- A basic calculator is allowed, although not needed. No programmable calculators or any other electronic devices are allowed.
- Write your answers directly on the question sheet. If you need scrap paper raise your hand.
- The University of Toronto's Code of Behaviour on Academic Matters applies to all University of Toronto Scarborough students. The Code prohibits all forms of academic dishonesty including, but not limited to, cheating, plagiarism, and the use of unauthorized aids. Students violating the Code may be subject to penalties up to and including suspension or expulsion from the University. (2)

7 Points

Convert the following IEEE754 double floating point numbers from their binary representation to the standard decimal format:

Also convert the following IEEE754 double floating point number to decimal notation. Here, it is ok to give an approximate answer.



Circle all numbers that can be approximated by a finite IEEE754 double floating point number:

 $\pi \cdot 10^{-428}$ 

42.00000

 $5^{-356} + 1$ 

 $2^{7734}$ 

Circle all expressions that result in a division by zero when implemented using IEEE754 double floating point numbers:

$$\frac{101}{(11 + e^{-10} \cdot 10^{-11}) - 10 - 1}$$

$$\begin{array}{c|c}
11 \\
\hline
(111+10^{-19})-100-11
\end{array}$$

$$\frac{101}{(111001+10^{-1})-111000-1}$$

In python 3, what are the result of the following expressions?

Question 2 5 Points

Instead of using an integer or floating point representation, some computer programs also use what is called a fixed point number representation. Assume we have 8 bits,  $b_i$ , i = 0, ... 7 to represent a number x in the following way:

$$x = (-1)^{b_0} \sum_{i=1}^7 b_i 2^{i-4}.$$

What is the largest number than can be represented (an approximate answer is ok)?

Add the following 8 bit fixed point numbers in their binary representation, then convert the result into the normal decimal format: 00110110, 01000011.

$$\frac{O(11 \text{ (loc)})}{\text{(binary representation)}} = \frac{15.125}{\text{(decimal format)}}$$

Mark all statements which are false:

- Any number between 0 and  $2^8$  can be represented exactly using this scheme.
- Additions and subtractions are exact when using this number representation.
- Negative numbers can be represented using this number representation.
- Some fractions can be represented using this number representation, for example  $\frac{1}{8}$ .
- An 8-bit fixed point number uses more memory than a double precision floating point number.

Question 3 4 Points

Write down the central finite difference formulas (second order) for the first, second and third derivatives!

$$f'(x) \approx \frac{\int (x+h) - \int (x-h)}{2h}$$

$$f''(x) \approx \frac{f(x+h)-2f(x)+f(x-h)}{h^2}$$

$$f'''(x) \approx \frac{f(x+2h)-2f(x+h)+2f(x-h)-f(x-2h)}{2h^3}$$

Hint: we did not discuss the equation for the third derivative in detail during the lectures. Try to construct it in an analogous way to the first two derivatives, then test it with a function where you know the answer, for example with a third order polynomial.

5 Points

Consider the follwing integral:

$$\int_0^{2\pi} \cos(x) dx.$$

What is the exact value of this integral?

		<u>し</u>												
Now,	discretize	the	integral	by	splitting	the	interval	$[0,2\pi]$	into	four	sub-intervals.	Then	use	the
trapez	zoidal rule	to c	alculate	the	value of t	he i	ntegral n	umeric	ally.					

Consider the integral once again, but this time over the interval  $[-\pi/2, \pi/2]$ . Split this integral into four sub-intervals and use the trapezoidal rule to calculate the value of the integral numerically. Hint:  $\cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$ .

平(12+1)

Explain the concept behind Simpson's rule in contrast to the trapezoidal rule! Would you expect a better approximation to the integral for the above case or not?

- Simpson's rule should be better.

Question 5 Use the bisection method on the function  $f(x) = x^4 - 1$ . Start with the interval [0, 16]. After 4

6 Points

iterations, what is the interval in which you expect the root?

How often would you have to iterate to reach machine precision?

Write down the update rule (how to calculate  $x_{n+1}$  from  $x_n$ ) for Newton's method and the specific function f(x) from above:

Xh+1= xn - xn+-1

Which method will converge to the true solution slower in the above case?

Si section method g(x) does Newton's method converge within 1 step?

Describe a real world situation where you might use a root finding method.

any type of aprimization problem

Question 6 3 Points

Below is a rather long assembler program. You'll find the syntax for the assembler commands used in the appendix (note that there are empty lines). What does this program do? Be concise and precise in your answer! Also add some comments to the program describing the various parts!

SET SET SET 0.8 r0 3. r1 2. r2 -30 r10 0 r7 2000 r8 1 r9 6.283185307179586 r4 SET SET ŠĒT SET 3 r5 -3 r6 SET SET SIN r2 r3 MUL r0 r3 ADD r3 r1 ADD r2 r1 actual standard map r3 r1 r2 IF r4 r1 JUMP r5 SUB r1 r4 r1 JUMP r6 IF r4 r2 JUMP r5 SUB r2 r4 r2 JUMP r6 Zpi on p and theta IF r1 r7 JUMP r5 ADD r1 r4 r1 JUMP r6 IF r2 r7 JUMP r5 ADD r2 r4 r2 JUMP r6 PRINT r1 PRINT r2 SUB r8 r9 r8 IF r8 r7 JUMP r10 Eimes

Question 7 2 Bonus Points

Convert the assembler program from the previous question into a python 3 program! Small syntax errors will not be penalized.

See assignment