20240913

Page 15, Problem 14. Page 19, Exercises 1, 2; Computer Problems 1,4.

14. Do the following operations by hand in IEEE double precision computer arithmetic, using the Rounding to Nearest Rule. (Check your answers, using MATLAB.)
(a) $(4.3 - 3.3) - 1$ (b) $(4.4 - 3.4) - 1$ (c) $(4.9 - 3.9) - 1$

a. $(fl(4.3) - fl(3.3)) - fl(1)$

$4_{10} = 100_2$

$3.3 = 11.0\overline{1001}$

$fl(1) = 100\cdots0$

$0.3 \mid \times 2$

0.6
1.2
0.4
0.8
1.6
1.2

$0.0\overline{1001}$

$0.3 \mid \times 2$

$4.3_{10} = 100.0\overline{1001}_2$

$$fl(4.3) = (-1)^0 \times 2^{1025-1023} \times 1.\underbrace{\boxed{0001001\,1001\cdots0011}}^{52\,bit}\,0011$$

$$fl(3.3) = (-1)^0 \times 2^{1024-1023} \times 1.\boxed{0100110\cdots01100110}\,0110$$

$fl(4.3) - fl(3.3)$

$= 100.01001\cdots0011 - 11.01001$

$= 1$

$(fl(4.3) - fl(3.3)) - 1$

$= 1 - 1$

$= 0$

b.) $(4.4 - 3.4) - 1$

$fl(4.4) = 100.0110 \cdots 0110$ $\quad$ $fl(3.4) = 11.0110 \cdots 0110$

0.4	×2
0.8	
1.6	
1.2	
0.4	

$$
\begin{aligned}
fl(4.4) - fl(3.4) &= 1.000110 \cdots 0011010 \times 2^2 \\
&\quad - 1.1\,0110 \cdots 00110011 \times 2 \\
&= 1.\underbrace{000 \cdots 01}_{51\text{ bit}}
\end{aligned}
$$

$$
\begin{aligned}
(fl(4.4) - fl(3.4)) - fl(1) &= 1.00 \cdots 01 - 1 \\
&= 0.\underbrace{0000 \cdots 01}_{51\text{ bit}} \\
&= 2^{-51}
\end{aligned}
$$

C. $(4.9-3.9)-1$

$fl(4.9) = 1.\boxed{0011\,1100\,1100\cdots1001}\,1001 \times 2^2$

1010

$fl(3.9)$

$= 1.\boxed{1111\,0011\,0011\cdots0011}\,0011 \times 2^1$

0011

0.9	$\times 2$
1.8	
1.6	
1.2	
0.4	
0.8	

$fl(4.9) - fl(3.9)$

$= 0.0100\cdots01 \times 2^2$

$= 1.00\cdots01$

$= 2^{-51}$

1. Identify for which values of x there is subtraction of nearly equal numbers, and find an alternate form that avoids the problem.

(a) $\dfrac{1-\sec x}{\tan^2 x}$ (b) $\dfrac{1-(1-x)^3}{x}$ (c) $\dfrac{1}{1+x}-\dfrac{1}{1-x}$

a. the loss of sign number will occure when x is close to 0

$$\tan^2 x = \frac{1}{\cos^2 x} - 1$$

$$= \sec^2 x - 1$$

$$= \frac{(1-\sec x)(1+\sec x)}{(1-\sec^2 x)(1+\sec x)}$$

$$= -\frac{1}{(1+\sec x)}$$

$$= -\frac{1}{1+\sec x}$$

b. the loss of sign number will occure when x is close to 0

$$(1-x)^3 = 1-3x+3x^2-x^3$$

$$\frac{\cancel{1}-\cancel{1}-3x+3x^2-x^3}{x}$$

$$= -3+3x+x^2$$

$$= x^2+3x-3$$

c. the loss of sign number will occure when x is close to 1

$$\frac{1}{1+x} - \frac{1}{1-x}$$

$$= \frac{(1-x)-(1+x)}{(1+x)(1-x)}$$

$$= \frac{1-1-2x}{1-x^2}$$

$$= \frac{-2x}{1-x^2}$$

2. Find the roots of the equation $x^2 + 3x - 8^{-14} = 0$ with three-digit accuracy.

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$x_1 = \frac{-3 - \sqrt{9 + 4 \cdot 8^{-14}}}{2} \approx -3.000$$

$$x_2 = \frac{-3 + \sqrt{9 + 4 \cdot 8^{-14}}}{2} \approx 7.58 \times 10^{-14}$$

1. Calculate the expressions that follow in double precision arithmetic (using MATLAB, for example) for $x = 10^{-1}, \ldots, 10^{-14}$. Then, using an alternative form of the expression that doesn't suffer from subtracting nearly equal numbers, repeat the calculation and make a table of results. Report the number of correct digits in the original expression for each x.

$$\text{(a)} \quad \frac{1 - \sec x}{\tan^2 x} \qquad \text{(b)} \quad \frac{1 - (1 - x)^3}{x}$$

$$\frac{1-\sec x}{\tan^2 x} \longrightarrow -\frac{1}{1+\sec x}$$

a.

	x	$\frac{1-\sec x}{\tan^2 x}$	$-\frac{1}{1+\sec x}$
10^{-1}	0.10000000000000	-0.49874791371143	-0.49874791371147
10^{-2}	0.01000000000000	-0.49998749979046	-0.49998749979166
10^{-3}	0.00100000000000	-0.49999987501429	-0.49999987499948
10^{-4}	0.00010000000000	-0.49999999362791	-0.49999999482500
10^{-5}	0.00001000000000	-0.50000000413685	-0.49999999999987
10^{-6}	0.00000100000000	-0.50004445029084	-0.50000000000000
10^{-7}	0.00000010000000	-0.51070259132757	-0.50000000000000
10^{-8}	0.00000001000000	0	-0.50000000000000
10^{-9}	0.00000000100000	0	-0.50000000000000
10^{-10}	0.00000000010000	0	-0.50000000000000
10^{-11}	0.00000000001000	0	-0.50000000000000
10^{-12}	0.00000000000100	0	-0.50000000000000
10^{-13}	0.00000000000010	0	-0.50000000000000
10^{-14}	0.00000000000001	0	-0.50000000000000

a.

	x	$\frac{1-(1-x^3)}{x}$	x^2+3x-3
10^{-1}	0.100000000000000	2.71000000000000	2.71000000000000
10^{-2}	0.010000000000000	2.97010000000001	2.97010000000000
10^{-3}	0.001000000000000	2.99700100000000	2.99700100000000
10^{-4}	0.000100000000000	2.99970000999905	2.99970001000000
10^{-5}	0.000010000000000	2.99997000015263	2.99997000010000
10^{-6}	0.000001000000000	2.99999969866072	2.99999700000100
10^{-7}	0.000000100000000	2.999997000015263	2.99999970000001
10^{-8}	0.000000010000000	2.999999969866072	2.99999997000000
10^{-9}	0.000000001000000	2.999999991515421	2.99999999700000
10^{-10}	0.000000000100000	3.00000024822111	2.99999999970000
10^{-11}	0.000000000010000	3.00000024822111	2.99999999997000
10^{-12}	0.000000000001000	2.99993361481964	2.99999999999700
10^{-13}	0.000000000000100	3.00093283556180	2.99999999999970
10^{-14}	0.000000000000010	2.99760216648792	2.99999999999997

4. Evaluate the quantity $\sqrt{c^2+d}-c$ to four correct significant digits, where $c = 246886422468$ and $d = 13579$.

$$= \sqrt{c^2+d} - c$$

$$= \frac{(\sqrt{c^2+d} - c)(\sqrt{c^2+d} + c)}{\sqrt{c^2+d} + c}$$

$$= \frac{c^2+d-c^2}{\sqrt{c^2+d} + c}$$

$$= \frac{d}{\sqrt{c^2+d} + c}$$

$$= 2.750 \times 10^{-8}$$